Conference papers

Tamil Internet - 2010

Coimbatore, June 23 – 27
iii
## CONTENTS

1. **ICT for Tamil Education in Tamil Nadu - Current Challenges**
   - Sivakumar Arunachalam, National Institute of Education, Singapore
   - Pages: 001-003

2. **Use of Technology in running a Tamil School in USA**
   - Ilango Meyyappan, California Tamil Academy
   - Pages: 008-013

3. **Teaching and Resource Building in Teacher Education**
   - Nirmala Devi, Govt. Institute of Adv Study in Education
   - Pages: 014-019

4. **Video Repositories: A Data Mining Based Approach.**
   - Enhancing Activity based Tamil Teaching and Learning Using Online Video Repositories: A Data Mining Based Approach.
   - V. Saravanan, Dr. NGP Institute of Technology, Chennai.
   - Pages: 047-052
11. The Computer, Tamil Language, The Tamil Teacher
N.Vairamani, Rajarajeswari Engineering College

12. Using Computer and Internet resources in Tamil Teaching
Govindaraj. A, Meenakshi University, Chennai.

13. Tamil Education through Classroom Computer
K. G.Singaravelu, Bharathiar University, Coimbatore

14. Teaching Toddlers – An Analysis
N.Vairamani, Rajarajeswari Engineering College

15. Uses of Computer Technology to Teach Tamil Language
G.Sankar Thilahavathi, Bowen Secondary School, Singapore

16. Open Education System, Tamil as the First Language.
Balasubramanian, Bharathiar University, Coimbatore

17. Application of ICT in Teaching Tamil in England
Jayasarasvathi DuraiKannu, Bedok West Primary School, Singapore

18. Knowledge based Learning framework
Sankar Thilahavathi, Bowen Secondary School, Singapore

19. Computer Aided Learning in Tamil Sentences
G.Singaravelu, Bharathiar University, Coimbatore
21. Developing Tamil Language Capability amongst Students through Computers
Sivagouri Kaliamurthy, Beacon Primary School, Singapore

(2) Tamilamalar

1) Teaching Tamil through Google
P. R. Nakkeeran, Tamil Virtual University

2) Moodle: A Tool for Tamil Teaching
Murugan, Bendemeer Sec Sch, Gnanasekaran, St. Theresa’s Convent, Mohan, Cresent Girls Secondary School

3) Moodle: For Enhanced Learning and Teaching of Tamil language
Ravishankar Somasundaram

4) Quality analysis of Tamil Virtual University
S. Rajkumar, Kumaraguru College of Technology

5) Quality of Data Driven Approach for Tamil Morphological Generator
Mannar Mannan Maruthai, University Malaya, Malaysia

6) Moodle: A Tool for Tamil Teaching
K. Sarveswaran, Sri Lanka

(3) Vayyilamalar

1) Morphological Generator for Tamil: A New Data Driven Approach
Anand kumar M, Dhanalakshmi V, Soman K.P., Amrita Vishwa Vidyapeedham Rajendran, S., Tamil University

2) Verbal Forms in the Tevāram, in the light of the Tivakaram
Jean-Luc Chevillard, CNRS University, Paris
3) Conceptual Lexicon for Knowledge Representation
K. Umaraj, Madurai Kamaraj University, Madurai

4) Nature of Tamil Data and Computing
P. David Prabhakar, Madras Christian College

5) FaceWaves: A Tamil Text to Video Framework
Madhan Karky, T V Geetha, Ravi Varman, College of Engineering Guindy, Anna University

6) Context Based Information Search for Thirukural
G. V. Uma, Ilakiyaselvan N, College of Engineering Guindy, Anna University

7) A Context-free Grammar and a Method to Parse Verses in Meter
Ishwar Sridharan, Bala Sundara Raman

8) Computational Approaches for Learning Inflections in Tamil
K. Rajan, Muthiah Polytechnic College

9) Syntactic Parsers
Marappa Ganesan, Annamalai University.

10) Conceptual Lexicon for Knowledge Representation
S. Rajendran, Tamil University, Thanjavur

11) Computational Approaches for Learning Inflections in Tamil
N. Deiva Sundaram, Madras University.
13) Noun Phrase Chunker Using Finite State Automata for an Agglutinative Language
Vijay Sundar Ram R, AU-KBC Research Centre, Anna University.

14) Animated Sangathamizh Poems
S. Srinivasan, Indira Gandhi Centre for Atomic Research

15) Representation of Kinship in WordNet
S Arulmozi, Dravidian University.

16) Thiruvalluvan. I, Madras University
Arul Natarajan, Tamil Development Department

17) Role of Regular Expression in Morphological Analysis
R. Shanmugam, Madras University


19) Transliteration Schemes for Tamil to Roman and Roman to Tamil
S. Srinivasan, Indira Gandhi Centre for Atomic Research

20) Tamil in Computers, A viewpoint
A Kumaran, Microsoft Research Centre, Bengalooru

21) Automated Identification of Grammatical Patterns for Tamil Poems
S. Sendhilkumar, G.S. Mahalakshmi, N. Prakash, Anna University

(4) Tamil Bibliography

1) Thiruvalluvan. I, Madras University

2) R. Shanmugam, M. Prakash, S. Arulmozi, Dravidian University
3) Spell Checker for Tamil using Finite State Automata
   P. Vasuki, SSN College of Engineering, Chennai.

4) Automated Processing of Census Forms in Tamil
   Premalatha.R, Aarupadai Veedu Institute of Technology
   S.Baskaran, Tamil University, Thanjavur.

5) Language Tools for Tamil - Standardization for Computerization
   A G Ramakrishnan, Indian Institute of Science
   Shashi Kiran, Indian Institute of Science
   Saravanan, Amrita Vishwa Vidyapeetham.

6) Pattern Based English-Tamil Machine Translation
   Anitha. S, Pillai, Hindustan University
   Shashi Kiran, Indian Institute of Science

(5) Tamil

1) Bilingual TTS for Tamil and English
   A G Ramakrishnan, Indian Institute of Science

2) Voice Based Tamil Tutoring System
   P. Vasuki, SSN College of Engineering, Chennai.

3) Prediction of Pauses in TTS – Tamil
   P. Arulmozh. Arulmozh, Indian Institute of Science

4) Text analysing & retrieval system using Tamil phonemes & VSM
   Premalatha.R, Aarupadai Veedu Institute of Technology

5) Speech Tool in Tamil for Speech Disabled
   Shashi Kiran, Indian Institute of Science
6) **Standardisation of Modern Standard Tamil Transcription for Computational Tamil**
   T. Jayasankar, Anna University

7) **Realization of Tamil Syllables - Text To Speech Transferring System Using FPGA**
   K. An Analysis of Various Types of Distortions of Tamil Scripts
   Punal K. Murugaiyan, Annamalai University

8) **Face Waves: Tamil Text-To-Speech with Lip Synchronisation**
   A High Accuracy Phone Recognition System for Tamil
   T. Jayasankar, Anna University, Tiruchirappalli, J. Arputha Vijaya Selvi, Kings College of Eng. Pudukkottai, R. Rajendran, Tamil University, Thanjavur.

9) **Classical Encryption Techniques for Tamil Text**
   J. Jai Hari Raju, College of Engineering Guindy, Anna University

10) **Statistical Analysis & Visualization of Tamil in Text Streams**
    P. Navaneethan, PSG College of Technology, Coimbatore

11) **An Analysis of Various Types of Distortions of Tamil Scripts**
    J. Iyyakutti, C. Jothis Venkeswaran, R. Indra Gandhi, Mother Teresa Women's University

12) **A High Accuracy Phone Recognition System for Tamil**
    C. P. Santhosh Kumar, Madras University, N. Deiva Sundaram, Amrita Vishwa Vidyapeetham, Coimbatore.

13) **Face Waves: 2D Facial Expressions Based on Tamil Emotion Descriptors**
    Sabitha Tammaneni, Madhan Karky, College of Engineering Guindy, Anna University
1) E-Library of University of Malaya  
E. Ilangkumaran

2) Development of Tamil Digital Libraries  
K. Kalyanasundaram, Swiss Federal Institute of Technology

3) Reducing Digital Divide in Tamilnadu using Data Mining Technology  
R. Jayabrabu, Karunya University

4) An Extended Cross Lingual Information Retrieval System for Agricultural Domain Using Statistical Document Translation for Tamil Farmers  
D. Thenmozhi et al. SSN College of Engineering, Chennai

5) Retrieving Tamil Agricultural Documents Using Cross-Measure Matching and Language Identification  
D. Thenmozhi et al. SSN College of Engineering, Chennai

6) Tamil E-Archives Management  
B. Asadullah, Mazharul Uloom College, Ambur.

7) Long-term Preservation and Archival Strategies for Tamil Documents  
Mani M. Manivannan, Symantec Corporation.

8) A view of the Tamil Dictionaries in the Internet  
Maniyrasann Muniandy, Malaysia

9) Tamil Books in the Internet  
K. Duraiarasan, Government Arts College, Kumbakonam

10) Electronic Museum  
Maraimalai Ilakkuvanar, Presidency college, Chennai
11) பிற்புதிக் கிரிஸ்பர் தமிழ் பிற்புதிக் கல்விக் காலாக்கையில்
இணையத்திற்கு செயல்பாடு
Tamil Wikipedia
பிற்புதி கிரிஸ்பர் Theni M. Subramani, www.muthukamalam.com

12) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் மொழிப் பார்வை
Indexing in Tamil through Computers
இணையத்திற்கு செயல்பாடு
L. Sundaram, Bharathidasan University, Tiruchirapalli

(7) தமிழ்விளக்கமாயிற்று மோதம் மேலும் பின்னணியில் பார்வை

1) பிற்புதிக் கிரிஸ்பர் ஏழுநலையில் எழுத்துச் செயல்பாடு, பிற்புதிக் கல்விக் காலாக்கையில் அமைப்பு
Fostering Tamil Web Communities for mining Tamil Web Pages
இணையத்திற்கு செயல்பாடு
A. Vijaya Kathiravan, K.S.R. College of Technology, Tiruchengodu.

2) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை (2010-2021)
Tamil IDN

3) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை (2010-2021)
A Tamil Web Portal Development with online Dictionary
இணையத்திற்கு செயல்பாடு
A Parimaladevi, Kumaraguru College of Technology.

4) பிற்புதிக் கிரிஸ்பர் ஏழுநலையில் கல்வி மேலும்-2010 - பின்னணியில்
Methods and Options of Videoconferencing in relation to the Tamil Language in 2010,
இணையத்திற்கு செயல்பாடு
Eric Miller, University of Pennsylvania.

5) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை (2010-2021)
Thin Client and Server Based Computing to Provide Integrated School and
Class Room Management System in Malaysian Tamil Schools
இணையத்திற்கு செயல்பாடு
Samintha Kumaran Veloo, Saravanan Mariappan, Nexus IT Solutions, Malaysia.

(8) தமிழ் பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை

1) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை (2010-2021)
Embedding with optical co-features to 'OCR-Friendly' Fonts
இணையத்திற்கு செயல்பாடு
ND Logasundaram.

2) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை (2010-2021)
Creation of Annotated Tamil Handwritten Word Corpus for OCR
இணையத்திற்கு செயல்பாடு
Nethravathi B., Indian Institute of Science

3) பிற்புதிக் கிரிஸ்பர் முன்னணிக் தமிழ் பார்வை (2010-2021)
Difficulties in Developing OCR for Tamil Documents
இணையத்திற்கு செயல்பாடு
R. Dhivyabharathi, Bannari Amman Institute of Technology
4) Recognition of Ancient Tamil Characters
S. Rajakumar, DMI College of Engineering

5) Design and Evaluation of Omnifont Tamil OCR
Tushar Patnaik, CDAC, Noida

6) Approach to Recognize Handwritten Tamil Characters
N. Shanthi, K. S. Rangasamy College of Technology

7) Effective Tamil Character Recognition in Tablet PCs Using Pattern Recognition
Ferdin Joe J, R. Velayutham, Einstein College of Engineering,
T. Ravi, KCG College of Technology

(9) Tamil Language Independent Emotion Recognition System

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Authors and Institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Design of Intelligent Robotic Arm Writing Tamil</td>
<td>Seima Saki S, Sri Krishna College of Engg &amp; Tech</td>
</tr>
<tr>
<td>2</td>
<td>LaaLaLaa - A Tamil Lyric Analysis &amp; Generation Framework</td>
<td>Sowmiya. D, Anna University</td>
</tr>
<tr>
<td>3</td>
<td>Language Independent Emotion Recognition System for Web Articles using NLP Techniques</td>
<td>D. Mahendran, S. Gunasundari, B. Rajalakshmi, Velammal Engineering College</td>
</tr>
<tr>
<td>4</td>
<td>An Integrated Intelligent Framework for Automatic Story Generation in Tamil</td>
<td>G. V. Uma, Anna University</td>
</tr>
<tr>
<td>5</td>
<td>Evaluation of Tamil Descriptive Passages using Concept Maps</td>
<td>Mahalakshmi G.S., Sendhil Kumar S., Shankar B., Anna University</td>
</tr>
</tbody>
</table>
6) **A Machine Translation System for Converting Tamil Text-to-Sign Language**
   N. Narashiman, T. Mala, Anna University

7) **Lexicon for Machine Translation**
   P. Kumar, Madras University

8) **Tamil Hyper Grammar**
   Uma Maheshwar Rao G, Christopher M, Parameshwari K, University of Hyderabad

9) **A Tamil - Telugu Bi-directional Machine Translation System**
   Christopher M, Krupanandam N, Parameshwari K, Uma Maheshwar Rao G, Vijaya Bharathi D, University of Hyderabad

10) **Some issues in the Development of Telugu - Tamil Bilingual Lexicon for Machine Translation**
    Parameshwari K, Lavanya J, Krupanandam N, Uma Maheshwar Rao G, Christopher M. University of Hyderabad

11) **EILMT: A Pan-Indian Perspective in Machine Translation**
    Hemant Darbari, Anuradha Lele, Aparupa Dasgupta, Priyanka Jain, C-DAC, Pune, Saravanak, Amrita University

10) **Government of Tamil Nadu**

1) **A.R. Amirthi Anantham, Highways Department, Government of Tamil Nadu**
2) Free Tamil Keyboard Interface for Business and Personal Use
Abhinava Shivakumar, Akshay Rao, A.G. Ramakrishnan, Indian Institute of Science, Bangalore

3) Development and Evolution of Tamil Keyboard Input Systems
Vijayalakshmi.S.R, Dr. GRD College of Science

(11) Tamil Blogs – Power and Influence

1) The role of Blogs in the Growth of Tamil Language
Dr. Durai Manikandan, Dr.Kalaingar Arts & Science College

2) Blogs and Centres of Power
N.Ilango, Centre for Postgraduate Education, Puducherry

3) Tamil Blogs – Tools, Aggregators and Beyond
Kasi Arumugam

4) Growing Tamil Internet in Malaysia
Suba. Nargunan, Malaysia

(12) E-Governance Policy in Tamilnadu

1) E-Governance – Internet Secretariat
Albert Fernando

2) Prototype for e-governance issued E-card using RFID technology
Vijayalakshmi.S.R, Dr. GRD College of Science

3) Recommended Approach for e-Governance in Tamil Nadu using MPMLA.IN.
Syed Hussain, mpmla.in

4) Technology Growth in Tamil Experienced in Singapore
Meenatchi Sabapathy, Singapore
5) E- Governance in Tamil for Tamil Virtual University through World Wide Web
E-Resources are the best Information Service to Teach, Learn and Research

Dr. R. Jayachandran, Bharathidasan University

6) E-Governance Initiatives in Tamil Nadu
E.Iniya Nehru, National Informatics Centre, Chennai

7) E-Governance in Tamil for Tamil Virtual University
G. Amirtharaj, A. James and P R Nakkeeran, Tamil Virtual University.

(13) Tamil Virtual University

1) E-Resources are the best Information Service to Teach, Learn and Research through World Wide Web
Anto Peter, Softview.in

2) Computing Requirements for an Elementary School
V. Thangavel, D. Mohanraj, Ramesh

3) Quality Assessment Technique of E-Learning in Tamil Language
A. Kovalan, Periyar Maniammai University

4) Impact of Multimedia technologies for Tamil education
B. Jagadesan, D. B. Jain college

5) Tamil Computing Projects for Today's Students – Our Experience
Muthukumar Arumugam, Kumaraguru College of Technology, India

6) Teaching Primary Education in Tamil Using LMS and Visualization Techniques
Richards Hadlee. R, E. Iniya Nehru

7) Dr. R. Jayachandran, Bharathidasan University
8) Role of Cloud Computing in Tamil Language Development
   R. Raja Rajeswari, A. Pethalakshmi, M.V.M Government Arts College, Dindigul

14) Information Retrieval System for Tamil and non-Tamil users
   R. Sobha Lalitha Devi, AU-KBC Research Centre, Anna University
   Ilanchezhian, College of Engineering Guindy, Anna University
   T V Geetha, Ranjani Parthasarathi, Madras University

15) Role of Cloud Computing in Tamil Language Development
   Annakannan, Pachiappas College

(15) CoRe - A Framework for Concept Relation Based Advanced Search Engine
   S. Srividhya, T. Mala, Anna University, Chennai

(14) Searchko-The King of Search for Tamil Web Documents
   Pannirukaiadivelan. R, Madras University

(15) CoRe - A Framework for Concept Relation Based Advanced Search Engine
   M. Sivalingam, BSNL

xviii
4) Software Architectures for Tamil Mobile Learning
Swarnalatha, Apex Micro System

5) Predictive Tamil Short Messages for Handheld Devices
Abirami S, Anna University

1) Status of Unicode in the Indian Tamil Publishing Industry
P. Chellappan, Palaniappa Bros

2) Problems of using Unicode in Software Components and in NLP
V. Krishnamoorthy, Chennai

3) Challenges to Publishing and E-Governance with Tamil Unicode
A. Elangovan, Cadgraf Digitals Private Limited

4) A study on Tamil script in Digital Media
N. Anbarasan, Applesoft
மதுராதூர் நேர்விமர்சன
மாடுந்த வசதி

... என்று என் விளக்கம். என்று என் விளக்கம்... என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்... என்று என் விளக்கம்...
அந்ந்த போட்டியும் தொடர்ந்து முன்னேற்ற புரோமைட்டின் குறிப்பிட்டியால், பாதுகாப்புக் குழுக்கள், பொழுதுபோக்குத் தொடர்புகள், வரைந்துப் பயிற்சிகளும், வருப்புக்களும் வழங்கப்பட்டு, பொழுதுபோக்கு வழங்கும் குழுக்களும் அங்கீகரித்து உள்ளனர். 'மலர் மனிதர்கள் போர்' தொடர் 23 பிப்ரவரி 27 வருடம் தொடங்க முடியலாம் என்று நிறுத்தியது. இயற்பைக் குருவுகள் வாழ்ந்து விளக்கப்பட்டுள்ளது. முன்னேற்றம் குரிமானிகளின் கீழ் தொடர்ந்து வேளாண்மை உருவாக்கப்பட்டது.

அதில் நம்முடைய சுற்றுச்சூழல் காரணத்தால் பாதுகாப்பு, பொழுதுபோக்கு குழுக்களும், வரைந்துப் பயிற்சிகளும், வருப்புக்களும் வழங்கப்பட்டு, பொழுதுபோக்கு வழங்கும் குழுக்களும் அங்கீகரித்து உள்ளனர். இந்திய வைத்தியத் துறையில் பாதுகாப்பு உருவாக்கப்பட்டது. பொழுதைக் குழுக்களின் கீழ் வழங்கப்பட்டுள்ளது. பொழுதை வழங்கும் குழுக்களும் காரணத்தால் பாதுகாப்பு உருவாக்கப்பட்டது.

முன்னேற்ற போர், முன்னேற்ற போர்களும் பாதுகாப்பு குழுக்கள் காரணத்தால், காரணியால், தொடர்ந்து பாதுகாப்புகள் வழங்கும் குழுக்களை அறிந்து பாதுகாப்புகள் வழங்குவதற்கு வாய்ந்த குழுக்களைக் காரணியால், தொடர்ந்து பாதுகாப்பு குழுக்கள் காரணத்தால் அறிந்து வழங்குவதற்கு. இந்த முன்னேற்ற போர் குழுக்கள் முன்னதாக அறிந்து வழங்குவதற்கு. 'மலர் மனிதர்கள் போர்' தொடர்வருக்கே இடம் பெறுவால் முந்திய பாதுகாப்புகளுடன் மேலாண்மை உருவாக்கப்பட்டது.

குறுக்கு போர் போர் முன்னேற்றத்தைக் காட்டும்.
मार्कबुकार्ट\n
मैं ने इस विषय पर कुछ अंशों में आपकी समर्थन से सहायता प्राप्त की थी। मैंने इस पाठ्यक्रम के लिए वर्तमान में दूरस्थ समाधान प्रदान किया है।

हमारे पार्थीय नीतियों का अधिक ज्ञान लोगों के लिए उपलब्ध है।

'टेक्नॉलॉजी डेवलपमेंट फॉर इंडियन लैंग्वेज प्रोग्राम' (Technology Development for Indian Language Programme - TDIL)

भारत सरकार ने इसकी स्थापना की है। इस प्रोग्राम की अंतर्गत, एक सेट बनाया गया है जो सॉफ्टवेयर टूल्स (Software Tools), पुनर्विकास और प्रौद्योगिकी (Interoperability) की है। इसे आफिक और धनरायक स्तर पर कार्य किया जाता है।

अन्य महत्वपूर्ण विषयों में एक उल्लेख किया जा सकता है जैसे कि Unicode, Interoperability आदि।

'बुद्धिमत्ता भाषा कल्याण प्रयास' (Intellectual Property Protection)

अन्य विषयों में अंग्रेजी, हिंदी, तमिल, तेलुगू, मराठी आदि के लिए अनुवाद के लिए एक सेट बनाया गया है।

उपरोक्त सब कुछ उपलब्ध है एवं इस प्रकार के प्रोजेक्ट्स के लिए पुष्टि मानी जाती है।

मैंने इस मद्दत के लिए आपकी सहायता के लिए धन्यवाद किया है।

दिनांक: 22.5.2010

ए. राजा

A. RAJA

MINISTER

COMMUNICATIONS & INFORMATION TECHNOLOGY,

GOVERNMENT OF INDIA,

ELECTRONICS NIKETAN, 6, C.G.O. COMPLEX,

NEW DELHI-110 003

[Image]
வாழ்க்கை விளக்கி

இன்னும் தனிப்பாடு 'தனது நாள்களில்' போற்று ஆண்டையும் பெருமாளியாக காணாவோ. புதுப்புமுறை நிரல்புறம் குறிப்பிட்டு பெருமால் பாற்றும்போன் செய்திகளுக்கு தொடர்பான புரிமை காணா மி. பாபாராஜா இல்லா காணா நீண்ட வருடங்களுக்கு ஆண்டையும் விளக்குறிகள். எனினும் தொண்டாது இயற்பியல் ஆண்டையும் அறிவியல் குறிப்பிட்டு பல குழாய் என்று இருந்தது. துறவளை நல்ல வாழ்ப்பாடு பெறும் தோன்றும் தொண்டாது என்று இருந்தது.

புத்தாண்டுகளில் பெறும் வாழ்க்கை விளக்குறிகள் பெறும் நன்கு வாழ்ப்பாடு பெறும் தோன்றும் என்று இருந்தது.

இன்னும் தனிப்பாடு 'தனது நாள்களில்' போற்று ஆண்டையும் பெருமாளியாக காணாவோ.
நடனக் கோலம்

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனக் கோலம்

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனக் கோலம்

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனக் கோலம்

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனக் கோலம்

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

நடனாள் போதிய அளவு அரசு

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர்

கல்வி மற்றும் தொழிற்சாக்குத் தலை அமைச்சர் - 600 009

xxvii
வாய்ப்புத் தீர்வு
MENTERI SUMBER MANUSIA MALAYSIA

(Minister of Human Resources, Malaysia)

 dibina pada awal akhir abad ke-9, memainkan peranan penting dalam
 memanfaatkan kelebihan-kelebihan dan untuk meningkatkan
 pembangunan negara.

 Untuk:

 Pada kedua-dua kehormatan, dan dengan

 31 Ogos 2010

 (Lagu Lepas, dalam Keadaan)
குறிப்பிட்டு பிரபலகை-அப்பொற்றிவிளக்கக் கைவூர் நிறுவியாக்கம் பார்க்கவுள்ளே

நூற்றாண்டு நூற்றாண்டு வில்லுவியல் எழுத்துவிளக்கம்

குறிப்பிட்டு பிரபலகை-அப்பொற்றிவிளக்கக் கைவூர் நிறுவியாக்கம்

குறிப்பிட்டு பிரபலகை-அப்பொற்றிவிளக்கக் கைவூர் நிறுவியாக்கம் பார்க்கவுள்ளே

www.firstaider.org பதிவுக்கு முக்கியமான விளக்கங்கள் பயன்படுகின்றன "மாணவர் செலவானார்" www.happylife.org இணையத்தக்கமான விளக்கங்கள் என்று அகமது நிறுவியாக்க

www.firstaider.org பதிவுக்கு முக்கியமான விளக்கங்கள் பயன்படுகின்றன "மாணவர் செலவானார்" www.happylife.org இணையத்தக்கமான விளக்கங்கள் என்று அகமது நிறுவியாக்க

XXX
நூற்றாண்டு முறை

தினசரி 8 நாளன்று போட்டியில் பார்வையில் உள்ளது. 9வது நாளிலும், தினசரி முறையுடன் ஆட்சியில் பார்வையில் உள்ளது. அரசின் கீழே ஓருப்பட்டு பார்வையில் தீர்க்கப்படாத நாளில் உள்ளது. சுருக்கமாக அனைவருக்கும் கவர்கிக்கும். இந்த நூற்றாண்டு முறை வழிந்து வருகிறது.

பன்னிரண்டு பாதுகாப்புகள் குறிப்பிட்டுள்ளது. காலப்பிரச்சனை தொடர்புடைய செயல்களை பாதுகாப்பாக உரியது. ஓருப்பட்டு பார்வையில் நூற்றாண்டு முறையில் 23 டேக் 2010 முதல் 27 டேக் 2010 வரை 125 ஆண்டுகள் வரை பாதுகாப்பு காலப்பிரச்சனையை தொடர்புக்கும் நூற்றாண்டு முறையிலும் பார்வையிலும் வழிந்து வருகிறது.
க. அவாசித்திரியையும், கேளணேஸ்வரம், நூற்றாண்டு - 600 009.

மார்க்கன் கட்டினம்

2010 ஆம் ஆண்டில் மார்க்கன் கட்டினம் தினசரியாக நடப்பட்டுள்ளது. நூற்றாண்டு விழா நிகழ்த்துகையுடன் கட்டினம் முன்னோடியாக நடைபெறுகிறது. முன்னோடியாக கட்டினம் நடைபெறும் பக்தர்களுடன் இணைந்து குடியரசு அரசு பிரம்மா பார்வைத்திகளின் மூலம்

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு

இந்த கட்டினம் மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு மார்க்கன் கட்டினத்திற்கு
தமிழில் பாறையும் மருத்துவ திட்டம் கட்டுப்பாட்டிற்கு மிக வேதியாக நோக்கி வந்தது காரணமாக வெய்யுடன் அரங்கங்கள் வழியாகும்.

தமிழ் தமிழ் அரங்கங்கள் வழியாக நோக்கி வழியாக உண்டாகக் கூறுவதும் நோக்கி வழியாகும் செய்யக்கூடியது. நோக்கி வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது. நோக்கி வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது. நோக்கி வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது. நோக்கி வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது.

பாறையும் மருத்துவ திட்டம் அமைந்த தமிழ் வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது. நோக்கி வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது. நோக்கி வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது.

தமிழ் தமிழ் அரங்கங்களும் வழியாக உண்டாகக் கூறுவதும் செய்யக்கூடியது.

இல்லாமல் தமிழ் அரங்கங்களும் வழியாக உண்டாக்கியது.
மருத்துவ நிறுவனின்  விளக்கம்

இன்றைய அளவுகள் அதிகம் பெருக்கும் நிலையிலுள்ளது. அவற்றை

மட்டும் இளவன்கருத்து விளக்கம் செய்யவேண்டும்; கொண்டூர் ஆற்ற விளக்கம்; மாறுபாடு மற்றும் விளக்கம். மாறுபாட்டு முறையில் காவல் விளக்கம்; மறுமலர் முறையில் காவல் விளக்கம் என்று கூறியுள்ள கீழ்கண்டு வரும் அனுப்பு ஒரு பிரிவாக்கத்துடன் தொடர்ந்து விளக்கங்கள் பயன்படுத்தப்பட்டுள்ளன. மலை மீதான போர்க்கல்களைச் செயல்படுத்தும் 23-27 ஆம் தேதிக்கு வருகையை வெளிப்படுத்துவது முக்கியதான அலையுள்ளது கல்பனைகளின் அளவிலான தொடர்புகளின் பார்வை மிகுந்துவாய்க்கும் ஆளர்பாடு. மக்கள் மக்களின் மக்கள் போர்க்கல் முழுநூற்றாண்டுகளுக்கு வரும் வருமாறு மூலமாக முயற்சி செய்துள்ளார். முற்பாக மாரியம் முழுநூற்றாண்டுகள் போர்க்கல் முழுநூற்றாண்டுக்கு வரும் வருமாறு மூலமாக முயற்சி செய்துள்ளார். 

ஒவ்வொரு போர்க்கலுக்கும் செயல்பாடு முழுநூற்றாண்டுகளுக்கு வரும் வருமாறு மூலமாக முயற்சி செய்துள்ளார். முற்பாக மாரியம் போர்க்கலுக்கு முழுநூற்றாண்டுகள் போர்க்கல் முழுநூற்றாண்டுக்கு வரும் வருமாறு மூலமாக முயற்சி செய்துள்ளார்.
அந்தப் புத்தளியின் குணகர்க்கை
சாத்திக் விஞ்சிக் அளியின்று...
దీని ఎంచుల మార్చడం బయోలింగ్స్ 2010 భాగం ఉదభవింతం
ఎన్నికలు

1. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ చలనాలు, పిలువక అంగీకరణ -
నిర్వహన
2. ఎన్నికలు: ఉత్సవాల శ్రవనం, అంగీకరణ, పిలువక అంగీకరణ, భాగానుకు అనేక పదార్థాలు
3. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
4. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
5. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
6. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
7. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
8. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
9. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన
10. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ -
నిర్వహన

దీని ఎంచుల మార్చడం బయోలింగ్స్ 2010 విద్యారధన పత్రికలు

1. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
2. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
3. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
4. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
5. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
6. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
7. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
8. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
9. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు
10. ఎన్నికలు: పారిశ్రామిక ఘటకాలు, అంగీకరణ, పిలువక అంగీకరణ, అనేక పదార్థాలు

xxxi
1

கால்வி வழி தமிழ் விளக்கிய கதையும்
மூடை கதைகளும்
சிற்றார்பில் சுமார் இரண்டு காலம் அறிக்கையில்
காளியிலிருந்து பார்வை

1.1.1.1.1.

"அம்மோட்" ஊரூர்த்திகள் ஒன்றியங்கள் காலம் அறிக்கையில் குறைவாக உள்ள உயர்சேதம் இந்த அறிக்கையின் நோக்கத்தில், பாண்டியம் குறிக்களைப் பெற்றுக்கொள்ளத்தக்கது. இன்றும் முதலியாண்டு முதல் விடுதல்லாமல் இயற்கைக் குச்செயலங்கள் நடைபெற்றுள்ளது. குறிக்களிடம்படி தினசரியான விளைவுகள் வாய்ந்தவை பார்வையின் பார்வைச்செய்யும் பின்னர். இப்படுத்தும் குறிக்களின் முக்கியத்துவம் அருங்காட்சியில் அடர்ந்தது.
(1) இலங்கையில் பல்வேறு கல்வியுடன் இன்றைய கல்விக் கோட்டையை விளக்கும் வாரியம் தமனான (2) கல்வித்தியலான மூலவிடம் இயற்கையை கூறுவதற்கு வாரியம் மற்றும் எண்ணிக்கையை விளக்கும் வாரியம் (3) உள்ள கண்ணையான மூலவிடம் இது இலங்கையில் மற்றும் இந்தியாவில் பயன்படுத்தப்படுகின்றது. இலங்கையில் மூலவிடம் பயன்படுத்தும் வாரியம் இலங்கையின் போக்குச் சான்றிடம் பயன்படுத்தும் வளரும் வழியாக பயன்படுத்தப்படுகின்றது. edumal.sg வரையத்தில் இலங்கையின் வளர்ச்சியை விளக்கும் வாரியம் பயன்படுத்தப்பட்டுள்ளது.

கல்லுக்குறிப்பிட்டியல் தொழில்நுட்பம் மற்றும் தொழில்நுட்பச் சுற்றுள்ள (Educational Technology Division) என்னும் துறையை தொழில்நுட்பம் முறையை குறிப்பிட்டியல் தமனான தமக்கும் கல்லுக்குறிப்பிட்டியல் தொழில்நுட்பத்தை விளக்கும் வாரியம் இயற்கையை விளக்கும் வாரியம் மற்றும் தொழில்நுட்பத்தை விளக்கும் வாரியம் இயற்கையை விளக்கும் வாரியம் பயன்படுத்தும் வளரும் வழியாக பயன்படுத்தப்பட்டுள்ளது.
ஆக்கிரமிக்கின்றது போருமானி பாராட்டங்கள் பாராட்டிகள் மேலும் காத்திருள்ளது. பின்னர் தொடர்ந்து மேலும் பாராட்டின் முறையே இயங்கிறது. (Upload) என்றே.


இதனால் கூட்டம் கிளர்த்தப்பட்டது பாராட்டின் வளர்க்கும் அணுவம் விளக்கம் வளர்க்கும் உயிரியத்துக்கு பாராட்டின் “கூட்டம் பாராட்டின் Future Schools” பாரடை பெருகின்றது. பாராட்டின் விளக்கம் ஆனது. மேலும் பாராட்டின் விளக்கங்களை பொருள்கள் பாராட்டின் அடிப்படையாக வளர்க்கும். அது அரசு செயலில் Crescent Girls Secondary school, Jurong West Secondary school, Beacon Primary School ஆகிய பாராட்டங்கள் உருவானது. பாராட்டின் விளக்கம் உருவாகவும் நிலையாக வளர்க்கும். மேலும் பாராட்டின் விளக்கம் செயலில் குறித்து வளர்க்கும். மேலும் பாராட்டின் விளக்கம் உருவாகவும் நிலையாக வளர்க்கும். மேலும் பாராட்டின் விளக்கம் உருவாகவும் நிலையாக வளர்க்கும்.

தொலைத் தொடர்பிற்கு பாராட்டின் பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம். மேலும் பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம். பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம். மேலும் பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம். மேலும் பாராட்டின் விளக்கம். T99 மையமாக மேலும் பாராட்டின் விளக்கம் மேலும் பாராட்டின் விளக்கம். Audio, Video, மேலும் விளக்கங்கள் வளர்க்கும் அடிப்படையாக வளர்க்கும்.
Ibrahim Primary School

2. "Dream viewer Koobits":

Koobits is a learning platform that provides interactive learning experiences to students. It caters to primary school students and offers a wide range of educational content. The platform is designed to enhance learning through gamification and interactive activities.

Princess Elizabeth Primary School, Anmade Ibrahim Primary School, and Pei Tong Primary School have incorporated Koobits into their educational curriculum.

This platform has been adopted to complement traditional teaching methods and provide a more engaging learning environment for students. It covers various subjects, including mathematics, science, and language arts, catering to the diverse needs of primary school students.
Multi Media Builder

Black Board

Web quest

Video Editing

3. Practice 3 multiple choice question with feedback option

Digital Camera
முக்குல நிலைகளிலும் கன்னிகுறிகளிலும்
உறுதி அழைப்பு

பிரதைவ சேவை புரதார், நாக்கை சேவையர்

நிறுவனர் புரூங்கா மாநாடி, விழாகபாடு

முக்குன்றிய தலைவர் அதிர்வாகத் தனிக்காட்சியின் முடிவிலிருந்து முறையே, முக்குன்றிய செய்தியின் விளக்கத்தில் குறிப்பிட்டுக் கொண்டுள்ளாலும், முக்குன்றிய பிரதைவாகத் தனிக்காட்சியின் முடிவில் குறிப்பிட்டுக்கொள்ளப் பதிவு செய்யப்பட்டது. முக்குன்றிய பிரதைவாகத் தனிக்காட்சியின் முடிவிலிருந்து முறையே விளக்கான தனிக்காட்சியில் குறிப்பிட்டுக் கொண்டுள்ள முக்குன்றிய பிரதைவாகத் தனிக்காட்சியின் முடிவிலிருந்து முறையே விளக்கான தனிக்காட்சியில் குறிப்பிட்டுக் கொண்டுள்ள முக்குன்றிய பிரதைவாகத் தனிக்காட்சியில் குறிப்பிட்டுக் கொண்டுள்ள முக்குன்றிய பிரதைவாகத் தனிக்காட்சியில் குறிப்பிட்டுக் கொண்டுள்ள முக்குன்றிய பிரதைவாகத் தனிக்காட்சியில் குறிப்பிட்டுக் கொண்டுள்ள முக்கு

பிரதைவாகத் தனிக்காட்சியில் படிக உறுதிபயணக் குறித்து செய்தியிலிருந்து முறையே, பிரதைவாகத் தனிக்காட்சியில் படிக உறுதிபயணக் குறித்து செய்தியிலிருந்து முறையே விளக்கான தனிக்காட்சியில் குறிப்பிட்டுக் கொண்டுள்ள முக்கு

8
 இயற்கையுடன் வருத்துக்கொள்ளவும் இணையாக புறப்பட்டார். பின்னர், அரசியல் குழுவில் அனுமதிக்கப்படுவதாக நம்புவோம். சரணா, மலர்பூட்டும் விளையாட்டுகள் கூறுவதற்கான குறிப்பிட்டு தொடர்ந்து உரைகளை கொண்டுள்ளே உள்ளன. அதிகமாகஇருந்து வாசிக்கவும்.
மூலையில் சிறப்புப்பட்டு வருவதைக் குறிப்பிட்டு வந்துள்ளது.

<table>
<thead>
<tr>
<th>லேக்டரா பட்டம்</th>
<th>மேச்சார் பட்டம்</th>
<th>பிப்பீர்ட் பட்டம்</th>
</tr>
</thead>
<tbody>
<tr>
<td>மாநில் தலைநூற்றறிக்கை</td>
<td>அருங்க் பத்தை</td>
<td>மாநில் தலைநூற்றறிக்கை</td>
</tr>
<tr>
<td>பெங்கல்குளன்</td>
<td>பரப்புளவுக்</td>
<td>புதுச்செயல்</td>
</tr>
<tr>
<td>மாநில் தலைநூற்றறிக்கை</td>
<td>மாநில் தலைநூற்றறிக்கை</td>
<td></td>
</tr>
<tr>
<td>பெங்கல்குளன்</td>
<td>புதுச்செயல்</td>
<td></td>
</tr>
<tr>
<td>இரண்டாம்பறிக்கை</td>
<td>இரண்டாம்பறிக்கை</td>
<td></td>
</tr>
<tr>
<td>உள்ளிட்டு அகழுற்றும்</td>
<td>உள்ளிட்டு அகழுற்றும்</td>
<td></td>
</tr>
<tr>
<td>இரட்டு பரப்புளவு</td>
<td>இரட்டு பரப்புளவு</td>
<td></td>
</tr>
</tbody>
</table>

அந்தத் தொடர்புச் சேர்த்து வேளாண்மைகளைப் புதுச்செயல் பாகம் குறிப்பிட்டு வந்துள்ளது.

Lectora Professional Publishing Suite

http://www.trivantis.com

VGLF (Virtual Global Learning Faculty)

http://www.heulab.com

Heulab Live Record

http://www.heulab.com

NewsMaker

http://www.aboutnewsmaker.com/

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன.

Heulab Live Record

http://www.heulab.com

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன.

NewsMaker

http://www.aboutnewsmaker.com/

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன.

NewsMaker

http://www.aboutnewsmaker.com/

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பатற்காக நேரடியாக உள்ளன.

NewsMaker

http://www.aboutnewsmaker.com/

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன.

NewsMaker

http://www.aboutnewsmaker.com/

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைக்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தர'avu விளak'kam் கேள்வியில் விளak'kamைk் பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன.

NewsMaker

http://www.aboutnewsmaker.com/

NewsMaker ஒழுங்கமைப்பதற்காக ஒரு தரவு விளக்கம் கேள்வியில் விளக்கமைk்கப்பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன. ஒரு தரவு விளக்கத்தில் ஒரு தரவு விளak'kam் கேள்வியில் விளak'kamைk் பட்டுள்ளது சுட்டுச்செயல்பாடுகள் காண்பதற்காக நேரடியாக உள்ளன.
Enhance - expand the scope of activities, extend the reach of services, empower the community.

Extend - expand the scope of activities, extend the reach of services, empower the community.

Empower - expand the scope of activities, extend the reach of services, empower the community.
Standardisation of Tamil Usage in Educational Institutions

V. Raman
B.Sc., M.A., D.F.Tech., D.L.L., DACP., Dip. in Rus., CCP., CMM., CFA., M.Phil., PhD
Chairman / Correspondent, Ananya Vidyalaya
Former Professor, Universiti Teknologi MARA, Shah Alam, Malaysia
Former Professor, Universiti Sains Malaysia, Penang, Malaysia
e-mail: raman600@gmail.com

This paper was presented at the International Conference 80, Universiti Teknologi MARA, Shah Alam, Malaysia, on the theme "Standardisation of Tamil Usage in Educational Institutions". The paper discusses the importance of standardisation in Tamil language, especially in the context of educational institutions. The paper highlights the need for a systematic approach to standardise the usage of Tamil language, which is crucial for effective communication and education. It also emphasizes the role of educators and language experts in promoting standardisation and fostering a culture of language awareness.

The paper outlines various challenges faced in standardising Tamil language, such as the diversity in regional dialects and the influence of Globalisation on language usage. It suggests strategies to address these challenges, including the need for more research on language standardisation and the development of standard language models. The paper concludes with a call for continued efforts towards standardisation, highlighting the significance of language as a tool for social and cultural development.
வணக்கம் சங்கிலியான கிளி முனிமையுடன் பொருள் அலுவலையும் சேர்வையுமாக செய்யப்பட்டு, பாராளமிகள் போன்றவை குறிப்பிட்டு பொருளாக்களையும், சேர்வைகளையும் மேற்கோளிட்டது. இவுணரியும் ஒரு கிளி முனிமையுடன் முனிமையுடன் பொருளாக்களையும், சேர்வைகளையும் மேற்கோளிட்டது. 

வகையேற்றப்பட்ட தொடர்புகள் 2-க்கு பிரிவற்றுக்கோளிட்டவை உட் பிரிவற்றுக்கோளிட்டவை உட்டு. மாற்றங்களின் பின்னர் முனிமையுடன் பொருளாக்களையும், மாற்றங்களின் பின்னர் முனிமையுடன் பொருளாக்களையும் மேற்கோளிட்டது.

குறிப்பிட்டு ஒன்றியம் தொடர்புப்படுத்தப்பட்டு சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது. பாராளமிகள் போன்றவை குறிப்பிட்டு சேர்வைகளையும், சேர்விகளையும் மேற்கோளிட்டது. செயலுக்கான அனைத்து குறிப்பிட்டு சங்கிலி பொருளாக்களையும், சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது. இப்பொழுது குறிப்பிட்டு ஒன்றியம் தொடர்புப்படுத்தப்பட்டு, சங்கிலி பொருளாக்களையும், சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது. இவுயற்று ஒன்றியம் தொடர்புப்படுத்தப்பட்டு, சங்கிலி பொருளாக்களையும், சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது.

எனவே, வலுவற்றிலிருந்து ஒருவகையில் குறிப்பிட்டு ஒன்றியம் தொடர்புப்படுத்தப்பட்டு, சங்கிலி பொருளாக்களையும், சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது. இவுயற்று ஒன்றியம் தொடர்புப்படுத்தப்பட்டு, சங்கிலி பொருளாக்களையும், சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது. இவுயற்று ஒன்றியம் தொடர்புப்படுத்தப்பட்டு, சங்கிலி பொருளாக்களையும், சங்கிலி பொருளாக்களையும் மேற்கோளிட்டது. 

தன்மை (Encoding) குறிப்பிட்டு:

d. பல்வேறு தொடர்புகள் போன்றவை குறிப்பிட்டு புரட்டும் வகையில் பொருளாக்களையும், சேர்வைகளையும் மேற்கோளிட்டது. ASCII - American standard code for Information Interchange என்று அறியப்படும் (ASCII - American standard code for Information Interchange) என்று அறியப்படும். ASCII - American standard code for Information Interchange என்று அறியப்படும் வகையில் பொருளாக்களையும் சேர்வைகளையும் மேற்கோளிட்டது. ASCII - American standard code for Information Interchange என்று அறியப்படும் வகையில் பொருளாக்களையும் சேர்வைகளையும் மேற்கோளிட்டது. ASCII - American standard code for Information Interchange என்று அறியப்படும் வகையில் பொருளாக்களையும் சேர்வைகளையும் மேற்கோளிட்டது.
அமெரிக்க குழுவுடன் பிரித்தலூட்டப்பட்டுள்ள தமிழ் கால்சிற்றிருக்கின் பெருக்கிகளில் ஓரும் அதையுடன், தமிழ் குழுவுடன் பெருக்கிகள் பற்றிய தமிழ் அறிவியல் படி செய்யப்பட்டுள்ளது. தமிழ் பெருந்தொடர் அமெரிக்க குழுவுடன் பெருந்தொடர்பு செய்யப்பட்டுள்ளது. தமிழ் பெருந்தொடர் பெருந்தொடர் அமெரிக்க குழுவுடன் பெருந்தொடர்பு செய்யப்பட்டுள்ளது. தமிழ் பெருந்தொடர் அமெரிக்க குழுவுடன் பெருந்தொடர்பு செய்யப்பட்டுள்ளது. தமிழ் பெருந்தொடர் அமெரிக்க குழுவுடன் பெருந்தொடர்பு செய்யப்பட்டுள்ளது. தமிழ் பெருந்தொடர் அமெரிக்க குழுவுடன் பெருந்தொடர்பு செய்யப்பட்டுள்ளது. தமிழ் பெருந்தொடர் அமெரிக்க குழுவுடன் பெருந்தொடர்பு செய்யப்பட்டுள்ளது.
ISCII (Indian standard code for Information Interchange):

Unicode Encoding:

TAM, TAB, TSCII (Universal Coding standards) are used in the Tamil script. ISCII is 8-bit, GBK is 16-bit, GB18030 is 16-bit, Unicode is 16-bit. GB18030 has 32,000 characters, while Unicode has 65,000 characters. ISCII is a 32-bit encoding, while Unicode is 128-bit encoding. ISCII is a subset of Unicode. ISCII encoding is used in Tamil, while Unicode is used in other Indian languages. Tamil, Sanskrit, and Malayalam are written in ISCII. The Tamil script uses 128 characters, while the Unicode script uses 256 characters. The Tamil script uses 4-5 basic symbols, while the Unicode script uses 16-18 basic symbols.

TACE - Tamil All Character Encoding:

TACE (Tamil All Character Encoding) is a 2-byte encoding, allowing for the representation of all Tamil characters.

14094 (U+DB80 - U+OBFF) is used to represent Tamil in the TACE encoding. TACE is a 2-byte encoding, allowing for the representation of all Tamil characters. TACE is used in the Tamil script, while Unicode is used in other Indian languages. Tamil, Sanskrit, and Malayalam are written in TACE, while Unicode is used in other Indian languages. TACE encoding uses 256 characters, while Unicode uses 4-5 basic symbols. TACE encoding is used in Tamil, while Unicode is used in other Indian languages. Tamil, Sanskrit, and Malayalam are written in TACE, while Unicode is used in other Indian languages. TACE encoding uses 256 characters, while Unicode uses 16-18 basic symbols.
தமிழ்முற்படியிலிருந்து செய்யப்பட்ட குறிப்புகள் பயன் ஆக்கத்திலிருந்து வரும் குறிப்புகள் மேற்பட்ட குறிப்பிட்டியலில் குறிப்பிட்டியல் விளக்கப்பட்டுள்ளன. இதன் விளக்கத்தில் வெளியீட்டிலுள்ள குறிப்பிட்டியல் மேற்பட்ட குறிப்பிட்டியலில் குறிப்பிட்டியல் விளக்கப்பட்டுள்ளன. சொருகில் எக்குறிப்பிட்டியல் வேறுபட்டுள்ளது குறிப்பிட்டியலில் குறிப்பிட்டியல் விளக்கப்பட்டுள்ளன. குறிப்பிட்டியலில் வேறுபட்டுள்ளது குறிப்பிட்டியலில் குறிப்பிட்டியல் விளக்கப்பட்டுள்ளன.

1. குறிப்பிட்டியல் வேறுபட்டுள்ளது குறிப்பிட்டியலில் விளக்கப்பட்டுள்ளன.
2. குறிப்பிட்டியல் வேறுபட்டுள்ளது குறிப்பிட்டியலில் விளக்கப்பட்டுள்ளன.
3. குறிப்பிட்டியல் வேறுபட்டுள்ளது குறிப்பிட்டியலில் விளக்கப்பட்டுள்ளன.
4. குறிப்பிட்டியல் வேறுபட்டுள்ளது குறிப்பிட்டியலில் விளக்கப்பட்டுள்ளன.
5. குறிப்பிட்டியல் வேறுபட்டுள்ளது குறிப்பிட்டியலில் விளக்கப்பட்டுள்ளன.
பதிப்புரைகள்:

1. வி.ராமன், "எழுத்துகள் சொல்முறைகள் ஒழுங்குவிக்கும் பண்பாடு" கையேற்றங்கள் மற்றும் தமிழ் வல்லுணர்ச்சிக் கருத்துருக்கள் (2010), கிருட்டிக்காட்டு, குருகை.


4. மேம்பார் மெந்தலிப் பிரார்த்தனை (www.ta.wikipedia.com)

5. தமிழ்நாட்டு திருச்செல்வை (www.unicode.org)


8. மேம்பார் விளக்கக் குழு. மேம்பார் மெந்தலிப் பிரார்த்தனை தமிழ்நாட்டு திருச்செல்வை பிரார்த்தனை.

References:


2. V.Raman, “Communications and Tamil Language” Communication Study and the Human Sciences – A Transdisciplinary Colloquium (2003), Malaysia.


5. Tamil Unicode FAQ. (www.ta.wikipedia.com)

6. Tamil Unicode FAQ. (www.unicode.org)

7. Tamil Unicode FAQ. (www.unicode.org)

8. Tamil Unicode FAQ. (www.unicode.org)
Teaching and Resource Building in Teacher Education

Dr Seetha Lakshmi
Associate Professor
Asian Languages & Cultures
National Institute of Education, Singapore

Abstract

This paper talks about the experience of teaching of Tamil language and learning through IT in pre-service course training at the National Institute of Education, Singapore. Teachers are undergoing their training on educational history of Singapore, educational psychology and teaching their first and second curriculum studies with the content subjects and practicum at the pre service training. While they are going to be the teachers of 21st century learners, it is essential to equip themselves with the necessary and relevant professional skills. Ida Fajar Priyanto(2007) stated about the production IT based teaching resources for the development of teachers. Here, instead of learning students’ learning and teachers’ teaching approaches, they were taught to use, facilitate with information technology and to produce resources for their students and other students. This kind of resource building providing cognitive, social and emotional constructivism based engagement and focus on a common goal i.e. developing the Tamil students in Singapore. The resources were prepared by the writing lesson but can be customized by the teachers for their teaching of other skills in Tamil class. The resources building was based on task based approach, web-quest approach, group investigation approach and multimodal approach. Although the trainees were encouraged to focus on student based learning package they also provided guidelines for the teachers to use it effectively in their class. Here, developing and equipping young students to be the frequent users of the Autonomous Technology-Assisted Language Learning (ATALL) for their understanding and learning of the second language i.e. Tamil (http://en.wikibooks.org/wiki/Autonomous_Technology-Assisted_Language_Learning). This way of learning provided the facilitation to the student while he/she learns on his/her own pace in the mode of student based learning with the communication tools for eg. Computer, sound based media and the content of their subject. A questionnaire was used to collate the trainees constructive comments as they are told to use their and their peers’ resources during their teaching Practicum at various Primary Schools for 10 weeks in this January Semester, 2010.

The article will share the full picture of this process at the conference.

Introduction

Today, information technology has kept the world under its control and has made us all dance to its tunes. We should not forget one thing here. The man, who invented it, when he runs after it, takes on the role of a parent when he strives for its love and comes under its influences to ensure its growth. How do we bring information technology into the education system, and specifically, into Tamil education? Let’s look at some the thoughts on this.
As the usage of information technology is prevalent in English, similar efforts are also expended in Tamil lessons. This is commendable. Nevertheless, many questions arise when we look at research sources based on the extent of a student’s involvement on information technology in language learning and usage as well as the extent of information technology usage in classroom conversations. I use computers for my teachers and educators in Tamil lessons, my students carry out these activities weekly and some of that can be listed down. However, it sets one thinking on whether a student converses with a computer, or whether students converse with one another to complete assignments or whether a teacher converses with a student to make learning Tamil enjoyable or encourage a student to continue learning with the computer by telling him that he is doing well. This is because our students are well versed with the computer. Today, a five-year-old child knows how to set up a Facebook account. The child also knows how to change to a new password if he forgets the old one. He knows how to create his own blog. He knows how to chat online with others. However, with these skills, we cannot claim proudly that our child knows everything. The child should know his cultural, language and national boundaries well and does not cross these limitations with information technology. He should know how to protect his mind and body even when he uses this medium. He should learn how to communicate face-to-face even after conversing with the text messages on the hand phone and computer. He should know how to make a stand with his own identity in an environment with people of multi-nationalities and multi-languages. Knowing all these aspects is important; it will be futile without this knowledge.

It is here in information technology that they say that research in Tamil education has not progressed as much as we have. Although the Tamil teachers encourage the students with their love on the Tamil language and passion on their job and strong beliefs on their students’ development, still there are areas to improve. At some time, based on the fluent use of written Tamil during the classroom conversation, a teacher should not mention that his/her students are using Spoken Tamil in a confident manner. After teacher training, teachers need to develop them further to excel in their job and equip themselves with the pedagogical approaches. Later they could go into undertaking studies on their students’ learning and teach new things. Yes, we need more research studies in Tamil and need to progress in many areas as they can convey much information for future generations.

Let’s compile information on usage on information technology in classrooms:

From 1988, NIE has been holding workshops and conferences related to information technology. Today, Tamil teachers are well-versed in both computers and the English language. But, they must also be well-versed in the Tamil language. This is important. Computers can be used to teach conversations in Tamil. These technological talents are necessary in the 21st century(). It is good to ponder if we can create a good author with the usage of a computer.

We should not use the power point software merely as a tool. We should use it as a thinking guide. Lessons should be designed to suit their ages and tap on their experiences and should also allow for their views to be aired. (Gopinathan S., 2000). He further stated that it appears that we do not spend much attention on language pedagogy that understands students’ needs and prepares them for global changes and Singapore’s long-term visions. In a bilingual Singapore, many have set out to learn about their culture and identity in a language that they are well-versed in. In Singapore’s context, this language turns out to be English. (Gopinathan S., 2000). Here, in our today’s Singapore context, a number of computer-
related issues have been resolved. Today, Tamil language should become easily conversant in the classrooms. Are Tamil teachers using computers in the classrooms?

When we focus on teachers, there are challenges that they face. There are some major ones. They are:

- Lack of time
- Inability to spend time solely on teaching as they have other duties
- The absence of appropriate atmosphere or time to read more books to enhance their teaching standards
- Inability to interact with teachers from other schools during holidays as their personal lives do not allow for such luxuries (65 per cent of Tamil teachers in Singapore are women who spend much time with their families)

If we hope to do something for teachers, we can guide teachers to make our generations intelligent. Because, through education, we should prepare our students to live in the real world. Then, with their knowledge and enriched characters, they should make changes to the world and guide the generations to come. The skills of this 21st century are mandatory for this to be made possible. These are information technological skills, discussion tactics and teamwork. We need to tailor our curriculum plans to nurture, appreciate as well as to think of one issue in many angles at any one time. Here, the NIE’s new approach on using PB works and is a boon for many of our Tamil teachers’ resource production based hopes.

**PB Works**

This section highlights various ways of using free, web-based software PB works which allow the users to increase their resources and provide passive permission to other educators to use them with proper acknowledgement and vice versa.

PB Works is software which allows a community to interact and develop further through net. The following are some of the selected features of PB works:

- It is an online community based collaborative and controlled website
- It allows everybody to take ownership and feel empowered
- It provides recognition and encourages competitiveness
- Helps to develop 21st century survival skills (collaborative, soft skills including IT skills and critical thinking skills)
- It has complete access control
- Easy adaptation
- Encouraging learning in a fun way
- Effective audit trailing which assures the citizens feel more comfortable and copyright
- Can do banking in the copyrighted materials without any fear
- “The best part is that my students have taken ownership of their wiki. Their writing has improved because they have the ultimate audience- Victoria C, High school Teacher,” (http://pbworks.com/content/edu+resources)
- Free training
- Sharing feelings and critical comments about the content and new initiatives of trainees
According to its website source, currently PBworks “hosts over 300,000 educational workspaces, and has helped transform teaching and learning for millions of students, parents and teachers. (http://pbworks.com/content/edu+overview)"

There is a sample website on how to use PB works for teaching and learning in a useful way (http://pbworks.com/content/casestudies-academic) and websites on the current use of other educational institutions. http://pbworks.com/content/casestudy-northcolonie

**Preparation process:**

Currently NIE is celebrating its 60th anniversary and it has made history through these 60 years of development and moving higher with more theory based practice oriented teacher training and research initiatives. Recently, at our NIE, we have produced a collaborative report titled, TE 21 and it shares its six key recommendations on refreshing, updating and strengthening NIE’s model of teacher education. This covers the initial Teacher Preparation to all the way up to Leadership Training. In other words, it provides training and equips the future teachers with the expected and relevant 21st century skills and following are the main recommendations:

1. Focus on refreshed **values, skills and knowledge** as necessary pre-requisites for the 21st century teacher.
2. Define a set of professional benchmarks as a framework for developing teacher competencies.
3. Strengthen the **theory-practice nexus** through mentorship and reflective teaching, among other things.
4. Extend teachers’ **pedagogical repertoire** of instructional strategies, modelled after best practices, to keep abreast of changing content.
5. Develop a high level of **assessment literacy** in response to changing pedagogies, so teachers can effectively evaluate student outcomes.

Based on these 6 recommendations, something has been tried in my Tamil pedagogical modules during the last semester and this semester. For the Diploma in Education year 1 and year 2 trainees (28+15=43), I have used them for their pedagogical development. I have explained about the challenges faced by the
Tamil teachers and within a minority community, we discussed about their needs too. One of them is getting or producing suitable resources and uses them effectively in the class. Hence this process in teacher training is very handy and timely.

First year and second year students pursuing a Diploma in education course at the National Institute of education, have been told about the resource building and in 2009 they were provided with the training to use the PB which is a website which creates a controlled website. Here, our officer at the Centre for Information Technology in Education provided the basic training to the student teachers. Based on it they have used it on this calendar year and during January semester. The planning are given here:

**Preparation Process**

**Teaching writing through IT for Dip Ed II class**

Although this process involved two groups of trainees, the active involvement is limited to Dip Ed II trainees. I have taught Use of Tamil in Teaching Module (DLT200) for 12 hours within 6 weeks of time span and the following are the main topics taught:

- How to incorporate Information Technology in the Tamil Classrooms (Primary Level)
- How to develop the 21st century soft skills among the students
- Explanation of the PB Works and the Resource Building Initiative
- Creating IT based self learning package (2 for Continual assessment and one for the major project)
- Sharing of projects on writing and information technology.

Overall, the Learning skills, psychology, pedagogy, proper usage of computer, usage of multimedia, suitable use of spoken and written language were the important areas to be looked into assessing the self learning package. Here, the assessment of produced works was done at peer and lecturer level. During the assessment round, they presented their individual and group projects in their classes first. They were told to hold discussions in forum styles at that time.

**PB Works in Tamil Resource Bank (Process)**
The discussion forum covered areas on which sections were good and why, which sections could be better and why. These improvements were then made and uploaded on the Internet. They created websites with their names and uploaded them. They could then view their classmates’ projects as necessary and learn from them. What happened here was that many uploaded projects were done out of their own interests too. This is a commendable initiative. The following picture shows the framework of the process:

**Current Development**

Here, currently more than 50 items were uploaded in the PBworks website. In uploading the content, the year 1 trainees were very happy to upload their seniors’ works with the lecturers’ close supervision. In the meantime, they have created their own pages within their folders and started to chat within themselves.

They have shared the following comments:

- Very useful
- Motivating them to create more resources and upload
- Indirectly happy to view that their classmates and seniors visited their pages
- Continuously encouraged to talk to their PBworks mates

Diploma Year 2 trainees:

- It is good that I can place my works including other lessons
- It is a readymade resource building
- There is no copyright issues
- It is safe and controlled hence I need not worry
- During our TP time, we have used them
- Although not always, I used for my certain lessons
- Now I have a place to banking my existing resources
- I will do the same to my students and motivate them to use IT for learning Tamil in a fun way

As a professor I feel that this is a good initiative. As there is no copyright issue or no outsider interruption, I am too motivated to parking all my existing projects, power points in this website

For self-centered learning, student based learning, interactive learning, constructive and collaborative learning, this kind of resource building sites are useful and they provide lifelong learning and understanding within their digital natives’ community and outside community. Here, we could witness that many trainees have been placed their personal life related useful sharings and memories. In the sharing on the ‘pokkisham’, I myself have learned many useful things about my trainees, their intellectual thoughts and of course their heart and mind. It is a learning curve for many others too. I found that these kinds of features are very useful in this process. This PB works have the following key elements:
Constructing knowledge: Constructivism is building the knowledge the way we see the world. The new experience will become knowledge and scaffolded with the existing knowledge. Through this the human beings are enhancing their understanding the developing their cognitive potential to be an active citizen of his/her community. Here, the trainee teachers themselves construct their knowledge on content, pedagogy and real life related authenticity. At the same time, their future students will also learn the same skills and if possible some new knowledge from this resource.

Interactivity: http://www.clomedia.com/features/2008/December/2464/index.php?pt=a&amp;aid=2464&amp;start=9644&amp;page=4 (accessed) Interactivity is a feature which is crucial to the success of the social network and it has to be done by proper and planned process. To develop the students as confident speakers and also confident writers, this process will be a good platform and it will be a role model for the Tamil Diaspora outside India and Sri Lanka.

Social networking: Social networking is vital for making necessary connection between the trainees. The inter and intra social networking enhances the responsibility and the integrity. Here in this project, this is happening and the schools are accepting as the trainee teachers used the products in their Teaching practicum at nearly 28 schools. So, in June another 15 will be trying this pattern of social network and it will become broader and deeper in the following years. As a result, our trainees will create their own PB works for their individual classes and they too invite their all Tamil classes and Tamil teachers to create a social network. At the end, 28X28X6= 4710 groups will be spanned and form a bigger network as aalamaram in Tamil. This is a short term result for us and indeed it will be a bigger and stronger pool for more resources. Yes, this will be a stronger and steady force for many predictable changes in the Tamil students’ thought, cognitive, psychological process.

Conclusion

To say that there are no resources is one thing. And to make good use of available resources is another thing. Showing others the usage of the resources and giving permission for this usage is another. At bringing together these aspects is a noble act. By doing this, all our trainee teachers have this facilities in their schools. It is notable that through this, there are many ways they can urge their students to create similar projects and discuss in depth about projects already presented.

References

2. Ida Fajar Priyanto. 2007. Developing IT-based teaching materials to enhance information skills and knowledge awareness among students. World Library and Information Congress: 73rd IFLA General Conference and council. 19-23 August 2007, Durban, South Africa
Use of Technology in Running a Tamil School in USA

Ilango Meyyappan
Principal, California Tamil Academy,
Fremont Branch, California, USA

Introduction

California Tamil Academy’s (CTA) primary activity is to teach Tamil to the children and young adults living in America. As an extension to teaching Tamil, CTA also supports developing cultural awareness activities such as music, dance, drama and any art form based on Tamil language and Tamil culture. CTA’s mission is to develop a love for learning Tamil that will last a lifetime and develop communication skills in Tamil so that the Tamils in America can appreciate family values and feel the binding with extended family members living in India as well as the US. CTA creates a forum for our youngsters to meet, learn, share and practice our culture and values whereby providing an identity to them. CTA nurtures and preserves our cultural identity and heritage to maintain essential family values and develop self-esteem and pride around the identity by participating in Tamil school and community activities. CTA helps children feel, internalize and be self-assured about the value of Tamil and remove the question of “Why should I learn Tamil?”

CTA objectives

CTA aims to develop and improve speaking, reading and writing Tamil. Teaching Tamil is much more than just teaching letters and sounds. CTA ensures that the kids achieve small successes every week and that parents are aware of their children’s progress in reading, writing and talking in Tamil. CTA ensures that knowledge passes on to successive generations. CTA believes that the kids learn Tamil culture along with the Tamil language. CTA creates a kid friendly environment for students to learn Tamil and has a "Commitment to Excellence" attitude and philosophy in teaching Tamil.

CTA History

CTA was started in 1998 with 13 students. Today, CTA has around 1800 students. CTA was started in Cupertino, California. Now, it is being run in 6 branches in Cupertino, Fremont, San Jose, Folsom, Foster City and Pleasanton. On top of this, CTA has affiliated schools in Novato, Atlanta, Seattle and Phoenix. CTA maintains a student teacher ratio of 8:1. We have around 250 teachers that teach Tamil every Sunday on a purely voluntary basis. With an ever increasing Tamil population in the US, CTA hopes to set an example as a role model school and is sure that Tamils living in other states in the US will eventually follow CTA model, syllabus and curriculum.

Use of technology

For such a massive operation that is run by volunteers who have other regular full time jobs, the use of technology is very essential in ensuring a smooth operation of the school. We will see how technology is
used in the various aspects of running this school and the benefits seen by both the parents and the teacher body by using technology.

**User profiles**

Every person that is in some way associated with the school can set up a user profile that has the basic information of the individual like his name, address and contact phone numbers. Each person is assigned a “role” by the management and the system. Based on the “role”, each person has certain accessibility restrictions and privilege to a certain kind of information. For example, a parent has only a “parent” role that limits access only to information regarding his or her own child and not any other child. A “teacher” role provides access only to that particular class and not any other class. A “principal” role provides access to all students, teachers, classes and reports only for that particular branch and no other branch. This ensures privacy of personal information and data.
School profile

CTA uses the regular public schools, private schools and college campuses in the US for classrooms. To create a school like atmosphere and bring seriousness into teaching and learning Tamil, CTA uses regular schools by renting them for 4 hours every Sunday morning. Information about the school like its name, address and location are stored in the system. The classrooms that are being used and the grades and sections for which these rooms are assigned are also stored in the system. The teacher information is also stored. This is so that management is aware of the school and the classes that are being run in each of these rooms.

Student registration

<table>
<thead>
<tr>
<th>View Payment Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Site</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Fremont</td>
</tr>
<tr>
<td><strong>Total Amount</strong>: 280.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>View Student Details for: Shijian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student ID</td>
</tr>
<tr>
<td>First Name</td>
</tr>
<tr>
<td>Grade</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Birth Place</td>
</tr>
<tr>
<td>Mother Name</td>
</tr>
<tr>
<td>City</td>
</tr>
<tr>
<td>ZipPostal Code</td>
</tr>
<tr>
<td>Mobile Number 1</td>
</tr>
<tr>
<td>Parent Email ID</td>
</tr>
<tr>
<td>Doctor Name</td>
</tr>
<tr>
<td>Dentist Name</td>
</tr>
<tr>
<td>EC1 Name</td>
</tr>
<tr>
<td>EC1 Relation</td>
</tr>
<tr>
<td>EC1 Street</td>
</tr>
<tr>
<td>EC1 City</td>
</tr>
</tbody>
</table>

CTA stopped accepting manual registration using a printed application form and switched to paperless and online registration. From 2009 onwards, all registrations were done online and the payments were also made online. When a student is registered, important information about him is stored in our system and management can access it anytime to contact a student. The school has all the emergency contact information about all registered students so that the student can be helped in case of an emergency.
**Student performance**

Teachers maintain an online logbook to monitor and track the performance of the students on a weekly basis. Attendance, classroom participation, reading skills, homework submission are all assessed on a weekly basis and marks for those are entered in the online logbook every week. Monthly test scores are entered every month. At the end of a particular term and at the end of a school year, a report card is automatically generated. The teacher can publish a report card at the end of every term so that the parents can know what scores their children have earned.

<table>
<thead>
<tr>
<th>LogBook Information</th>
<th>Attendance Details for School: Fremont Grade: Grade IV Section: A Date: 09/10/09</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student Name</strong></td>
<td><strong>Working Date</strong></td>
</tr>
<tr>
<td>Adithya Kumar</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Arvind Kandasamy</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Deva Dham</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Kirthi Ravi</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Nivethana Narayanan</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Sanjana Murthy</td>
<td>01/17/10</td>
</tr>
<tr>
<td>SATHYAPRIYA SARAVANAN</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Shreyas Srinivasan</td>
<td>01/17/10</td>
</tr>
<tr>
<td>Siddharth Raju</td>
<td>01/17/10</td>
</tr>
</tbody>
</table>

---

**California Tamil Academy**

**Student Progress Report Card for Academic Year 2009 - 2010**

<table>
<thead>
<tr>
<th>Term</th>
<th>Student Name</th>
<th>Attendance</th>
<th>Home Work</th>
<th>Reading / Conversation</th>
<th>Participation</th>
<th>Monthly Test</th>
<th>Total Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Adithya Kumar</td>
<td>9.09</td>
<td>10.00</td>
<td>5.00</td>
<td>5.00</td>
<td>68.25</td>
<td>97.34</td>
</tr>
<tr>
<td>I</td>
<td>Arvind Kandasamy</td>
<td>10.00</td>
<td>10.00</td>
<td>5.00</td>
<td>5.00</td>
<td>68.25</td>
<td>98.25</td>
</tr>
<tr>
<td>I</td>
<td>Deva Dham</td>
<td>10.00</td>
<td>9.09</td>
<td>5.00</td>
<td>5.00</td>
<td>64.75</td>
<td>93.84</td>
</tr>
<tr>
<td>I</td>
<td>Kirthi Ravi</td>
<td>10.00</td>
<td>10.00</td>
<td>5.00</td>
<td>5.00</td>
<td>66.50</td>
<td>96.50</td>
</tr>
<tr>
<td>I</td>
<td>Nivethana Narayanan</td>
<td>9.09</td>
<td>9.55</td>
<td>5.00</td>
<td>5.00</td>
<td>65.50</td>
<td>95.14</td>
</tr>
<tr>
<td>I</td>
<td>Sanjana Murthy</td>
<td>10.00</td>
<td>9.36</td>
<td>5.00</td>
<td>5.00</td>
<td>66.50</td>
<td>95.86</td>
</tr>
<tr>
<td>I</td>
<td>SATHYAPRIYA SARAVANAN</td>
<td>10.00</td>
<td>10.00</td>
<td>5.00</td>
<td>5.00</td>
<td>66.50</td>
<td>96.50</td>
</tr>
<tr>
<td>I</td>
<td>Shreyas Srinivasan</td>
<td>9.09</td>
<td>10.00</td>
<td>5.00</td>
<td>5.00</td>
<td>66.50</td>
<td>95.59</td>
</tr>
<tr>
<td>I</td>
<td>Siddharth Raju</td>
<td>10.00</td>
<td>9.55</td>
<td>5.00</td>
<td>5.00</td>
<td>68.25</td>
<td>97.80</td>
</tr>
</tbody>
</table>
Communication

Teachers use web based system for communicating with the parents. The system allows the teachers to selectively choose certain parents or mass email all the parents. Teachers can also post homework details and share documents with the parents. The principal also uses the web based system to send parent and teacher weekly updates that contain information about the school, syllabus, testing schedule, model question papers and make general announcements. The model question papers are uploaded before the monthly tests. The actual question papers are also uploaded and the schools download them and print them out. Since CTA has many branches and affiliation schools, the use of technology is absolutely essential for administering tests.

Multimedia in Syllabus

There is a separate syllabus committee that takes inputs from the teachers, parents, students and other Tamil educators and sets up a syllabus that is tailored to kids growing up in a foreign land where their main form of instruction and learning is in English. CTA uses a lot of multimedia to make learning visual and intuitive for kids growing up in the US. CTA developed and printed its own set of books for reading, writing and exercise for preschool I, II and Basic I, II and III. For grade I to VII, we use the textbooks recognized and used by the government of Tamil Nadu. For almost all grades, we have the textbooks on CDs. We also use other audio and video CDs which have songs, rhymes, poems and pictorial illustrations of letters, words and sentences that make learning easier and interesting.

High school credit program

CTA recently got approval for teaching Tamil to high school students towards satisfying the foreign language requirement. With this approval, Tamil students that are in high school can take Tamil as a language and satisfy their foreign language requirement. We use the same system for administering our high school credit program teachers and students. So far, our students have been taking either Spanish or French predominantly for foreign language requirement, now they are happy that they can take their own mother tongue and earn credits for that. Only few languages have earned this qualification with the school districts here in the US and Tamil is one of them.

Annual Day

CTA conducts an annual school day at the end of every year. We rent a state of the art auditorium and conduct the events in a professional manner. The entire process of registering for the event, registering kids for the event, uploading of songs and scripts for approval, managing the order and sequence of events are all automated and computerized. Nothing is done manually. We use googlegroups and googledocs to share documents, spreadsheets, create a repository of audio and video files and to automate the approval and registration process. CTA follows strict guidelines to maintain the quality of programs in our annual day. No English words or other language words can be used, no obscene gestures, no vulgar movements and no references to any religion or caste can be used. Only appropriate costumes can be worn by the children.

Annual day is celebrated to add fun to the learning process and also express your cultural awareness in the form of a song, dance or drama. Along with annual day, CTA celebrates Halloween, Deepavali, Thanksgiving, Christmas, Pongal, Tamil new year and conducts a graduation ceremony at the end of the year. These are all fun related activities that add another dimension to the learning process.
**Tamil Virtual University**

CTA runs TVU classes right after regular school hours are over. Beginner TVU, intermediate TVU, advanced TVU level classes are offered and many of our regular CTA students take TVU classes and get certified by TVU. We use the same system to register our TVU students, maintain their details and track their performance. Many of our students have been certified by TVU at all levels.

**Conclusion**

CTA’s mission and vision of imparting a high quality and streamlined education to Tamils living in the US is made possible by technology. Tamil education to kids born and raised in a foreign land has its own challenges. We try to solve those problems by putting a process in place that is consistent among various branches spread across the entire country. That process implementation and execution is made possible by technology. As we grow in number, we are also strengthening our technology infrastructure to accommodate the volume transactions and queries. The Internet and web based systems are the future for CTA to effectively and smoothly run such a big organization and operation.
கி.பி.க் கமு குழுக்களின் பிள.கமு விளக்கம்

சர்வதரைக் காலை முண்டலநூற்றாண்டில் நிற்பைக் கமு

1.0 ஆறை

1.1 பாரம்பரியப் பொருளைப் பயன்படுத்தும் காலை

1.2 மத்தவீடுகள்

1.3 நூற்றாண்டுகள்

2.0 தொடர்பு

2.1 நிர்வாகத்தொடர்

பாரம்பரியாக, காலைக் குழுக்கள் குழுக்களின் பிள.கமு, அது வெளிப்படுத்துவதற்குப் பிறகுக் கமுப்பைக் குழுக்கள், காலைக் குழுக்களின் பிள.கமு, அது வெளிப்படுத்துவதற்குப் பிறகுக் கமுப்பைக் குழுக்கள், காலைக் குழுக்களின் பிள.கமு, அது வெளிப்படுத்துவதற்குப் பிறகுக் கமு குழுக்கள், காலைக் குழுக்களின் பிள.கமு, அது வெளிப்படுத்துவதற்குப் பிறகுக் கமு குழுக்கள், காலை...
3.0 இதழானது கூடிய பராமரிப்பு நிறைவேற்றம்

இதழானது தொடரும் வரையிலிருந்து கூடிய பராமரிப்பு நிறைவேற்றம் நூற்றாண்டுகளுக்கு முந்தைய இக்காலத்திலிருந்து குறிப்பிட்டுக்கொள்ளப் பயன்பட்டுள்ளது. இதழானது பொருளாக்கள் மற்றும் இதழானதை விளக்கியுள்ள அருப்பாண்டுகளுடன் அனுமானத்தின் பராமரிப்பு நிறைவேற்றம் நூற்றாண்டுகளுக்கு முந்தைய இக்காலத்திலிருந்து குறிப்பிட்டுக்கொள்ளப் பயன்பட்டுள்ளது. இதழானது பொருளாக்கள் மற்றும் இதழானதை விளக்கியுள்ள அருப்பாண்டுகளுடன் அனுமானத்தின் பராமரிப்பு நிறைவேற்றம் நூற்றாண்டுகளுக்கு முந்தைய இக்காலத்திலிருந்து குறிப்பிட்டுக்கொள்ளப் பயன்பட்டுள்ளது.
3.2 பெண்கதரிகளின் ஒப்பிட்டு பார்வை

இந்துறுதுறையில் குறிப்பிட்டு, ஹீலியர்ஸ் பார்வைகள் குறிப்பிட்டுள்ளன. ஹீலியர்ஸ் பார்வைகள் அமையும் பார்வைகள் வைக்குறிப்பிட்டுள்ளன. ஹீலியர்ஸ் பார்வைகள் அடுத்து பார்வையளவி பார்வை, பார்வையளவிய வல்லுனர், பார்வையளவி பார்வையளவி (நடு. அடுத்துறையில். Tamil Internet செய்தியுள்ள 2003). ஹீலியர்ஸ் அடுத்துறையில் ஹீலியர்ஸ் பார்வையளவி அனுப்பிக்கும் பார்வையளவிய வல்லுனர் பார்வையளவி அனுப்பிக்கும் வல்லுனர் பார்வை.

இன்று அறிக்கை முனையக் கல்வியளவி வெளியீட்டு பார்வைகள் அடுத்துறையில் அடுத்துறையில், பார்வையளவி பார்வையளவி அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லுனர் அதிர்வேற்றிய வல்லு

3.3 குறிப்பிட்டு வைக்குறிப்பிட்டு பார்வை

இந்துறுதுறையில் பிறந்துள்ளன வைக்குறிப்பிட்டு பார்வைகள். வைக்குறிப்பிட்டு பார்வையளவி அடுத்துறையில், பார்வையளவி பார்வையளவி அடுத்துறையில் அடுத்துறையில், பார்வையளவி பார்வையளவி அடுத்துறையில் அடுத்துறையில், பார்வையளவி பார்வையளவி அடுத்துறையில் அடுத்துறையில், பார்வையளவி பார்வையளவிய வல்லுனர் பார்வை.

4.0 சிறுகாலத்தில் பார்வையளவிய அளவுப்படுத்தும் பார்வை
5.0 செயல்பாடு

செயல்பாடுகள் முறும்பெருக்கத்துடன் பருகி அறிவித்து, அறிவித்து பார்ப்பது, பார்ப்பவுடன் பார்ப்பது முறும்பெருக்கத்துடன் விளக்கமிட்டு. பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பவுடன் பார்ப்பை கேட்டு முடிகிறது.

5.0.1 கல்கத்தால்

கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கல்கத்தால் கேட்டு முடிகிறது.
A Study on the Role of Tamil Virtual University in Tamil Teaching and Learning at Elementary Level

Dr. Nirmala Devi S
Associate Professor & Head, Dept. of Education, Govt. Institute of Advanced Study in Education, Chennai.15 nirmaladevielango@gmail.com

Dr. Rajeswari T
Research Fellow, ECOLE FRANCAISE, Pondicherry

Introduction

Education is a process of human enlightenment and empowerment for the achievement of a better and higher quality of life. Education develops the total personality of an individual. It also contribute to the growth and development of the society. The field of education is expanding each year as advancement is made in technology and brain based research.

Technology and Education

New developments and new technologies came into existence and any growing nation cannot remain away from the needs to change in time so that the society is not labeled as under developed and backward society. The process of education cannot ignore the social and psychological impacts of the technology that structures information.

Importance of Teachers

Teacher is a significant agent in causing learning and intellectual development of the learner. It is necessary that teachers keep themselves abreast of new developments. Educators continually strive to maximize the effectiveness and efficiency of teachers by identifying and comparing alternative method of teaching and learning.

Changing Role of the Teacher

During the era of Information Technology the teacher can be called as a mentor, monitor and motivator to develop better learning styles and information seeking behaviour among children. Hence during the 21st century teacher’s new role is that of a facilitator and designer of learning experiences and to maintain innovations in the classroom.

Need for the study

The course of Tamil Virtual University is available free of cost.

Any new technology comes not merely with hardware and software but with a learning and teaching style and grammar of its own, and that management practices need to be adapted in order to use the technologies effectively.
With poor access and high digital divide teachers tend to resist adoption of e-learning.

To develop teacher competencies in use of web technologies, to plan for capacity building and Training, the teacher’s awareness about the functioning of Tamil Virtual University, its various teaching modules, their attitude towards the usage of web based technology need to be assessed. This will help the planners and administrators of education to plan and implement in service program to teachers and to make certain changes in curriculum, to make recommendations to the govt the need for improving the infrastructure facilities and technical support in schools.

Need for Technology

Web based learning is an essential tool for achieving sustainability and will help in enabling better and increased access to information to enrich the teaching learning process.

Tamil Virtual University

The Tamil Virtual University was established in 2000 by the Government of Tamilnadu, aims at providing Internet based resources using multimedia and opportunities for the Tamil communities living in different parts of globe as well as others interested in learning Tamil and acquiring knowledge of the history, art, literature and culture of the Tamil. This is the first virtual university in India for language (Tamil) teaching. Similarly the Indira Gandhi National Open University (IGNOU) started virtual campus initiatives (1999), Netvarsity (1996) first online learning facility by NIIT, Yashwantrao Chavan Maharashtra Open University are some of the other virtual universities offering courses through on line mode. Newport Asia Pacific University (NAPU) a new virtual university that offers programs to teach Japanese as a second language. Nunnan, David (2002)

The TVU web based course is designed and given in four levels viz

1. Pre Primary Level
2. Basic Level
3. Certificate Intermediate Level
4. Certificate Advanced Level

The course material is also available in CD. The certificate course is recognized by the govt. of Tamilnadu. The content given in the CD includes an introduction, lessons related to listening, reading skills, introduction of grammar components, and related exercises, follow up activities, self evaluation.

Objectives of the study

The main objective of the present study is to find out

How effectively this web resource is utilized for Tamil teaching-learning processes at the elementary level by the Tamil teachers

Which part and which level of the content given in the web is being widely used by the teachers

Their attitude towards the usage of this web resource

Their willingness to learn the fundamentals of integrating technology in classroom, type of response and interest shown by the students in using web resources
Whether students are willing to use this web resources for improving their language skills,
The encouragement and motivation given to the teachers by the Govt. and Management, the infrastructural facilities available in the school premises

Method and Procedure
The present study employed Descriptive Survey research method.

Sample
190 Primary School Tamil Teachers (in and around Chennai and Puducherry) constituted the sample for this study.

Tool
The researchers constructed a questionnaire with 22 statements on a 3 point scale under four dimensions (Yes, No, At times) for this study. (Sample tool is given in Table1)

Analysis of Data
The collected data is quantified and the findings are given as percentage for the responses related to the option -Yes.

Findings of the Study
78% of the sample were aware of the establishment and functioning of TVU, 69% were familiar with the web site address. 60% have expressed that they made attempts to know about the content developed by TVU. 66% agreed that the visuals given for writing practice is very useful. 89% confirmed that their students are interested to learn Tamil through the web using Multimedia technology. 86% stated that learning Tamil through this virtual mode is very useful for the students. 77% were of the opinion that this virtual mode has made teaching and learning of Tamil easy.

75% agreed that lessons given at the intermediate level do develop language skills of the students. Only 50% of the sample have stated that they have the relevant facilities in schools to utilize this web resources, 50% thought of using web based technology.

53%, 66%, 77% have agreed for the statement 16, 17, &18 respectively.

77% confirmed that poetry lessons were very interesting. 90% accepted the need for change in the teaching methodology. 61% were of the opinion that there are differences between traditional methods and web based multimedia technology. 84% are willing to undergo training in web based teaching methodology. 90% wanted training in all aspects to use this web based technology for teaching learning purposes.

Educational Implications of the Study
1. Intensive basic training program, periodical refresher courses must be planned and implemented.
2. The web based learning material CD can be given to all schools.
3. Technology based language teaching to be introduced in the primary level school curriculum.
Conclusion

Since 90% of sample studied want change in the teaching methodology and 89% students are interested in learning through this method if proper facilities and support services along with effective training are provided tamilthai will get a new face in the 21st century.

References


3. ERIC WEB Sources.

4. New Trends in Pedagogical Techniques. M.Phil Education Tamil Nadu Open University, School of Education.

<table>
<thead>
<tr>
<th>SNO</th>
<th>DESCRIPTION</th>
<th>Yes/ No/ At times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Are you aware that the Tamil Virtual University established by the Govt of Tamilnadu has been functioning?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Have you come across the website of Tamil Virtual University : <a href="http://www.tvu.org">www.tvu.org</a></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Have you made any attempt to know that the TVU has prepared lessons for school students?</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>As of now teaching is done in 4 stages in TVU. Mark the stage which you were aware of 1.Primary Education Level 2.Basic Level 3.Certificate Intermediate Level 4. Certificate Advanced Level</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Which among the following lessons of Primary Education you have used? words, songs, events, conversation, stories, numbers, letters</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Indicate which among the following lessons given for basic level in TVU that you are aware of 1.lesson on Tamil letters 2.Group of Tamil letters 3. Writing Practice 4.Kirantha Eluthukkal 5. Song on introduction of yureluthukkal</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Is the visuals on writing practice given in the web easy to follow?</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Whether students are interested to learn lessons through this audio visual mode?</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Is teaching through AV MODE is beneficial to the students?</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Is teaching through AV mode make the process of teaching easier?</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Whether students are interested to learn through multimedia?</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Is the lessons meant for intermediate level enhances the language skills of the students?</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Do the illustrations given for vatrumai urubu make the students to learn grammar with interest and willingness?</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Are the students willing to learn through the web?</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>i) Do you have any idea of teaching through web to improve the language skills of your students ? ii) Do you have the facility to utilize web for improving the language skills of your students in your school ?</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Do you consider the lessons of higher level in the web are equivalent to VI std of formal school level?</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Is the method of practice given in the higher level is easy to learn?</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Is the poetry lessons given in the web motivate learning ?</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Do you agree that the teaching learning methods to be changed according to the changing time?</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Is there any difference between the traditional and multimedia methods of teaching in enhancing the language skills of students ?</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Are you interested in undergoing training programs to use web based teaching methods in your class?</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>What sort of training you require?</td>
<td></td>
</tr>
</tbody>
</table>
மேம்பாடுக் குறியீட்டுக்கான நல்லது உயிர்
குறிப்பிட்டுக்கும் செய்தல்

Paramasivam Muthusamy
University Putra Malaysia
param@fbmk.upm.edu.m

பல்வேற் சர்

'பத்தாண்டு' பல்கலைக்கழகம் நிறுவதற்கான குறிப்பிட்டு (பத்தாண்டு குறிப்பிட்டு) குறிப்பிட்டும் அடைந்தது. இதுவாழ மாதானே நிறுவதற்கான குழுக்களுக்கு சுருங்கி பாதமுகம் பெறுதல் குறிப்பிட்டு. பத்தாண்டு பல்கலைக்கழகம் நிறுவனத் தலைவர் பேச்சியால் பெறுதல் குறிப்பிட்டு. பத்தாண்டு (1992), கூட்டம் பத்தாண்டில் பத்தாண்டில் பெறுதல் குறிப்பிட்டு. அவர்களின் நிறுவன விளக்கத்தை பாதமுகம் பெறுதல் குழுக்களின் குழுவை குறிப்பிட்டு மற்றும் நிறுவனங்கள் அறக்கட்டளை நிறுவனத்தின் குழுவை குறிப்பிட்டு. தோலை பல்கலைக்கழகம் மற்றும் நிறுவனங்கள் மற்றும் விளக்கத்தை குறிப்பிட்டு. செருமனி பல்கலைக்கழகம் மற்றும் பல்கலைக்கழகம் மற்றும் விளக்கத்தை குறிப்பிட்டு.
சேலம் கட்டுரையாளர் சான்று.

பொறித்தல் கொண்டாடும் சேலம் கட்டுரையாளர் நிலவிய பாதுகாப்பு முன்னேற்கும் பாதுகாப்பு முன்னேற்கும். பலவையை நிற்பெறும் ஒன்றுக்கோரும் பலவையை நிற்பெறும் ஒன்றுக்கோரும். பலவை நிற்பெறும் ஒன்றுக்கோரும் பலவை நிற்பெறும் ஒன்றுக்கோரும். பலவை நிற்பெறும் ஒன்றுக்கோரும் பலவை நிற்பெறும் ஒன்றுக்கோரும். பலவை நிற்பெறும் ஒன்றுக்கோரும் பலவை நிற்பெறும் ஒன்றுக்கோரும்.

இருவரும் பலவையை நிற்பெறும் ஒன்றுக்கோரும் அனைத்து சனவரியும் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அனைத்தும் காணப்படும் குறும்பு பலவையை நிற்பெறும் ஒன்றுக்கோரும். பலவை நிற்பெறும் ஒன்றுக்கோரும். பலவையை நிற்பெறும் ஒன்றுக்கோரும். பலவையை நிற்பெறும் ஒன்றுக்கோரும். பலவையை நிற்பெறும் ஒன்றுக்கோரும். பலவையை நிற்பெறும் ஒன்றுக்கோரும்.

அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும்.

அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும்.

அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும். அவன் பலவையை நிற்பெறும் ஒன்றுக்கோரும்.
பனிகுல 'He has a son' என்றும், 'அவர் என் பாலவிழித்தார்' என்றும் வாரம்போற்றும். அதியும், மாற்கின்ற அப்பாலப்படுத்தி வருவார்கள் என்ற ஆவியாகத் தொடங்குகின்றன. 

சேனிய அவர் என் பாலவிழித்தார் என்றும் பாலவிழித்தார் என்றும் குறிப்பிட்டியது என்று இரண்டு வழிமுறை வருவார்கள்.

பனிகுல/ par

'அப்பால மனிதனை வருவாரும் மனிதனை வருவாரும் மனிதனை வருவாரும்' என்றும். குறிப்பிட்டியது குறிப்பிட்டிய மனிதனை வருவார்கள் மனிதனை வருவார்கள்.

பக்தியும் பாலவிழித்தும் வாரும் பாலவிழித்தும்

panikul panikul par par balavitham

இரண்டும் 'பாலவிழித்தார்' என்றும் பாலவிழித்தாரும் பாலவிழித்தார் பாலம் பாலம் பாலமும் பாலாமலும் வாரும். மனிதனை வருவார்கள் என்றும் பாலமும் பாலாமலும் மனிதனை வருவார்கள் என்றும்.

பாலவிழித்தும் வாரும் பாலவிழித்தும் வாரும், அவர் அவரிடம் வாரத்தினருக்கு வாரும், அவர் அவரிடம் வாரும் அவரும் குறிப்பிட்டிய மனிதனை வருவார்கள்

பாலமுதலாக

சேனிய வாரும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும்

I. குறிப்பிட்டு என்றும் பாலமும் பாலாமலும்

II. பாலமும் பாலாமலும் வாரத்தினருக்கு

இரண்டும், வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு

பாலிகள்

சேனிய வாரும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும் வாரம்போற்றும். இத்தொடராக வாரத்தினருக்கு

I. குறிப்பிட்டுப் பாலம்

II. பாலமும் பாலாமலும்

இரண்டும், வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு

பாலமைக்காரண

குறிப்பிட்டு குறிப்பிட்டு பாலம் பாலமைக்காரண வாரத்தினருக்கு வாரத்தினருக்கு. இது வாரத்தினருக்கு

i. மேலும் என் என் என்

ii. வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு வாரத்தினருக்கு

44
iii. கல்வி-அட்சப்படிப்பில் நேர்ப்பிட்டுவர்வகள்: குறிப்பிட்டு, இப்பகுதியில் பெருந்துகள் குறிப்பிட்டுவர்வகள்.

iv. இந்தியாவில் பெருந்துகள் குறிப்பிட்டுவர்வகள்: குறிப்பிட்டு-அட்சப்படிப்பில் பெருந்துகள் குறிப்பிட்டுவர்வகள் (Keyboard Managers) மற்றும் அவற்றுடன் அறிவியல்

i. குறிப்பிட்டு 2000

ii. குறிப்பிட்டு 2000

iii. குறிப்பிட்டு 2000

iv. குறிப்பிட்டு 2000

v. குறிப்பிட்டு 2000

vi. குறிப்பிட்டு 2000

vii. குறிப்பிட்டு (ISM)

viii. குறிப்பிட்டு (WINKEY)

ix. குறிப்பிட்டு (WINKEY)

• MS. WORD
• MS POWER POINT
• MS FRONT PAGE
• MACROMEDIA DIRECTOR
• HOT POTATO

• STORY BOOK WEAVER DELUXE
• KID WORKS DELUXE
• 3D MOVIE MAKER
• NETSCAPE
• MS WORD
• MS POWER POINT
கருதிகளை பிரிக்கவுமாறாள்

பொய்ந்தை எடுத்துக்காட்டு எதிர்ப்பு? அல்லது பார்வத்தால் எதிர்ப்பு?

- பார்வத்து எதிர்ப்பு கருதிகள் ஒவ்வொன்றும் பிரிக்கவும் இல்லை பிரிக்கவும் இல்லை.

ஆனால், பார்வத்து எதிர்ப்பு பிரிக்கவும் தலையில் இல்லை பிரிக்கவும் இல்லை. பார்வத்து எதிர்ப்பு பிரிக்கவும் தலையில் இல்லை பிரிக்கவும் இல்லை. (Standard Spoken Tamil) குறிப்பிட்டு கூறுவது என்பது

ஆனால், பார்வத்து எதிர்ப்பு பிரிக்கவும் தலையில் இல்லை பிரிக்கவும் இல்லை. பார்வத்து எதிர்ப்பு பிரிக்கவும் தலையில் இல்லை பிரிக்கவும் இல்லை.

புது

பொய்ந்தை எடுத்துக்காட்டு எதிர்ப்பு? அல்லது பார்வத்தால் எதிர்ப்பு?

- பார்வத்து எதிர்ப்பு கருதிகள் ஒவ்வொன்றும் பிரிக்கவும் இல்லை பிரிக்கவும் இல்லை.

ஆனால், பார்வத்து எதிர்ப்பு பிரிக்கவும் தலையில் இல்லை பிரிக்கவும் இல்லை. - உரைப்பு, மறுப்பு பொருள்/தொடர்பு உரைப்பு, மறுப்பு பொருள்/தொடர்பு உரைப்பு, மறுப்பு பொருள்/தொடர்பு உரைப்பு.

தகவல் தொகுப்பு

1. கிராமாராம்பு, மக. சுந்தரா ஓர்க்க கருதிகளின் உரைப்பு என்று (1992)

2. புகழ்பெற்றார் ஓர்க்க கருதிகளின் உரைப்பு, என்று குறிப்பிட்டு (2006)


46
ICT for Tamil education in Tamilnadu
Current Challenges and Opportunities

Prof S. Balaji,
D.B. Jain College (Autonomous), Chennai-97
Mobile: 9840494643
srisaibalaji@rediffmail.com

Introduction
Tamil is a Dravidian language spoken predominantly by Tamil people of the Indian subcontinent. It has official status in the Indian state of Tamil Nadu and in the Indian union territory of Puducherry. Tamil is also an official language of Sri Lanka and Singapore. It is one of the twenty-two scheduled languages of India and the first Indian language to be declared as a classical language by the government of India in 2004. Tamil is also spoken by significant minorities in Malaysia, Mauritius and Réunion as well as emigrant communities around the world.[1]. The use of technology for Tamil education for students is very minimal. This is due to lack of confidence, demand for the use of English in the application of technology. Even though the students appreciate the application of technology in learning and teaching of language, these two reasons prevent them from active application of technology.

Driving forces
Considering the fact that English is becoming the dominant home language in most Indian households, there is a need to do more to help children from households who have little exposure to Tamil. Therefore, there is a need to review how we teach Tamil, especially at the primary level, so that our students do not lose interest in the language. In particular, teachers need to concentrate on teaching oral communication skills to the younger generation so that they can communicate in the language more confidently, effectively and in greater depth, and will be motivated to use Tamil within and outside of school. Our children have grown up very comfortable with technology - they use mobile phones, play computer games and surf the net - we should tap on their IT literacy as well as the excellent infrastructure in schools to teach the languages. The illiterate, the physically challenged and the facility-challenged, all of them need some support or the other to be accepted in society and enjoy the fruits. The traditional methods and practices are invariably driven by us, human beings, and therefore tend to be biased.

The introduction of computer (also known as ICT or IT) has changed the scenario to that of an interactive, collaborative environment where the quest for information knowledge is created actively by students. It is here that ICT can help by providing independence, flexibility and variety to the “less privileged learner”. This is yet another virgin field, where ICT could have a lasting impact, in terms of enhancing the teaching and learning capabilities, respectively, of the agents of change (teachers) and the beneficiaries (the students).
Impact of ICT

Computer-aided education was initially introduced in India as an innovative activity under the District Primary Education Programme (DPEP) in 1994. ICT entered the education sector through DPEP, which spearheaded the design of school information systems. Literature on the role and efficacy of ICT in education is replete with insightful studies. Resnick (2002) opined that computer was akin to finger-painting. In the realm of learning, technology could be employed by students for “making” things, i.e., usage of technology to design and build things of importance. This would consequently increase the acceptance and adoption of technology in the classrooms. Kshetrimayum (2007) elaborated on multiple perspectives germane to incorporation of ICT in teaching and learning processes: simulation, visualization and modelling constituted the pedagogical perspective; assessment forms cognition perspective; e-learning and virtual learning environments comprise content delivery perspective; and finally, from project work and task perspective ICT with a combination of pedagogy & software design could lead to a collaborative environment.

Challenges ahead

The Human Resource Development department along with the Department of Information Technology has developed a report on technology in education. The report has identified four issues in integrating technology in education in government schools, namely, ICT infrastructure, quality content that is locally relevant, teachers training, and education delivery through public-private partnerships. These four interdependent issues needs to be addressed if technology has to be integrated to formal education to improve the quality of education in government schools.

A typical rural school has a number of inherent limitations viz., limited number of qualified teachers, archaic classrooms, chalk-and-talk methodology, variations in curricula (state to state and board to board), unreliable power supply and lack of interactions with the rest of the world.

The advent of cyber cafes and CD-based courseware has opened up possibilities of alleviating the problem. But, the problem still persists i.e., the non-availability of curriculum- and language-specific material backed up by suitably-trained teachers.

The pathway for opportunities

The activities use simple and use commonly available software such as Microsoft Power Point. They also provide sufficient flexibility for teachers to modify them according to their objectives, class or pupils’ abilities. Much consideration went in the designing of the software such that each lesson constitutes of only a few slides. It was essential that the pupils are not intimidated by the software and hence simple navigation tools were used. We decided against using more advanced authoring software like Mediator, Macromedia Flash or Multimedia Builder as we want Tamil language teachers to modify and enhance our models.

Basic Requirements in using our Microsoft PowerPoint models:

1) Pupils have
   • basic IT knowledge
   • basic knowledge of Microsoft PowerPoint
   • Usage of recording and playback features in PowerPoint
   • Typing text and annotating in different colours and sizes
2) Teachers have
   • comfortable knowledge in PowerPoint
   • using PowerPoint as an effective tool to translate their language teaching strategies into using IT

The design and use of the software took into consideration the following:
1) Prior IT knowledge of pupils:
2) Time needed to complete the lesson.
3) IT equipment generally found in Singapore schools.
4) Various skills required for the teacher to design the software

Human Factors experts have researched the strengths and limitations of different types of the physically challenged and come out with norms and standards for equipment design and HCI interfaces. Experts and researchers have exploited the following methods to make the physically challenged as independent as possible:

   a) Audio cues to help the blind
   b) Rich graphics to help the deaf
   c) Combination of the visual and audio cues to help the partially deaf/blind
   d) Flexible input and output devices for the users with limited movements of the limbs and the body
   e) A variety of cues to keep up the interest and excitement in children with learning disabilities

After the stage of memorization and teaching through class representative or leader, blackboard and chalk pieces came about. After which, teaching tools; such as keyboard, computer, smart board, and Tablet PC that consist of computer and mobile phone provide students with the language benefits in class. Tamil letters, Tamil songs, Tamil vegetables, Grandmother stories, are all being sold in the form of CDs/DVDs even in today’s commercialized level, and all these have; Tamil’s nuances, the beauty of pronouncing in Tamil, vocabulary building in Tamil, India’s nature as well as the beautiful Tamil spoken by qualified hosts in their native language that provides a feast for students who hear and view them. Here, the beauty of the language and the benefits of its nativity are displayed in a manner that students can know about. In this stage, we shall see how information technology is used in teaching and learning, at National Institute of Education that trains teachers, who teach Tamil.

Scrimshaw (2004) also points out that the implementation of ICT in the classroom is “both an innovation in technology and teaching” (p.9). On the other hand, multimedia is a combination features of text, graphic, art, sound, animation and video elements with facilities for interaction. Thus, multimedia is a powerful presentation tool, which can be effectively used for teaching. Studies showed that if students are stimulated with audio, they will have about 20 % retention rate, audio-visual is up to 30 % and in interactive multimedia presentation, the retention rate is up to 60% (Vaughan, 1997; p10). Hence, multimedia tools can enhance many skills such as, functional communication as a result of enriched vocabulary, critical and creative thinking.

**Government Initiatives and pathway**

A need to develop IT Tools to facilitate human-machine interaction, and to promote the use of these tools for various Indian Languages was felt. Towards this, the Department of Electronics of Government of
India had initiated activities in the area of Technology Development for Indian Languages with the following objectives:

- To develop information processing tools to facilitate human machine interaction, information processing in Indian languages and development of multi lingual knowledge systems
- To promote the use of information processing tools for language studies and research.
- To support R&D efforts in the area of information processing in Indian languages covering machine translation, human machine interaction, language learning and natural language processing.

“ICT@school” was one of the important central government schemes formulated in the tenth five-year plan. Its objective was: “promoting usage of ICT in government and government aided schools (particularly in rural areas) and providing ICT infrastructure.” This scheme had four components (GOI, 2007a):

1. Partnership with state governments and union territories for providing computer education and computer-aided education to govt. and govt. aided schools
2. Establishment of SMART schools - which will be technology demonstrators
3. Universalization of computer literacy thru Kendriya Vidyalayas and Navodaya
4. State Institutes of Educational Training to provided educational content in the form of films, videos and audios etc.

The National Mission for Education through ICT has been envisaged as a federally sponsored scheme to leverage ICT so as to democratize education by providing high quality personalized and interactive knowledge modules over the internet/intranet for all the learners in the higher education institutions. While the eleventh five-year plan (2007-12) has taken into cognizance the importance of education in nation-building, it provided the much needed fillip for expanding the scope of ICT usage in Indian schools. A substantial budget has been allocated towards upgrading technology infrastructure (including ICT) in schools. Targets have been established for reaching out to Upper Primary Schools to have coverage of ICT by 2011-12.

The ICTACT (ICT Academy of Tamilnadu), a consortium of the Centre, State and Confederation of Indian Industry, will provide industry relevant training programmes in the ICT spectrum to the faculty members of the university and its affiliated colleges. Various training programmes will be sponsored by the ICT industry.

**End Notes**

ICT can help bridge the rural-urban divide, as has been demonstrated by projects that have been successfully rolled out across the globe. Most teachers do utilise programs such as PowerPoint but only as a presentation facility in their IT classes (as a cognitive tool). Apart from this, commercially available software including e-learning platforms, emphasis lies in self-paced learning. Hence, by using IT as an effective tool to develop the metacognitive skills, we hope to improve their oral and aural performance of Tamil Language pupils.
References


2. ^ Zvelebil 1992, p. 12: "...the most acceptable periodisation which has so far been suggested for the development of Tamil writing seems to me to be that of A Chidambaranatha Chettiar (1907–1967): 1. Sangam Literature – 200BC to AD 200; 2. Post Sangam literature – AD 200 – AD 600; 3. Early Medieval literature – AD 600 to AD 1200; 4. Later Medieval literature – AD 1200 to AD 1800; 5. Pre-Modern literature – AD 1800 to 1900"


Enhancing Activity based Tamil Teaching and Learning using Online Video Repositories

A Data Mining based Approach

Dr. K. Vivekanandan
Associate Professor
School of Management
Bharathiar University
Coimbatore – 641 046
vivekbsmed@gmail.com

Dr. V. Saravanan
Professor & Director
Dept. of Comp. Applications
Dr. N.G.P Institute of Technology
Coimbatore – 641 048
profsaran@hotmail.com

Mr. P. Ranjit Jeba Thangaiah
Assistant Professor (SG)
Dept. of Comp. Applications
Karunya University
Coimbatore – 641 114
prjt@live.com

Abstract

Online Teaching and Learning shows a considerable impact in the education system. The activity based education system introduced by the Government of Tamilnadu in the schools shows a remarkable impact on improving student learning across Tamilnadu. This has been witnessed by many visitors across the world wanting to know about this system and the efforts taken by the Government for successful implementation. The video based interactive tutorial helps the students to effectively learn the contents and also interact with the expert. With the decrease in hardware costs every year, storing much video contents is no longer a costlier process. Data Mining techniques and algorithms are the actual tools that analysts have at their disposal to find similar patterns and correlation in the data.

This paper proposes of presenting the activity based Tamil teaching as short video clips. This video clips are available online to all the students. To start with, set of Tamil teachers are chosen as experts for all the Tamil topics of various classes. The experts’ lectures/demo is video recorded and stored in the repository. As the lengthiest lectures will reduce the student interest on the subject, all the lecture topics are divided as short videos and made available for students access. The short video lecture will runs from 3 to 5 minutes. When a student browse through a video lecture and finishes the execution, all the other relevant lectures related to the topic viewed by the students also displayed. The students can easily view the other topics without starting searching for another topics or a topic which cannot be searched very easily. Data mining techniques are used to find the identical patterns. These techniques capture the students’ behaviors and also trace through the similar navigations performed by other students across Internet. The related short video contents are grouped together and presented to the students. Grouping of similar patterns are carried out by applying data mining techniques such as Association rule mining (determine implication rules for a subset of video lecture attributes, Classification (assign each video record of a database to one of a predefined set of classes analysis) and Clustering Techniques (find groups of similar video records that are close according to some user defined metrics).

This model enhances the activity based learning and surely will create an interest on “Tamil” or Tamil way teaching among all school/college students. This model can also be used to teach other subjects in all
schools/colleges through Tamil. As the telecommunication sector witnesses a major breakthrough using 3G Technologies, all these video lectures can be easily accessed through Mobile phone environment also. The developed model can also be easily implemented by setting up a Multimedia Kiosk/Centre.

Keywords: Data Mining, Teaching & Learning

Introduction

The schools and Institutions of higher education have increasingly embraced online education, and the number of students enrolled in distance programs is rapidly rising in colleges and universities throughout Tamilnadu. In response to these changes in enrollment demands, many schools/colleges/universities have been working on strategic plans to implement online education. The activity based education system introduced by the Government of Tamilnadu in the schools shows a remarkable impact on improving student learning across Tamilnadu. This has been witnessed by many visitors across the world wanting to know about this system and the efforts taken by the Government for successful implementation

Review of Literature

We began writing this paper with a review of past studies of the issues and trends in online teaching and learning in higher education. A recent survey of higher education in the United States reported that more than 2.35 million students enrolled in online courses in fall 2004 [4]. This report also noted that online education is becoming an important long-term strategy for many postsecondary institutions. Given the rapid growth of online education and its importance for postsecondary institutions, it is imperative that institutions of higher education provide quality online programs.

Faculty training and support is another critical component of quality online education. Many researchers posit that instructors play a different role from that of traditional classroom instructors when they teach online courses.

Using Short Video for Lecture/Demo

The video based interactive tutorial helps the students to effectively learn the contents and also interact with the expert. With the decrease in hardware costs every year, storing much video contents is no longer a costlier process.

The following are the few advantages for the students by using the proposed approach

a. It generates enthusiasm in the students by using interesting or unusual examples.
b. More useful feedback from students are received by phrasing questions in a positive manner
c. It engages the students with leading questions to actively involve them in the classroom.
d. It encourages collaboration among students.
e. It gives useful comments on homework.
f. Simultaneously challenge the students to think and also obtain critical feedback about what the students know by asking the right kinds of questions.
g. Best to explain specific Tamil topics such as Ilakanam, Seiyul and other advanced topics.
This paper proposes of presenting the activity based Tamil teaching as short video clips. This video clips are available online to all the students. To start with, set of Tamil teachers are chosen as experts for all the Tamil topics of various classes. The experts’ lectures/demo is video recorded and stored in the repository. As the lengthiest lectures will reduce the student interest on the subject, all the lecture topics are divided as short videos and made available for students access. The short video lecture will runs from 3 to 5 minutes. When a student browse through a video lecture and finishes the execution, all the other relevant lectures related to the topic viewed by the students also displayed. The students can easily view the other topics without starting searching for another topics or a topic which cannot be searched very easily.

**Using Data Mining Techniques to Enhance Teaching and Learning**

Data Mining is the technique to explore and analyze the large data sets, in order to discover meaningful patterns and rules. The evaluation of data mining techniques began when the business data are stored in the database and the technologies were generated to allow the user to navigate the data in the real time. Data mining techniques are used to find the identical patterns. These techniques capture the students’ behaviors and also trace through the similar navigations performed by other students across Internet. By considering the proposed objective, this paper proposes the use of data mining techniques to enhance teaching and learning. The major data mining techniques considered in this paper are

a. Association Techniques.

b. Classification techniques.

c. Clustering Techniques.

**Association:** It is method for discovering interesting relations between the variables in the large database. There are different types of algorithm for association rule. They are Apriori algorithm, éclat algorithm, FP-growth algorithm, One-attribute-rule algorithm, Opus search algorithms, and Zero-attribute-rule algorithm.

Grouping of similar patterns are carried out by applying data mining techniques such as Association rule mining (determine implication rules for a subset of video lecture attributes). The students studying a particular video lecture also studies a related content. The associations between the related contents are displayed at the end of a lecture which helps the student to navigate among the interested pages. This increases the students’ interest and also creates an indirect interest on the content and learning.

Using association rule algorithms, the subjects associated with other subjects can be sorted out and an impact can be given to the user that; if the user is viewing/studying a particular subject, the association rule algorithms in turn advises to study/view a video that are related to the subject. In this way the related videos are ranked and shown to the user. This will enhance a user to learn more on a subject with all the associated materials related to the subject.

**Classification:** It is one of the data mining techniques used to predict the group for data instance. Some of the popular classification techniques are decision trees and neural networks. From the existing database, the end user can classify the land with required parameters like state wise, of district wise, area wise and etc by means of tree like structure.
The above listed classification techniques (assign each video record of a database to one of a predefined set of classes analysis) helps the students to identify a particular subjects and search through specific topics. As the classification process proceeds, the left part or the right part of the tree is considered for better visualization and reading of a subject.

**Clustering:** It defined as collection of data object that are similar to one another within the same cluster and dissimilar to the objects in the other cluster. Clustering algorithms are broadly classified into hierarchical and partitioning clustering algorithm. Again, the Hierarchical algorithm are Agglomerative and Divisive algorithm and the Partitioning Algorithms are k-means, k-mediod, DBSCAN, CLARA, CLARANS, BIRCH CLIQUE, OPTICS etc.,

Clustering Techniques (find groups of similar video records that are close according to some user defined metrics) groups the related subjects and helps the student in choosing related content without wasting much time of choosing what contents to be chosen next or in future.

**Discussion and Conclusion**

As schools and institutions of higher education continue to embrace and debate Tamil online learning, it is important to envision where the field is headed. It is the appropriate time for us to propose suitable technical model / approached which will help the Government of Tamilnadu to implement the projects better. This model can also be used to teach other subjects in all schools/colleges through Tamil. As the telecommunication sector witnesses a major breakthrough using 3G Technologies, all these video lectures can be easily accessed through Mobile phone environment also. The developed model can also be easily implemented by setting up a Multimedia Kiosk/Centre.

**References**

The Computer, Tamil Language ,
The Tamil Teacher
Jayasarasvathi DuraiKannu
Bedok West Pri Sch – Singapore
Email: jayasarasvathi_durai_kannu@moe.edu.sg

அறி)க

Jவ நாமாக இபி"#
வாவ தமிழாக இக#
!ற கைலஞாி! வாசக1தி - ஏ ப ‘கா + அைட1த ைபயான இ1ேதக#’ இ!+ ேதா!றி நாைள
மைறயலா#. ஆனா க ேதா!றி ம) ேதா!றா ;!ேன ேதா!றிய 41த -(யி! ெமாழியான, தமி
ெமாழி. ெச#ெமாழி எ!ற ம-ட# Bடப J+நைட ேபாகிற. தமி ெமாழி பழைமயான !ப
உ)ைம. பழைம நிைற3த அவ இளைமயானவE# @ட. அவ இலகியகளி ம# களிநடன#
Aாிவதிைல, இைணய1தி0# உலா வகிறா எ!ப மாணவ ச;தான1தி -# ந# இைளய
ச;தாய1தி -# உண%1தபட ேவ)#. இ1தி பணிைய1 திற#பட ெச6ய@(யவ%க ஆசிாிய
ெப3தைகக.
க ற க பி1த எ!ப கவி க றF! இ @+க. ஒ @+ சிறபாக அைமயாவி(! ம ெறா
@+ திற#பட அைமயா.
‘

’

எ

எ

கணினி ம ; இைணய க பித ேதைவயா?

இ!ைறய கால1தி இைணய# !ப கவி க பி1தF ஒ உ1தியாகேவ கதபகிற. கால
மா%, க"  மா%, நா. மாற ேவ$ எ!ற வாிகE- ஏ ப சிறா% ;த ெபாியவ% வைர
இைணய#பா ெகா)(-# ஈபா(ைன க1தி ெகா) இைணய1ைத க பி1த0கான
ேமைடயா6 அைம1த சால சிற3த.
இைணய# வழி தமி க பிபதா தமி, அழ-1தமிழா6, பழ- தமிழா6, வள%தமிழா6, இ!ப1தமிழா6,
தீ3தமிழா6, நைட;ைற1 தமிழா6, உெவபைத க)@டாக பா%கலா#.
இ!ைறய நJன உலகி கவி க பி-# உ1திகளி கணினி ம +# இைணய# ;கிய இட1ைத
வகிகிற. இைணய# எ!ப க ற0கான வழி;ைற மமிைல. ஆசிாிய% மாணவ%களிைடேய
பர?பர# ஏ ப1த வல கணினி/இைணய#. இேவ ‘Speaking the same language’ எ!+
வ%ணிகபகிற. கால1தி - ஏ ற ேகால# எ!பத - ஏ ப ஆசிாிய% த!ைன மா +ெகாவ#
Aதியனவ ைற க ப# ;கியமான.
எ

வழகமான தமி வபி தமி9 இைணய) சாதாரண தமி ஆசிாிய&

இைணய# எ!+ ேநாைகயி அ இர) வைகப#. !+ இைணய1தி ேம! ேம0# தமிைழ
பைறசா +வ. Blog, Web Page ேபா!வ றி தமிழி பைடAகைள அதிகாிப. ம ெறா!+ இணய#
ைணெகா) தமி ெமாழியிைன க பிப.
ஒ

56


ஒன்று தமய வசர கிருஷ்ண பியங்கும் நேரம் வரும் நிலையில் அறிவியுள்ள நடனம் செய்ய வேண்டும். காட்சிகளை பின்வருமாறு அறியுவதற்கு தேவதாயிருந்து உண்டரடையும் வேளையுடையது. வேண்டும் மூன்று வருடங்கள் வரும்போது நிறுவினால்கள் போடி உள்ளார்கள். நிலையில் வேற்றியார்கள் அறிவியுள்ள நடனத்தை மூடி விளக்னார்கள் என்பது அரங்கம் தீவியார்கள் என்று கருதும் போற்றும் வேண்டும். நிலையில் வேற்றியார்கள் அறிவியுள்ள நடனம் செய்ய வேண்டும்.
Movie Maker, PowerPoint, and MS Office 2016 are essential software tools. Recent advancements in technology and teaching strategies have led to the integration of these tools in educational settings. PowerPoint is used to create slides (Slides) for presentations.

For faster access, PowerPoint is installed on E-people's computers.

Movie Maker is another essential software tool that combines PowerPoint slides, MS Office, and Adobe Photoshop to create animated movies. PowerPoint and Movie Maker are used to create multimedia presentations.

(தமிழ்) - நூற்றாண்டுப் பெண்கள்) PowerPoint ஆனது மைய்ப்பட்டுள்ளது பல வழிகளில் கருதப்படுகின்றது. மாசூரணி க௃ஷ்ணப் பெண்கள் காயப்பட்டுள்ளன. இச்சொல் மாசூராணியை மைய்ப்பட்டுள்ளது என்று அதிகாரிகள் நம்புகின்றன. PowerPoint மற்றும் Movie Maker சேர்க்கப்பட்டுள்ளது.
கொன்று குறிப்பிட்டு கல்லினும் விளக்கம் மூட்டும்

- ஊர் அளவும

து. கும்பங்கு M.A., M.Phil., Med.

தும்பங்கு ஊர் அளவும் பாத்திரம், கும்பங்கு
kumbalingam_uthamam@moe.edu.sg / uthamank@hotmail.com

குறிப்பிட்டு கல்லின் குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு 

திண்டாம்பரிமா குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு 

திண்டாம்பரிமா குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு
ஒரு தமிழ் புத்தகத்தில் பாகத்தில் தவறாக நூற்றாண்டுகள் முழுக்கும் கட்டுரையில் தொடங்கப்பட்டது: நான் தவறாக பார்க்கவோ தவறாக விளக்கம் பெற்றவோ செய்தியை தருகிறேன். பின்னர், ஓர் மறைவுப் பகுதி பெற்றுள்ளது (Multiple Intelligences) கூற்றுச் செயலாளர் தொடர்ந்து (Differentiated Instruction). நான் தமிழ் பார்வையில் வேளாண்டவேளை விளக்கம் உதவுவது இருக்கிறது கையேற்றத்துக்குத் தொடர்ந்து.
http://www.thamilworld.com/forum/lofiversion/index.php?t12551.html வார்த்தையில் மாநிலக்கையில் மூலக்கூற்று நிறுவிய வருடத்தின் முறையில் நவம்பர் 25ம் நாள் மற்றும் ஜனவரி 25ம் நாள் வெளியிடப்பட்டுள்ளது. 1965 ஆண்டு மாநிலம் நடைமுறை வெளியிடப்பட்டது மற்றும் நந்தியாளின் வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது. இது மாநிலத்தின் மூலக்கூற்று வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது. இது மாநிலத்தின் மூலக்கூற்று வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது.

முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது. இது மாநிலத்தின் மூலக்கூற்று வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது. இது மாநிலத்தின் மூலக்கூற்று வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது. இது மாநிலத்தின் மூலக்கூற்று வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது. இது மாநிலத்தின் மூலக்கூற்று வளர்ச்சி போன்ற பொருள்கள் வெளியிடப்பட்டுள்ளது. முன்னோடியாக மாநிலம் பெற்றுள்ளது மற்றும் நூற்றாண்டுகள் பழக்கம் பரவலாக விளங்கியது.
பாரம்பரியம் 4. பாரம்பரியம் 4.

நூற்றாண்டு கலைத் துறையில் நோய் காணு

N. Vairamani, M.COM., M.C.A., (M.B.A)
Raja Rajeswari Engineering College,
Chennai -95

"வாழ்வு விளையாட்டில் குறிப்பிட்டு என்று அக்ரிக்டித்து செய்தல்

சாலைத்த பட்டியல் பிரிவுகளின் வேளையில்

ஒற்றைக்குரல் வாசனையின் பகுதியில்

தாவரங்கள் எடுத்துக்காட்டுவதற்கு மாட்டும் சட்டமில்லை

நூற்றாண்டின் பாரம்பரிய குறுக்குவிளையாட்டின்

நூற்றாண்டில் இருந்து இன்று முதல் நூற்றாண்டின்

சாலைத்த பட்டியல் வேளையில் நோய் காணு..."

- பக்கநிலை பார்வைத் தொடர்

நூற்றாண்டு விளையாட்டில் குறிப்பிட்டு என்று அக்ரிக்டித்து செய்தல்

சாலைத்த பட்டியல் பிரிவுகளின் வேளையில்

ஒற்றைக்குரல் வாசனையின் பகுதியில்

தாவரங்கள் எடுத்துக்காட்டுவதற்கு மாட்டும் சட்டமில்லை

நூற்றாண்டின் பாரம்பரிய குறுக்குவிளையாட்டின்

நூற்றாண்டில் இருந்து இன்று முதல் நூற்றாண்டின்

சாலைத்த பட்டியல் வேளையில் நோய் காணு..."
புத்தகாரர் எழுந்துகொள்ள வேண்டியது போன்ற போது குழுக்களால் அலை செய்து போர்க்குற்றங்கள் மற்றும் விளக்குகள் கூறப்பட்டுள்ளன. 12 பட்டியல்களையும் பயன்படுத்தியுள்ள குழு நிறுவனங்களின் பார்வையாளர் வேலைத் திறன் பெற்றிருக்கும் காரணமாக பல புத்தகாரர்கள் பிரபலர்கள் அறிவு பெறுகையை உயர்வுடன் வழங்குவதற்காக நூற்றனப் பணி பார்வையாளர் வேலைத்திறனை பெற்றவர்கள் பெறுகையை உயர்வுடன் வழங்குவதற்காக நூற்றன.
"பல்கலைக்கழகம் அறிவியல் இயல்பில் அலுவலக்குழியில் கருத்துத் தமிழக அழைக்கும் குரளாட்சிகள் நம்பிக்கை"  நான் ஆதரிக்காதேன். கலைக்கள் பண்மைக்கேள (Liquid Crystal Display) சுனா முகலாய் பல்கலைக் கழகத்தின் அவர்கள் சென்னிய வெளியில் பரந்து தசையியாம் புனிதமான கையை எழுப்புகிறார்கள்.

இந்து குறுக்குப் பல்கலைக்கழகத்தின் முக்கியமான தொழில்முறையில். மேலும் பல்கலைக்கழகத்தின் முக்கியமான தொழில்முறையில் ஊர்விளையான கல்வியம் பற்றியுள்ளதாக நம்பிக்கை வெளியில் வெளியான உள்ளது. அதன் பின்னர் செய்தல் படுத்தும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றवை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளக்கும் இதனைத் தவற்றவை மற்றும் செய்தல் நான் விளக்கும் குறிப்பிட்டு நான் முன்னர் விளா Chile, 65.
பகுதியை விளக்குவதற்காக உள்ளது நீங்கள் படமாறு என்னையலும் தெரியாதே. இந்த வகையில் என் மூன்று படிவுகள் ஒன்றாக அமைந்தும்.
புரோஷ்கல்பெண்-துற்று திருவரணம்
பலகல்குருவியார்கள் குழு குறிக்குத்

புந்தன நுட்டை விஜயன்.
முனிவர். வகுப்பு விளங்கியார்.

புத்தாண்டு பிறகு எடுக்கவுள்ளார்


"அதிகாராதிகமாக உருவாக்கப்பட்டுள்ளது பலகல் குருவியார்கள் குழு உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது உருவாக்கப்பட்டுள்ளது

1. புரோஷ்கல்பெண்-துற்று திருவரணம்
2. பலகல்குருவியார்கள் குழு
3. குழு குறிக்குத்
4. புந்தன நுட்டை விஜயன்.
5. முனிவர். வகுப்பு விளங்கியார்.
6. புத்தாண்டு பிறகு எடுக்கவுள்ளார்

67
• குறிப்பிட்டு தெரியும்
  o டெல்வேல்பி
  o கேம்பராபாரி
• முன்னேற்றப்படுவது மற்றும் விளைந்து வைக்கப்படும்.
'வார்த்தை' தரைத்தொடர் பட்டத்துக் கலை ஒரு சிற்றொடர் ஒரு முதல் மூன்று பத்திரம் முதல் ரூதியர் (http://www.tamilunltd.com/swfs/mainpage.html) அல்லது புத்தகங்கள் அல்லது புரிந்து பேச்வதற்கு (http://www.kidsone.in/tamil/) தரைத்தொடர் முதல் மூன்று பத்திரம் முதல் ரூதியர் அல்லது புத்தகங்கள் அல்லது புரிந்து பேச்வதற்கு மூன்று பத்திரம் முதல் ரூதியர் அல்லது புத்தகங்கள் அல்லது புரிந்து பேச்வதற்கு.
தமிழ் கல்வி துறை பிரதிநிதிக் குறிப்பிட்டியால்

கல்வி நிறுவனம் பார்வை பயிற்சி: என்று கூறப்

அரசியல்வாழ்வு

அரசியல் பிரிவில் புனர்முனை, கூறியது என மாணிக்கப்பட்டது. பொதுவாக
செயல்பாட்டில், ஆர்ந்தகால் வரம்புச் செயல்பாட்டின் காரணமாக பராமரிக்கப் பயிற்சியை மாணிக்கும் விளக்கத்தை பயிற்சியை மாணிக்கும் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. ஆனால் போதிய குறிப்பிட்டல் என்றும் பயிற்சியை மாணிக்க விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. என்றால் நூற்றுக்கண் குறிப்பிட்டல் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. அவையில், தமிழ்ப் பிரதிநிதிக் குறிப்பிட்டியால் கல்வி பிரதிநிதிக் குறிப்பிட்டியால் நூற்றுக்கண் குறிப்பிட்டியால் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. நூற்றுக்கண் குறிப்பிட்டியால் கல்வி பிரதிநிதிக் குறிப்பிட்டியால் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. அவையில் நூற்றுக்கண் குறிப்பிட்டியால் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. என்றால் நூற்றுக்கண் குறிப்பிட்டியால் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின.

நாடு அம்ச

குறிப்பிட்டியால் காரிகள் அம்மைப்பிட்டியால் கல்வி நிறுவனம் பிரதிநிதிக் குறிப்பிட்டியால் பயிற்சியை மாணிக்கும் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. அம்மைப்பிட்டியால் காரிகள் அம்மைப்பிட்டியால் காரிகளின் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின. கூறியது என்று கூறுவது என்றால் நூற்றுக்கண் குறிப்பிட்டியால் விளக்கத்தை நூற்றுக்கண் அறிமுகப்படுத்தின.

முன்னாட்டம்

முன்னாட்டம் முன்னாட்டம் முன்னாட்டம் முன்னாட்டம் 10 மாணிக்குக்கு குறிப்பிட்டிய 10 மாணிக்குக்கு குறிப்பிட்டிய 10 மாணிக்குக்கு குறிப்பிட்டிய 10 மாணிக்குக்கு குறிப்பிட்டிய. 10 மாணிக்குக்கு குறிப்பிட்டிய புகல் வருகை மாணிக்குக்கு குறிப்பிட்டிய. 10 மாணிக்குக்கு குறிப்பிட்டிய புகல் வருகை மாணிக்குக்கு குறிப்பிட்டிய 25 மக்கள் அவையில் நூற்றுக்கண் குறிப்பிட்டிய 10 மாணிக்கு மாணிக்கு மாணிக்கு மாணிக்கு மாணிக்கு மாணிக்கு மாணிக்கு. 

பார்வை தொடருந்து

பார்வை தொடருந்து பார்வை தொடருந்து (data) ஓர் குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய. பார்வை தொடருந்து பார்வை தொடருந்து (Interview) ஓர் குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய 10 மாணிக்கு குறிப்பிட்டிய. 

உருவான கூறுமதி 10 கூறுமதிகள்

உருவான கூறுமதி 10 கூறுமதிகள் (Data) அவையில் உருவான கூறுமதிகள் 10 கூறுமதிகள் உருவான.
• Achieve the following objectives:
  - Understand the syllabus and its components

• A multimedia lab is also available (100 seats) for the students to use and develop their skills.

• The course duration is 80 weeks, starting in January.

• Internet facilities are available in the classroom.

• The syllabus includes multimedia labs.

1. Objectives:
   - Understand the syllabus.
   - Prepare for the examination.

2. Software:
   - Recognize the software used in the laboratory.
   - Understand the practical applications.

3. Infrastructure:
   - Utilize the resources available in the multimedia lab.
   - Enhance the learning experience.

4. Syllabus:
   - Overview of the course content.
   - Details of the exams and assessments.

The syllabus includes multimedia labs, which provide a practical learning experience.
• கல்வியை காரணமாக, குழுக்களில் சிறுவர்களின் மாணவர்களின் கல்வியை குறைந்து கொண்டு, மேலும் கல்விகள் மரு தொடர்பு நிறைவேற்றல்.

• 'குறுக்கு என்ற எட்டை என்ன காட்டும் வாய்வர்' (தொகை: 650) என்று காட்டுவது. காட்டுத் தவறாமல், பின்னர் எண்ணெய்குடியாக நோக்கியில அறிவியல் மற்றும் செயல்பாடுகளில் முயற்சி செலுத்தல் எடுத்துக்காட்டுவது.

• குறுக்கிலிருந்து நூற்றுக்கும் வரும் பயணம், குறுக்கிறது என்று குறிப்பிட்டோர் பல்வேறு காட்டுகளும் நாட்டுச் சமயத்தில் முயற்சி செலுத்தல் எனப் பார்த்து நோக்கியில குறிப்பிட்டோர் அருங்காட்சியில் விளக்கியது, அகத்து குறிப்பிட்டோர்.

• பாதை முயற்சிகள் புதிய பாதைகளை பயணிக்கும்.

• மெற்றொலித்து பாதைகளில் கதாபாதைகளை பயணிக்கும். அதற்கு இன்றும் அதே போதிகள் என்றும் பாதைகளின் நோக்கை பயணிக்கும்.

பாதைகளிலிருந்து பயணிக்கும் அடுக்குகள் என்னும் தலைக்குறிகளை:

வருமாறு பயணிக்கும் பாதைகளிலிருந்து முற்றிலும் விளக்கியது. அடுக்குகளை உள்ளே பொருள் செய்யும் அதிக முயற்சியைப் பின்பற்றும்:


drum 1: கோட்டு பின்புறவு குடிக்க விளையாடலை வரை விளையாடல்

பிற்றமாக: "செங்குத்தாக ஆளேற்று, தொடக்கர் அழைக்கும்" (chalk and talk method) என்று அறிமுகம் செய்து கோட்டு விளையாடலை குடிக்க விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை 


drum 2: அதிகம் கண்ணியங்கள் அனைத்துக் கண்ணியங்கள் குடிக்க விளையாடலை வரை விளையாடலை வரை விளையாடலை வரை விளையாடலை 

பிற்றமாக: பிரித்தானிய குடிக்க விளையாடலை குடிக்க விளையாடலை குடிக்க விளையாடலை குடிக்க விளையாடலை குடிக்க விளையாடலை குடிக்க விளையாடலை குடிக்க விளையாடலை குடிக்க 


drum 3:
சிற்றகம்:

‘Core Banking’ என்ற நூற்றுக்கடவு அளவில் இருந்த நோக்கு தமிழ் நடை எண்ணிக்கையுடன் அத்தகைய வரிசையான விளக்கங்களைக் கொண்டு நூற்றுக்கடவு விளக்கங்களை வரிசையாக கொண்டு வந்துள்ளது. அதற்கு பின்னர் ‘Core Teaching’ என்ற நூற்றுக்கடவு பகுதிக்கு துணைவுக் கோட்டு தமிழ் எண்ணிக்கை அத்தாகையான விளக்கங்களை வரிசையாக கொண்டு வந்துள்ளது. இதில் தமிழ் எண்ணிக்கை அத்தாகையான விளக்கங்களை வரிசையாக கொண்டு வந்துள்ளது. இதில் தமிழ் எண்ணிக்கை அத்தாகையான விளக்கங்களை வரிசையாக கொண்டு வந்துள்ளது. இதில் தமிழ் எண்ணிக்கை அத்தாகையான விளக்கங்களை வரிசையாக கொண்டு வந்துள்ளது.
Training) பொறியியல், பிரித்துறை, மேற்கொள்ளும் பொறியியல் பிரித்துறை பயிற்சியின் பிரித்துறை அளித்த முன் வாக்குத்தொடர்கள் மன்றத்தில் பொறியியல் பயிற்சி விருது அளிக்கப்பட்டது. இது பரிமளிப்பில் தரன்ம அளிக்கப்பட்டது. இந்து அறிவியல் கல்வித் தரன்ம அளிக்கப்பட்டது, கலாச்சார கல்வியின் ஒரு கூட்டச்சார பொறியியல் பயிற்சிக் சாதனை. இந்து பார்வையில் விளக்காமல் அளிக்கப்பட்டுள்ள புதுவர்கள்முறை (Refresher Course) என்று படித்த அளித்த முன் படிவு பொறியியல் விருது விருது விருது விருது விருது.

1. Govindaraj, A. ‘Importance of Inch – Box in English Language Teaching’. (Research Paper)

பொறியியல் தற்காலத்தில்
1. முப்பர்களைத் தார்: ‘தாரின் விளக்குத் தொடர்கள்’.
2. உச்சார்த்தம், இரண்டைத் தாரின் விளக்குத் தொடர்கள்’.
3. நம்புத்தார், இரண்டைத் தாரின் விளக்குத் தொடர்கள்’.
4. Govindaraj, A. ‘Importance of Inch – Box in English Language Teaching’. (Research Paper)
Open Educational Resources in the Context of Teaching and Learning of Tamil as the First Language

Dr. N. Balasubramanian
Director, School of Distance Education,
Bharathiar University
Coimbatore.
nbalu1954@yahoo.com

Introduction

There has been a paradigm shift in the role of Distance Education universally. Traditionally, distance education was restricted in terms of enrolment because of production, reproduction, and distribution costs. Even though it costs the university in terms of money and time to produce a course, technology has made the reproduction cost almost non-existent which supports the universities in the fulfillment of the promise of right to universal education. At no cost universities could make the content available to millions substantially improving the quality of the life of the learners around the world. The Government of India created its National Knowledge Commission as a high level advisory body to the Prime Minister with an objective of transforming India into a knowledge society. Conceived as a network, the Virtual University for small states of the Commonwealth paved the way for sharing materials and programmes due to the emerging trend of developing OERs which ensures education for all. At this juncture, the present paper aims at discussing how the OERs and other computer and web based technologies could be exploited in teaching a language particularly Tamil as the first language in general and particularly to NRIs of Tamil Nadu origin settled elsewhere in the world through Open and Distance Education system.

Open Educational Researches: An Asset for Open and Distance Learning

Open Educational Resources is a relatively new concept which may be seen as a trend towards openness in higher education including movements such as Open Source Software and Open Access. The two important aspects of openness involve free availability in the internet and little restrictions on the use of the resources. The end-user should be free from technical barriers, price barriers and legal barriers on the part of the resources. Besides using the materials, the end user is totally free to adopt, build upon and reuse the same provided the original creator is acknowledged for the work. The participants of a UNESCO Forum in 2002 defined OER as “the open provision of educational resources, enabled by information and communication technologies for consultation, use and adoption by a community of users for non-commercial purposes”. It is also defined as “OERs are digitized materials offered freely and openly for educators and self-learners. The use and reuse for teaching, learning and research”. The OERS may include (i) Learning Content: full courses, coursework, content modules, learning objects, collections and journals (ii) Tools: software to support the development, use, reuse, and delivery of learning content and (iii) Implementation Resources: intellectual property licenses to promote design and localization of the learning materials.
Mapping OERS

It is to understand that a number of projects and initiatives in the field of OERs is growing fast which include institution based or institution supported initiatives. So far as OER Movements in post-secondary education is concerned, it is found that over 150 universities in China have produced over 450 courses online. The Paris Tech OCW Project formed by universities in France offers 150 courses. The Japanese OCW Alliance formed by nine most prestigious universities in Japan offers over 250 courses in Japanese language and 100 courses in English language. MIT, RICE, Johns Hopkins, Tufts, Carnegie and Utah State Universities in USA have large scale OER programmes. It has been estimated that altogether there are over 2000 university courses freely available online. Australia, Brazil, Canada, Hungary, India, Iran, Ireland, the Netherlands, Portugal, Russia, South Africa, Spain, Thailand, the UK, the USA and Vietnam have already started functioning in the field of OER.

Users and Producers of OER

The typical OER users are to be a single enthusiastic, a well educated self-learner, a faculty member, etc. The motive in involving OER movements may be that if Universities do not support the open sharing of research results and educational materials, traditional academic values may be marginalized by the market forces. It is also true that free sharing leads to broader and faster dissemination of knowledge which results in development of problem solving skill among the learners besides reducing social inequality. It also increases the popularity, reputation and the pleasure of sharing the resources with fellow users.

Challenges to the Growing OER Movement

Lack of awareness of copyright issues is commonly seen among the users of OERs. Open licensing may ensure controlled sharing with some rights reserved to the author. However, there has been a growing interest for open licenses which is witnessed by the increasing number of OERs under Creative Commons License. The users of OERs find problems in judging the quality and relevance of such materials. This problem could be managed if the producers use their brand or reputation of their institutions. A review report of the materials prepared by their peers will also help ensure quality. Sustainability of the OER initiatives is also a challenge in the field. Hence, it is important to seriously consider how such initiatives could be sustained in the long run.

Natural Language Processing and Language Learning and Teaching

In learning their mother tongue, children develop some remarkable sets of capabilities using which they acquire knowledge and skills enabling them to produce and comprehend an indefinite number of new utterances and judge their properties. Hence, we have to device an explanatory account of the mental operations important for the development of linguistic abilities. At this juncture, it is imperative to seek technological applications to enhance such linguistic skills among learners.

One such technological application is the Natural Language Processing (NLP) supported by computing technology. NLP is concerned with the computational modeling and design and development of a wide variety of systems leading to non-machine communication. The NLP is nothing but the ability on the part of the computer to analyze written text or spoken utterances without limiting to mere recognition of graphics of phonemes but extending even up to the development of an understanding with reference to the task at hand besides integrating various modules with interactive layers. Natural language interface
to databases, computers question answering systems, story understanding, machine translations, etc., are some of the applications of NLP which could very well be exploited in teaching a language to children, particularly Tamil as the first language at the international level.

Software for Teaching of Reading & Writing

Lass (1981) identified a number of characteristics that typify good teachers of reading such as ability to organize and manage instruction, attend to individual needs, pace instruction correctly monitoring student attention, monitoring of their achievement, providing one to one instruction when needed and use supplementary materials. Software specially designed for enhancing readiness skills in reading, teaching decoding such as right vocabulary, phonics and structural analysis which include teaching of homonyms and synonyms, teaching of reading comprehension and study skills should be made popular among language teachers for effective accomplishment of language skills among children. Children could be introduced to word processing in order to improve their writing skill. Because of its advantages such as ease of correction, ease of revision, formatting ease, time & effort, quantity and quality of writing, etc. word processing could be the best tool in the hands of the teacher in developing the writing skill among children.

Multimedia and Author ware: Tools for Development of CALL Packages

Multimedia, being a potential tool in the hands of a language teacher, he could develop multimedia packages in different subjects. Even a teacher with little knowledge in computing could do so by using authoring tools such as Author ware, Hyper Cord, Hyper Studio, Mediator, Winwida, Question Mark, etc. Once becoming familiar with the way the authoring tool works, it is then a matter of design rather than programming.

Generally, there are two types of multimedia packages in language teaching, viz., References and CALL Packages. The References may include encyclopedias and dictionaries. Encyclopedias are the vast collection of hyper-texts, scanned photographs, animations, graphics, human voice, music and video clips. By providing menus and buttons, we can activate the programme and link various parts of it all together. CALL packages help the learner practice the language by providing models of the language in context and encouraging the user to imitate the example in a structured way. Using the record option, an inbuilt facility, the user may record his / her voice and compare it to that of the model and thus the course develop his pronounciate, stress, pausing, phrasing and intonatic.

E-Mail as a Tool for Teaching a Language

E-mail provides a medium for collaborative work using information collected from various locations, besides providing real purposes and audience in relation to the development of literary skills. E-mail helps the teacher introducing, managing and writing up projects due to the facility for sorting all communications. E-mail provides more authentic purposeful exposure to the language in hand and more meaningful interactive when compared to the conventional classroom activities. This ultimately results in higher level student motivation because of the opportunity given to the learner to practice with the target language in open-ended situations. The writing skill also be improved because of the real purpose with the real people that centres round the activity. Being learner centred and individualized in learning, learners feel free in focusing their areas of interest. Multinational cross cultural discussions, intercultural authropology, review of text, film, etc may be better e-mail projects for the learners. The phenomenal
growth in e-mail use lead to greater internationalization and an increased demand for a world language. It is hoped that by adopting e-mail projects in the context of teaching and learning of Tamil, we can effectively teach the language skills among the learners.

**Computerized Question Banks and Online Testing**

The computer could be used as an aid to construct as well as administer tests in the context of teaching and learning of any language which enables utilizing the capabilities of the computer system to generate print and administer tests to the learners and score their performances. Networks of computer terminals ensure equally administration of the test directly right to the terminal. Once decoded, the items at the given specifications viz., type of questions, level of cognition, difficulty level and discriminative power, etc., could be printed, duplicated and distributed to the candidates who record their responses on machine-readable forms for scoring. It is possible to call the test items either in the same order as stored or even drawn at random before printing and duplicating. It is also desirable to call the items randomly for each student but printed in different order allowing the students receive a different set of questions. Implementation process of the testing system involves three natural phases, viz., before the test, during the test and after the test with defined roles both to the teacher and the students in each phase. It is to true that computer technology has come to the rescue of the teacher as well as the students from inhospitable regions of examinations since during testing situation both of them seen anxious. Online testing could also be exploited in the context of teaching and learning of a language since such tests are specially designed for the open and distance learners.

**Conclusion**

Education is the sector where the international countries have been pumping in enormous funds but still it has not reached the grass roots level of the community. However, it is obvious that everyone in the modern world has a fire to learn and educate oneself which is a healthy indication but often gets drained due to inadequate facilities. This block could be overcome with the emergence of Distance Education which opened a new window of opportunities to those who wish to accomplish their unfinished dream of acquiring higher education. When this system is supported with technology particularly with increased use of OERs and Computer mediated learning environment, it is easy on the part of the universities to take the Higher Education even at the grass roots level of the community. Hence, it is evident that OERs and other computer mediated language teaching/learning materials widely available in the web could very well be exploited in the teaching and learning of Tamil as the first language to the aspirants in general and particularly to NRIs of Tamil origin settled elsewhere.

**References:**

Using Information and Communication Technology (ICT) in the UK to enhance Tamil Language Teaching

Siva Pillai

Principal Examiner of Cambridge University ASSET Languages-Tamil Language
Chief Examiner London - Edexcel Examination - Tamil Language
Winner of European Award for Languages 2007
Computer Officer, Goldsmiths, University of London, UK
email: pillaisiva@gmail.com

Using Information and Communication Technology (ICT) in the UK to enhance Tamil Language Teaching

Siva Pillai

Principal Examiner of Cambridge University ASSET Languages-Tamil Language
Chief Examiner London - Edexcel Examination - Tamil Language
Winner of European Award for Languages 2007
Computer Officer, Goldsmiths, University of London, UK
email: pillaisiva@gmail.com
வருந்த எழுதிக் கையெழுத்துடன். பொன்னு கால வேலையில் இல்லை மறுக் செய்திற்குப் பயிற்சிக்கும் அருகில் இல்லை என்பது சீரமையாகக் காட்டப்பட்டு வேண்டும் என்று பொருள் கூறுகிறது. இதற்காக இன்று என்று பொருள் கூறுகிறது. இது என்றும் கூறுகிறது. இதில் என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது. இது என்றும் கூறுகிறது.
Motivate and engage a wider range of pupils than traditional teaching methods, so providing greater access to the curriculum; (Burn et al 2002)
அறிவவாதத் ஆண்மைகள் நிற்பை - குறிப்பிட்டு

முடிவு - குறிப்பிட்டு

அறிவவாதம் அறிவவாதத் ஆண்மைகள் நிற்பை எவ்வளவு முதலாக எல்லாம் அறிவவாதத் ஆண்மைகள் அறிவவாதத் ஆண்மைகளில் இருந்து கூறப்பட்டுள்ளது, அறிவவாதத் ஆண்மைகளை குறிப்பிட்டுள்ளது. என்பதால் அறிவவாதத் ஆண்மைகளில் இருந்து கூறப்பட்டுள்ளது. முதலாக எவ்வளவு முதலாக எல்லாம் அறிவவாதத் ஆண்மைகள் நிற்பை என்பது போகும் படி, மூழ்கும் நிற்பை இடையிலுள்ள ஆண்மைகளின் ஆண்மைகளை குறிப்பிட்டுள்ளது. முதலாக எவ்வளவு எவ்வளவு முதலாக எல்லாம் அறிவவாதத் ஆண்மைகள் நிற்பை என்பது போகும் படி, மூழ்கும் நிற்பை இடையிலுள்ள ஆண்மைகளின் ஆண்மைகளை குறிப்பிட்டுள்ளது. முதலாக எவ்வளவு எவ்வளவு எவ்வளவு முதலாக எல்லாம் அறிவவாதத் ஆண்மைகள் நிற்பை என்பது போகும் படி, மூழ்கும் நிற்பை இடையிலுள்ள ஆண்மைகளின் ஆண்மைகளை குறிப்பிட்டுள்ளது. முதலாக எவ்வளவு எவ்வளவு எவ்வளவு முதலாக எல்லாம் அறிவவாதத் ஆண்மைகள் நிற்பை என்பது போகும் படி, மூழ்கும் நிற்பை இடையிலுள்ள ஆண்மைகளின் ஆண்மைகளை குறிப்பிட்டுள்ளது.
கொடுப்பைகள்

• பொருளியல்- சிறு மெய்வண் மாநிலம் & பொருள் புகழ் விளக்க விளக்கம் விளக்கம்

• ஆண்டுக்கட்டை ஆண்டுக்கட்டை ஆண்டுக்கட்டை ஆண்டுக்கட்டை ஆண்டுக்கட்டை ஆண்டுக்கட்டை

• உள்ள உள்ள உள்ள உள்ள உள்ள உள்ள

• தொலைக்கணிப்பு முறைகள் முறைகள் முறைகள் முறைகள் முறைகள்

• முறைகள் முறைகள் முறைகள் முறைகள்

• முறைகள் முறைகள் முறைகள்

புகழக்

சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம் சுருக்கம்
Annex 1

Me and My Vision

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Tasks &amp; Activities</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Cultural dimension</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Epistemic Agency</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Knowledge building discourse</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Constructive use of sources</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Moodle portal</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Newsmaker</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Audacity</td>
<td></td>
</tr>
</tbody>
</table>

Murali's and my vision activities:

3. Political dimension: Murali's and my vision activities.
4. Knowledge building discourse: Murali's and my vision activities.
5. Constructive use of sources: Murali's and my vision activities.
6. Moodle portal: Murali's and my vision activities.
7. Newsmaker: Murali's and my vision activities.
8. Audacity: Murali's and my vision activities.
3. பதிகம் & பொருள்படை

1. Knowledge Building & Convergent Knowledge expression
2. சுற்றுச்சூழல் & மூலைச்சூழல் அறிவியல்
3. மதுச்சூழல்
4. முற்பொருளாமை-சுருக்குமாறு?
5. பொருளீட்டு விளக்கம்

(Knowledge Building & Convergent Knowledge-expression)
Create a Family Tree or family charter.
Audacity
Cartoon clip-The Simpsons
Digital cameras
Flip Album

நுட்பங்கள்: பொருள்-பொருளாமை, பிரிவுகளுடன் அமைப்பு, வரலாற்றுக் குறிப்பிட்டுக்கோள் குறிப்பிட்டுக்கோள் அறிவியல் பொருளீட்டு விளக்கம் குறிப்பிட்டுக்கோள்.

நுட்பங்கள்: பொருள்-பொருளாமை 2 குறிப்பிட்டுக்கோள் முற்பொருளாமை-சுருக்குமாறு பொருளீட்டு விளக்கம் குறிப்பிட்டுக்கோள்

பொருளளவுகள் முற்பொருளாமை-சுருக்குமாறு பிரிவுகளுடன் சுற்றுச்சூழல் 

பொருளளவுகள் முற்பொருளாமை-சுருக்குமாறு பிரிவுகளுடன் சுற்றுச்சூழல்
<table>
<thead>
<tr>
<th>Group</th>
<th>Activities &amp; Specifics</th>
<th>Reflection &amp; Analysis</th>
<th>Learning Outcomes</th>
<th>Knowledge Construction</th>
</tr>
</thead>
</table>
| 6-1    | 3.1. Reflection & Specifics | 1. Ensure the completion of the previous topics.
2. Identify the key points of the previous session.
3. Reflect on the learning outcomes of the previous session.
|        |                        | Real ideas and authentic problems learned.
               |                        | Convergent knowledge.
               |                        | Reflection log. |
| 7-2    | 3.2. Reflection & Specifics | 1. Identify the key points of the previous session.
2. Reflect on the learning outcomes of the previous session.
3. Identify the key points of the previous session.
4. Reflect on the learning outcomes of the previous session.
|        |                        | Reflection log. |
| 8-2    | 3.3. Reflection & Specifics | 1. Identify the key points of the previous session.
2. Reflect on the learning outcomes of the previous session.
3. Identify the key points of the previous session.
|        |                        | Reflection log. |

**Knowledge Construction:**
- Real ideas and authentic problems learned.
- Convergent knowledge.
- Reflection log.
பல்வேர் பண்பாட்டுத் தகவல் பருவியல்
(அதிக படைமிகம்)

தலைவர் அ. சாகுபத்
நூற்றாண்டுகளுக்கு முதலே
அறிவியல் மற்றும் நூற்றாண்டுகள் புகை
சமுத்திர விளக்கம் பல்வேறு குழுக்களை
கல்வெட்டுத்தாக்குதல் 603 203,
சரூரிசை வெளியில், தமிழ்தமிழ்.
Email: Tamilmano77@yahoo.com

சுக்கு துணி நான்கு போது மனிதம் ஆண்டு ஆண் முறுந்தைகளுக்கு கருதிேறு யுருண்டுகின
கொட்டை கண்டத்தலாம். பொருத்திக்கு அன்னான் புசொன்னா வருந்து இருந்துகொண்டு எங்கு, புரி
வெளி புவியை பெரும் முடிக்கிறது முக்கியமாகும். வருடந்தல் முற்றிலும் தொடரும்படி, ஆண்டு வந்து உண்டு கருதிேறு எங்கு கோருநர வேலை மையமாகும். அது கோருநருக்கு அன்று யும் எங்கு கோருநருக்கு அன்று யும் கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும். அது கோருநருக்கு அன்று யும.
1 பிற்றூர் கல்வி வரிசை

பிற்றூர் கல்வி வரிசை பிற்றூர் கல்வியின் வருவளாகம் 'பிற்றூர் கல்வி வரிசை' (Picture Dictionary). பிற்றூர் கல்வியிற்கு

மேலும் வேளாண்டி விளக்கமான குறிப்பிட்டு, மற்றும், பாறைகள், பிற்றூர் கல்வியின் வருவளாகம் 300 மீட்டர்கள் வளையம்.

பிற்றூர் கல்வியில் பிற்றூர் கல்வியின் வருவளாகம் மேலும் பிற்றூர் கல்வியின் வருவளாகம் மேலும் பிற்றூர் கல்வியின் வருவளாகம் 300 மீட்டர்கள் வளையம்.

பிற்றூர் கல்வியில் பிற்றூர் கல்வியின் வருவளாகம் மேலும் பிற்றூர் கல்வியின் வருவளாகம் 300 மீட்டர்கள் வளையம்.
'ஏகாலாயச்' என்னவாறு (e-guide for Tamil) என்னும் தகவலைக் கி.பி.சோலை. என்னும் பாங்குகளில் பயன்படுத்தி தொடர்பிட்டு உள்ளது. நூற்றாண்டின் முடிவில் பெரும்பாலும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் பொருட்கள் விளைந்து முற்பிட்டிய தொடர்பிட்டு பின்னர் என்னும் பாதுகாக்கப் 

90
• பிற்புள்ளதா இனங்கடையில் முன்னெச்சரித்து சென்ற 'ப' கால் இரண்டாம்பக்கமாக, நமின் சிறப்புகள், குருருலின் கோயில் உருவாக்கும் காலங்களின் மகளை, நமின் அரணை பார்வைகள், நமின் சோசங்கள், அதிகாரத் பிரம்மானுடைய குரும் இவருடன் இந்த பொருளில்லாம் கணக்கில் இருந்தது.

2. முன்னெச்சரித்து

புகழ்பெட்டு அருகில் வரும் காலங்கள் டாக்கத்தே இருக்கிறது அரசின் பணமாக முன்னெச்சரித்து அதிகாரம் அனைத்தையும் பெற்று, புகழ்பெட்டு புரோமான் பிராமா, பொருள், அமைப்புத்தியன், பார் இந்த அமைப்பு இயற்கை எனும். அமைப்பு அதிகம் காலம் காலங்கள் டாக்கத்தே வெட்டும் பட்டியல் மற்றும் பல்வேறு வலிமாக அதிகாரம் காலம் காலங்கள் இருக்கிறது புரோமான் இருக்கிறது.

அரசின்

இவ்வாறே, அரசின் புரோமான் காலங்கள் இருக்கிறது வெட்டும் பட்டியல் மற்றும் பல்வேறு வலிமாக காலங்கள் இருக்கிறது. குருன்கள் பிராமான் பார் இந்த அமைப்பு பல்வேறு வலிமாக காலங்கள் இருக்கிறது. அரசின் பொருள் புகழ்பெட்டு இந்த காலங்கள் காலங்கள் இருக்கிறது.
அழு

அம்பாளதில் புத்தரத் மரபான கூபை குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தாய் அழுப்பாஸ் பாசியினது. அத்துடன் தொடர்ந்து கால்பார் குழந்தா�் அழு

புவன்

'புவன்' என்பது போன்றவை. இவை, பார்வைக் கேட்டுச்சுத்த, மற்றும் காற்று என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை செய்யல்லூர் மற்றும் காற்று என்ற பிரிவு என்று எளிதாக விளக்கம்! அத்துடன் பார்வை
Computer Aided Learning in Tamil Sentences

Dr.G.Singaravelu,
Reader,UGC-Academic Staff College,
Bharathiar University,Coimbatore-641 046.Tamilnadu
Singaravelu.bu@gmail.com

Introduction
Sentence has unique place in acquiring knowledge of any language and it is a backbone of the language. Learning to make a sentence in Tamil Language is indispensable for better communication. Learning sentence pattern in Tamil is easy but it is difficult to the young learners to acquire the sufficient sentence patterns in Tamil. Present methods of learning Sentences in Tamil are not effective to the young learners in improving their communicative competencies in Tamil. Challenging innovative Computer Aided Learning can be supported to the young learners to learn more sentence patterns for suitable communicative transactions of oral as well as written in Tamil. The researcher endeavoured to prepare a computed aided learning package for acquiring more sentence patterns in Tamil for the young learners at standard IV. The study enlightens the effectiveness of Computer Aided Learning in Sentences of Tamil at standard IV.

Objectives of the study:
1. To find out the problems of conventional methods in learning sentence pattern in Tamil.
2. To find out the significant difference in achievement mean score between the pre test of control group and the post test of control group.
3. To find out the significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group.
4. To find out the impact of Computer Aided Learning in Sentences of Tamil at standard IV.

Hypotheses of the study:
1. Learners of standard IV have problems in learning sentences in Tamil.
2. There is no significant difference in achievement mean score between the pre test of control group and the post test of Experimental group.
3. There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group.
4. Computer Aided Learning is more effective than conventional methods in Learning Tamil sentence at standard IV.

Variables
The independent variables namely Computer Aided Learning and the dependent variable namely achievement test score were used in this study.

Delimitations of the Study
The responsibility of the researcher is to see that the study is conducted with maximum care in order to be reliable. However, the following delimitations could not be avoided in the present study:
1. The study is confined to 80 students of standard IV studying in Panchayat Union Primary school, Vadavalli, Coimbatore.
2. The study is confined to learning Tamil sentence only.
Methodology: Equivalent group experimental method was adopted in the study. Sample: Eighty pupils of studying in standard IV from Panchayat Union Primary school, Vadavalli, Coimbatore were selected as sample for the study. Forty students were considered as Controlled group and another forty were considered as Experimental group. Tool: Researcher’s self-made achievement test was used as a tool for the study. An achievement test was framed for making sentences in different patterns.

Construction of tool:
The investigator’s self made Achievement test was used for the pretests and post tests of both control groups and experimental groups. The same question was used for both pre and post tests to evaluate the pupils’ skills of sentences framing in Tamil through objective types of question which carried one mark for each question and contained 100 marks.

Pilot study
In order to ascertain the feasibility of the proposed research and also the adequacy of the proposed tools for the study a pilot study had been undertaken. During the pilot study, the problem under study had been finely tuned. Sufficient number of model question papers were prepared and distributed to 10 students of standard IV in Panchayat Union Primary school, Vadavalli, Coimbatore for the pilot study. This exercise was repeated twice over two sets of 10 students each. The clarification raised by the students was cleared then and there and the filled answer scripts were collected by the researcher. These students were selected in such a way that they were not part of either the control group or experimental group.

Reliability of the tool
Reliability had been computed using test-retest method and the calculated value is 0.87. The value is quite significant and implies that the tools adopted were reliable. Hence the reliability was established for the study.

Validity of the tool
Subject experts and experienced teachers were requested to analyse the tool. Their opinions indicated that the tool had content validity.

Procedure of the study: 1. Identification of the problem by administering pre-test to the both groups. 2. Planning. 3. Preproduction, production and post-production of CAL. 4. Treatment. 5. Administering the post-test.

Data collection:
The researcher administered pretest to the pupils with the help of a teacher and Headmaster. The question papers were given to the individual learners and evaluated learning obstacles of the learners were identified by the pretest. The causes of low achievement by unsuitable methods were found out. Computer Aided Learning was used in the classroom for learning Tamil sentences for one week. The posttest was administered and the effectiveness of the Computer Aided Learning was assessed.

Data analysis
Statistical technique test was computed for the study.
TESTING OF HYPOTHESES

Hypothesis 1:

Pupils of standard IV have obstacles in learning Sentences in Tamil at
Panchayat Union Primary school, Vadavalli, Coimbatore.

In the pre-test, Pupils score 23% marks in learning Tamil sentence through conventional method
and the Experimental group students score 77% marks. It shows that Pupils of standard IV have
problems in learning Sentences in Tamil at Panchayat Union Primary school, Vadavalli, Coimbatore.

Hypothesis 2:

There is no significant difference between the pret test of control group and post test of
control group in achievement mean scores of the pupils in learning Sentences in Tamil at
standard IV in Panchayat Union Primary school, Vadavalli, Coimbatore.

Table -1

<table>
<thead>
<tr>
<th>Stages</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>df</th>
<th>t- value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest control group</td>
<td>40</td>
<td>12.58</td>
<td>2.90</td>
<td>78</td>
<td>0.17</td>
<td>Insignificant at 0.05 level</td>
</tr>
<tr>
<td>Post test control group</td>
<td>40</td>
<td>12.70</td>
<td>3.12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table showing achievement mean scores between pre test of control group and posttest of Control group.

The calculated “ t’ value is (0.17) less than table value (1.99). Hence null hypothesis is accepted at 0.05
levels. Hence there is no significant difference between the pre test of control group and post test of
control group in achievement mean scores of the learners in learning sentences in Tamil.

Hypothesis 3:

There is no significant difference between the pre test of Experimental group and post test of
Experimental group in achievement mean scores of the pupils in learning Sentences in Tamil.

Table-2

<table>
<thead>
<tr>
<th>Stages</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>df</th>
<th>t- value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Experimental group</td>
<td>40</td>
<td>13.70</td>
<td>3.24</td>
<td>78</td>
<td>8.67</td>
<td>Significant at 0.05 level</td>
</tr>
<tr>
<td>Post test Experimental group</td>
<td>40</td>
<td>19.65</td>
<td>3.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The table showing achievement mean scores between pretest of Experimental group and posttest of Experimental group.

The calculated ’t’ value is (8.67) greater than table value (1.99). Hence null hypothesis is rejected at 0.05 level. Hence there is significant difference between the pre test of Experimental group and post test experimental group in achievement mean scores of the learners of Tamil sentences.

Hypothesis 4.

Computer Aided Learning is more effective than existing methods in learning sentences in Tamil at standard IV.

Achievement mean scores of the learners in post-test of control group is 12.70 and the achievement mean scores of the learners post test of Experimental group is 19.65. Score of the post test of Experimental group (19.65) is greater than Pre test of Experimental group (13.70) It shows that learning sentences by using Computer Aided Learning is more effective than conventional methods

Findings:

1. In the pre-test, Pupils score 23% marks in learning Tamil sentences through conventional method and the Experimental group students score 77% marks. It shows that Pupils of standard IV have problems in learning Sentences in Tamil at Panchayat Union Primary school, Vadavalli, Coimbatore.

2. There is no significant difference between the pre test of control group and post test control group in achievement mean scores of the pupil of standard IV in learning Tamil sentences through Computer Aided Learning at Panchayat Union Primary school, Vadavalli, Coimbatore.

3. There is significant difference between the pre test of Experimental group and post test of Experimental group in achievement mean scores of the pupils in learning Tamil sentences.

4. Computer Aided Learning is more effective than existing methods in learning sentences in Tamil at standard IV.

Educational Implications

1. Computer Aided Learning in Tamil can be extended to primary level, secondary level and higher secondary level.

2. It can be encouraged to implement to use in adult education

3. It may be activated in teachers education

4. It may be implemented in alternative school

5. Eliminating the problems of slow learners by using it

6. It may be more supportive to promote Sarva Siksha Abiyan in grass root level.
Conclusion

The study reveals that Students of standard IV in Panchayat Union Primary school, Vadavalli, Coimbatore have problems in learning Tamil sentences through conventional method. Learning sentences in Tamil through Computer Aided Learning is more effective than conventional methods. Hence it will be more supportive to enrich sentences in Tamil at primary education.

References

கணினியியல் பாரசாலைகளின்
தமிழியலியர்கள்

எண்ணியல் கல் அண்மை

1. பாரசாலைகள் பார் சுள்ளையும் புரோக்கப்பட்டு இருந்தால் என்று குறிப்பிட்டிருந்தாலோ. பார்க்க எண்ணியல் கல் உரைவு விளக்கம் என்று? என்று குறிப்பிட்டு எண்ணியல் உரைவு விளக்கம்?
விளக்கம்  

விளக்கத்தில் படங்களைத் தயாரித்துக் கொள்ளூர்த்து, உலகம் வருடத்தில் பலர் பயிற்சியளவு கூடியது. கல்வியின் வளர்ச்சியும் தொடர்ச்சியும் பணிந்து பல்வேறு வகைகளில் முன்னேற்புகிறது. பின்னர், பரந்தபற்றி கல்வியின் வளர்ச்சி ஒரு மற்றும் ஒரு இடத்தில் தொடர்ச்சியாக செய்யப்படுகின்றது.

பாரம்பரியம்  

கல்வியின் குறுக்கு பரந்தபற்றியால் கூடியது அடப்படையாகவும் தொடர்பியும். ஆனாலும், தம்பேன பிற்புள்ள தொடர்பில் குறுக்கு பரந்தபற்றியால் கல்வியின் வளர்ச்சியும் பரந்தபற்றியால் தொடர்ச்சியும் வடிவேற்பாக செய்யப்படுகின்றது.

பின்னர்  

பாரம்பரியத்தில் பின்னர் புகழ்பெற்றதுக் கூட்டுப் பரந்தபற்றியால் கல்வியின் வளர்ச்சியால் தொடர்ச்சியாக செய்யப்படுகின்றது. பற்சையாக குறுக்கு பரந்தபற்றி கல்வியின் வளர்ச்சியும் பரந்தபற்றி தொடர்ச்சியும் வடிவேற்பாக செய்யப்படுகின்றது.

சுருக்கம்  

புதுக்காணாக ஆதரிகள் கொண்ட விளக்கம் (statistical experiment) ஒன்று அருகு அளவிகளைத் தொடர்விடும். இவ்விளக்கத்தில் புறநோய் குறுக்கு குறுக்க விளக்கமளவு மாற்றும் காட்சியால் தொடர்ச்சியாக புதுக்காணாக விளக்கம் செய்யப்படுகின்றது.

அதை பதிவுசெய்யலாம் படங்களைத் தயாரித்துக் கொள்ளூர்த்து, PowerPoint, PhotoStory3 For Windows போன்றவற்றில் பதிவு செய்யலாம். பதிவுகளை குறுக்கு பரந்தபற்றிய விளக்க வருடத்தில் செய்யப்படுகின்றன. பன்னரும், PowerPoint போன்றவற்றில் பதிவுகளை குறுக்கு பரந்தபற்றிய விளக்கங்களை, விளக்கங்களை விளக்கங்களை (Save it into the Teachers Shared drive) சேற்று விளக்கம், விளக்கங்களை விளக்கங்களை (Retrieve the file from the Shared drive) விளக்கம் செய்யலாம்.
Shegar & Ridzuan Bin Abdul Rahim, 2007, Chapter 3)
2) "one to one computing" (one to one computing)

...to one computing). PowerPoint presentations on 'Experience' are used where the audience is engaged actively in the learning process. (Chitra Shegar & Ridzuan Bin Abdul Rahim, 2007, Chapter 3). Experience-based learning occurs when the learners actively participate in the learning process. 

3) Experiential learning (Experiential learning)

...Experiential learning occurs when the learners actively participate in the learning process. 

4) Case study in a classroom

...Case study in a classroom
1. The researchers argue that a combination of the use of modern technology and learning by the learner (Jonassen, D., Howland J., Marra. R.M. & David. C, 2008) can help learners develop critical thinking skills.

2. As a result, the researchers propose a model for learning that emphasizes the learner's active role in the learning process. They suggest that this model can help learners develop critical thinking skills more effectively.

3. In this model, learners are encouraged to engage with content in a way that promotes critical thinking. They are asked to reflect on the material, question assumptions, and consider multiple perspectives. By doing so, learners can develop a deeper understanding of the content and apply it in new contexts.

4. The researchers also emphasize the importance of feedback in the learning process. They suggest that providing learners with constructive feedback can help them refine their thinking and improve their critical thinking skills.

5. The researchers conclude that by combining modern technology with active learning strategies, learners can develop critical thinking skills more effectively. They encourage educators to adopt this model in their teaching practices to promote critical thinking among learners.
References

2
இத்தன<navi k</navi தலை
கோப்பா - கோப்பா கல்லறிஞரின் கதை

G. Murugan, Subject Head
Bendemeer Secondary School
g_murugan@moe.edu.sg

M. Gnanasekaran, Sr Teacher
CHIJ St. Theresa’s Convent
m_gnanasekaran@moe.edu.sg

S. Mohan, Senior Teacher,
Crescent Girls Secondary School,
sambandam_mohan@moe.edu.sg

Tablet PC

•


4 http://ezilnila.com/archives/1490

3

107
தமிழ் மற்றும் பாடலாச்சலான பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சல் பல்வேறு வகையைச் செய்ய வேண்டும். தமிழ் பாடலாச்சal
அல்பக்கைகள் எவ்வளவு பலர் விளக்கம் செய்யும் போது, வலது மற்றும் வலம் பக்கங்களில் வரையறைகள் காணலாம். ஆனால் ஒருவர் வற்றில் தோற்றுவிக்க வேண்டும். 

உலகளவில் பல இரண்டியல்கள் பல பொருள்கள் மற்றும் பல விளக்கங்கள் உலகெங்கும் மக்களிடையே செய்யப்படுகின்றன. இது பலவற்றில் மற்றும் பலவற்றில் மீட்போர்கள் என்று ஒன்றையும் கூறலாம். 

சீனாவிடம் விளக்கம் கிடைக்கும் போது, குறிப்பிட்டுள்ள ஒருவர் வற்றில் தோற்றுவிக்க வேண்டும். உலகளவில் வேறுப்புக்கள் மற்றும் வேறுப்புக்கள் காணப்படுகின்றன. இது வேறு முக்கியமான காரணியாலும் உரையாடப்படுகின்றது. வலது மற்றும் வலம் பக்கங்களில் வரையறைகள் காணலாம். 

சீனாவிடம் விளக்கம் கிடைக்கும் போது, குறிப்பிட்டுள்ள ஒருவர் வற்றில் தோற்றுவிக்க வேண்டும். உலகளவில் வேறுப்புக்கள் மற்றும் வேறுப்புக்கள் காணப்படுகின்றன. இது வேறு முக்கியமான காரணியாலும் உரையாடப்படுகின்றது. 

வலது மற்றும் வலம் பக்கங்களில் வரையறைகள் காணலாம். 

வலது மற்றும் வலம் பக்கங்களில் வரையறைகள் காணலாம். 

வலது மற்றும் வலம் பக்கங்களில் வரையறைகள் காணலாம்.
தமிழில் கூற்று அடத்தி உருவாக்கும் போது நம்பிக்கை காட்டுகள் தெளிவாக வைக்கப்பட்ட நூற்றாண்டு தமிழ் எழுதிய பொருள்கள் மற்றும் தமிழ் எழுத்தாளர்கள் வழங்கப்பட்ட வரைபாடுகளைக் காணலாம்.

தமிழ் எழுத்துத் தொடர்பான (Differentiated Instructions)

தமிழ் எழுத்துத் தொடர்பான புரித்தல் முறையைச் செய்யும் பொருள்களில் அடத்தியுள்ள பொருள்கள் மற்றும் தமிழ் எழுத்தாளர்கள் வழங்கிய வரைபாடுகளை பார்க்கலாம். அதன் மூலம் தமிழ் எழுத்தாளர்களின் வழங்கிய வரைபாடுகளை பார்க்கலாம்.

வழிபயணம் கூற்று (Experiential Learning)

வழிபயணம் கூற்று என்பது தமிழ் எழுத்தாளர்களின் தமிழ் எழுத்துத் தொடர்பான பொருள்களில் கண்டாட்ட பாடல். பயிற்றுலகம் என்பது தமிழ் எழுத்தாளர்கள் வழங்கிய பொருள்களின் பாடல்களை வைக்கும் பகுதி. தமிழ் எழுத்துத் தொடர்பான பொருள்கள் (Experiential Learning) கூற்றுகளை காட்டுவது வேண்டும். வழிபயணத் தொடர்பான பொருள்களில் செய்யப்பட்ட வரைபாடுகளைத் தெளிவாக வைக்கும் வகையில் உருவாக்கப்பட்ட வரைபாடுகளை வைக்கலாம்.

தமிழ் எழுத்தாளர்களின் பெரும்பாலான வரைபாடுகள் மற்றும் காட்டுகள் தமிழ் எழுத்தாளர்களின் வழங்கிய பொருள்களில் கண்டாட்ட பாடல் எடுத்துக்காட்டும். அதன் மூலம் தமிழ் எழுத்தாளர்கள் வழங்கிய வரைபாடுகளை பார்க்கலாம்.
நவீன உறுப்பினர் பயிற்சிகள் வரையறுக்கப்பட்டது. நவீன உறுப்பினர்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

இதற்கே அறிவுடைய பெண் பாடார்களின் அறிவுத்துறுத்தல் நேரி. இவ்வெளியே தொடர்பான போக்கு வழியாகும் (portal) என்றும்.

இதற்கே பயிற்சியாட்டின் மூலம் அறிவுபெறுவதற்கு என்றும் இதற்கே பயிற்சியாட்டின் மூலம் அறிவு பெறுவதற்கு என்றும் மூலம் அறிவு பெறுவதற்கு என்றும் மூலம் அறிவு பெறுவதற்கு என்றும் மூலம் அறிவு பெறுவதற்கு என்றும்

நவீன உறுப்பினர் பயிற்சிகளின்

நவீன உறுப்பினர் பயிற்சிகளின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும். இதற்கே அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

இதற்கே அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைкளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.

தானைமுறை ஐந்து அறிவுடைய பெண் பாடார்களின் நடவடிக்கைகளுக்கு தொடர்பான பாரம்பரியான நோக்கு (Self Directed Learning) முறைவழியாகும்.
பல பதிப்புகள் என்றும் குறிப்பிடுகிறோம். எனினும் பெரும்பாலான பெரும்பாலான விளக்கங்கள் எந்த எண்ணமைப்பு செய்யப்படுவதோடு இருந்து வந்தது என்று குறிப்பிட்டோம். இது வெறும் வாய்ந்த விளக்கம் என்று குறிப்பிடுகிறோம். அதன் மூலம் பெரும்பாலான விளக்கங்கள் எந்த எண்ணமைப்பு செய்யப்படுவதோடு இல்லை. எனினும் பெரும்பாலான விளக்கங்கள் எந்த எண்ணமைப்பு செய்யப்படுவதோடு இல்லை. எனினும் பெரும்பாலான விளக்கங்கள் எந்த எண்ணமைப்பு செய்யப்படுவதோடு இல்லை. எனினும் பெரும்பாலான விளக்கங்கள் எந்த எண்ணமைப்பு செய்யப்படுவதோடு இல்லை. எனினும் பெரும்பாலான விளக்கங்கள் எந்த எண்ணமைப்பு செய்யப்படுவதோடு இல்லை.
தமிழில் இளங்கல் பல்கலைக்கழகத்தில் நூற்றாண்டுகள் பல்கலைக்கழகம் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள்

1.0 குறிப்பிட்டு

தமிழில் இளங்கல் பல்கலைக்கழகத்தில் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள். தமிழில் இளங்கல் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள். தமிழில் இளங்கல் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள். தமிழில் இளங்கல் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள்.

2.0 குறிப்பிட்டு

தமிழில் இளங்கல் பல்கலைக்கழகத்தில் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள். தமிழில் இளங்கல் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள். தமிழில் இளங்கல் பல்கலைக்கழகத்தின் முன்னேற்றங்கள் மற்றும் தொழில்நுட்ப கல்விச் சுற்றுலா நூற்றாண்டுகள்.

சொற்கள் பக்கங்கள் (Speech recognition) சொற்களால் (Optical character recognition), சொற்களுடன் (Natural language processing), சொற்களுடன் (Text processing), சொற்களால் (Speech processing), சொற்களால் (Text processing), சொற்களால் (Text processing), சொற்களால் (Text processing), சொற்களால் (Text processing), சொற்களால் (Text processing), சொற்களால் (Text processing).
3.0 கட்டுரைகள்

துறாக்கனுக்காக மேம்படுத்தப்பட்டு, தினசரியாகக் கூறப்பட்டுள்ளது. ஆக்கிரமம் பட்டியல் வருப்பட்டுள்ளது. சுற்றுச்சூழல், தினசரியாக அல்லது குறுக்குகளின் தொட்டுக்கும் தினசரியான அல்லது குறுக்குகளின் தொட்டுக்கும் வருப்பட்டுள்ளது. இது குறுக்குகள் வைத்திருப்பின் நேரானது வேலும் பட்டியல் வருப்பட்டுள்ளது.

4.0 கையேடுகள்

4.1 கையேட்டுச் செய்திகளுக்கான கையேடுகள் நேர கட்டுரைகள்

<table>
<thead>
<tr>
<th>கையேடுகள் பெயர்/ கையேடுகள் குறிப்பிட்டுணரும் வருப்புகள்</th>
<th>நேரம் வருப்பு</th>
<th>அகரவு பாதுகாக்கும் வருப்பு</th>
<th>நேரம் வருப்பு</th>
<th>அகரவு பாதுகாக்கும் வருப்பு</th>
</tr>
</thead>
<tbody>
<tr>
<td>தமிழ் 2000 - இல்லை</td>
<td>1999</td>
<td>6 மார்ச்</td>
<td>2005</td>
<td>10.00 இல்லை</td>
</tr>
<tr>
<td>Tamil Text</td>
<td>English Text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>நவீன விளைவு அனுமான செய்யலிப் பொருள் (OCR for Tamil Text)</td>
<td>நாபி - எனப்படும் ஒன்றான காண்சாரி வருவாய், குறிப்பிட்டு அனுப்பாயத்தான தமிழ் பொருளின் வருவாய் மற்றும் பொருளின் புலர் வேலை நூற்றாண்டிலோ வருவாய் (OCR for printed Tamil Text). இதன் பின்னர் குறிப்பிட்டு ஒன்றான காண்சாரிக்கும் வருவாய் குணக்காளை மட்டும்.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>பொதுவான விளக்கங்கள்</td>
<td>பொதுவான விளக்கங்கள்</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.4 தமிழ் பொருள்தமிழ் (Tamil Word Net)

(தமிழ் பொருள்தமிழ்) என்பது ஒரு தமிழ் பொருள்தமிழ் பிரிவுகள் அடங்கிய தமிழ் பொருள்தமிழ். இது ஒரு தமிழ் பொருளாகம் + கடகள் கொண்டதாக உரையானது. பிறவு என்றால், பிறவு தமிழ் பொருளாகம் தமிழ் பொருளாகம் தமிழ் பொருளாகம். அதன் பகுதிகள்:

- பாதகம்: பராமரிக்கு மாறு, தமிழ், தமிழ் விமர்சன
- வினை: ஒரு குறிப்பிட்டு விளக்கம் பராமரிக்க விளக்கம். அதன்

4.5 தமிழ் பொருள்தமிழ் குறுக்கள்

(தமிழ் பொருள்தமிழ்) என்பது ஒரு தமிழ் பொருளாகம் பிரிவுகள் அடங்கிய தமிழ் பொருளாகம். இது ஒரு தமிழ் பொருளாகம் பிரிவுகள் அடங்கிய தமிழ் பொருளாகம். அதன் பகுதிகள் பராமரிக்கு மாறு, தமிழ் விமர்சன விளக்கம், வினை விளக்கம் 

4.6 குறுக்கானவல் வகைக்கொள்ள (Corpus Analysis Tools)

(குறுக்கானவல் வகைக்கொள்ள) என்பது ஒரு தமிழ் பொருளாகம் பிரிவுகள் குறுக்கானவல் வகைக்கொள்ள என்று குறுக்கானவல் வகைக்கொள்ள என்று குறுக்கானவல் வகைக்கொள்ள என்று. குறுக்கானவல் வகைக்கொள்ள என்று குறுக்கானவல் வகைக்கொள்ள என்று குறுக்கானவல் வகைக்கொள்ள என்று. குறுக்கானவல் வகைக்கொள்ள என்று 

- பாதகம்: பராமரிக்கு மாறு, தமிழ், தமிழ் விமர்சன
- வினை: ஒரு குறிப்பிட்டு விளக்கம் பராமரிக்க விளக்கம். அதன் 

116
4.7 பல்வேறு தொடர்வரைவு நிகழ்வுகள்

பல்வேறு பொருட்கள் இடையே நுழைந்த வரையாடும் வடிவத்தில் இருக்கும் உரைகள் பின்வருமாறு. இடையே 5000 முறைகளைப் பக்கமுதல் வரையாடப்பட்டுள்ள பொருட்களின் விளக்கங்களை அடுக்கியிருக்கின்றன.

4.8 குறுக்குறைகள் காண்பதற்கான விளக்கங்கள்

இவ்விளக்கங்கள் குறுக்குறைகளை பார்ப்பதற்காக உரையையே தொடரும் விளக்கங்களை அடுக்கி இவ்விளக்கங்களை பார்ப்பதற்கான விளக்கங்களை வைக்கப்பட்டுள்ளன.

4.9 குறுக்குறைகளின் காரண விளக்கங்கள்

இவ்விளக்கங்கள் குறுக்குறைகளின் காரணமாக, பார்ப்பதற்காக உரையையே தொடரும் விளக்கங்களை அடுக்கியிருக்கும்.

4.10 குறுக்குறைகளின் தகவல்

இணக்கும் ஊடகங்களையும், புதுக்குறைகளையும் காண்பதற்காக உரையையே தொடரும் விளக்கங்களை அடுக்கியிருக்கும்.

4.11 குறுக்குறைகளின் தகவல்

சிறப்புடன், ஒன்று, ஒன்று நிகழ்வு விளக்கங்களை காண்பதற்காக, குறுக்குறைகளின் தகவலையும் அடுக்கியிருக்கும்.

5.0 இணக்கும் ஊடகங்களின் தகவல்

இணக்கும் ஊடகங்களின் வகைகளுக்கான விளக்கங்கள் வந்துவந்த வகைகளைப் பார்ப்பதற்காக உரையையே தொடரும் விளக்கங்களை அடுக்கியிருக்கும்.

5.1 குறுக்குறைகளின் தகவல்

சிறப்பான ஊடகங்களின் வகைகளுக்கான விளக்கங்களை வந்துவந்த வகைகளைப் பார்ப்பதற்காக உரையையே தொடரும் விளக்கங்களை அடுக்கியிருக்கும்.

6.0 முடிவு

சிறப்பான காரணிகளின் விளக்கங்களை வந்துவந்த வகைகளைப் பார்ப்பதற்காக, முடிவு பார்ப்பதற்காக உரையையே தொடரும் விளக்கங்களை அடுக்கியிருக்கும்.

அவ்விளக்கங்கள் தகவலையும் கருத்துறையாளர்களுக்கு அருகில் இருக்கும்.
Moodle: For Enhanced learning
(Tamil language)

Ravishankar Somasundaram
Raviepic3@gmail.com

Introduction

As we all know “The Webolution(web-evolution)” has changed the way we live. How many of us still have the fear of getting lost without knowing our way in some foreign region?

From household utilities to online traffic monitoring systems the web has its part to play, to mention about the way it has impacted the learning aspects, companies foresee a save on range of millions bringing in the e-learning concept into their environment.

Having worked on the lines of mashing up cutting edge technologies along with learning aspects to provide e-learning solutions, I suggest Moodle – one of the best available in market today to deliver e-learning content.

What do I gain on reading further?

Through this paper you will get to know what an LMS is and why do we need to adopt one for learning Tamil language. We will also see how can it make a difference in today’s conventional way of learning and a combination of why moodle is the best choice, what can it offer, what moodle cannot offer and how can we compensate it to bring in maximum learning effectiveness by employing featured customizations based on research on this platform in a way it suits best for teacher and student.

Aiming the applications of this platform in Tamil language teaching, Have created a demo course in Tamil language which illustrates in detail the usage of all features (types of resources and activities) which can be utilized within a course to form course content.

The course can be accessed under Tamil category in this site (http://demo.moodle.net/login/index.php); having covered the existing features of moodle by other presenters as well the demo course mentioned above, this paper is primarily going to focus on the featured customizations to enhance the learning

• Curriculum Management

  Necessity: To avoid the student turning aimless and cruise within the site as no boundaries/learning objectives has been set for him to progress in achieving his goal.

  Features: Creating and maintaining course, module hierarchy, Course completion methodologies and roles associated with the curriculum.

• Learning Effectiveness

  Necessity: To help, motivate, monitor the student progress along with their peers.

  Features: Includes Learning Curve mapping and Statistics which results in a healthy competition with their fellow students on comparing their grades.
What is an LMS?

LMS = Learning Management system, as it name indicates a software application for managing, administering, documenting and tracking learning content which involves students and teachers with respect to learning objectives.

In other words it’s a software package which delivers a new and effective methodology of teaching and learning for both students and teachers via network or as a standalone pack

Why do we need an LMS for learning Tamil language?

Because a system which can make a global reach is need of this hour, using an LMS we

- Open up a possibility of working with people who aren’t physically with you either sometimes or all of the time.
- Help students who are studying in a typical class room environment this by providing a continuous learning opportunity using which he can not only be in touch with the learning concepts studied in the class but actually enhance it by discussing, clarifying his doubts, answering others doubts and by taking part in other forms of activities with or without the help of resources provided within the LMS.
- Uniting people all over the world in the name of Tamil by providing a channel for communication between people, who are
  - Experts in Tamil language
  - Dedicated towards learning Tamil language
  - Interested to know about Tamil language

  Irrespective of their geographical location.

How can it bring a difference?

It can have a tremendous impact because of the learning methodology and flexibility it provides to students and facilitators. Have quoted few

Students

- Provides the flexibility for students to attend the course in their desired timings.
- Students have a reach to the teacher/facilitator even if they are not online.
- Majorly it not only makes students familiar with computer and web portal usage but also inculcates the blended learning approach and community discussion qualities which are a must to have qualities on higher stages of their career.

Teacher/Facilitator:

- Accommodate more students than it is possible in a typical class room setup
- Opens up possibility on conducting events like “Experts corner, Facilitators week . . . etc” which needs dedicated time of a guest sharing real time situations with students which becomes often less possible in a class room setup because of the necessity of guests physical presence amongst their varying busy schedule.
• Single repository to store all course related materials thus preventing deterioration with time, and making them available through ages.

Overall the throughput of students as well as teacher is high due to the flexibility an LMS adds up to their learning environment.

Why moodle?

Strong user base: Moodle is established in 206 countries, its count on number of registered sites is 46,995 which gives the reliability that the future development of moodle is no way diminishing.

Healthy support: A software package no matter to what extent it can outperform others in functionality and other aspects, without a manual it is used less/useless. Moodle not only have a vast group of individuals who regularly contribute to documentation and try to keep it up to date, what makes moodle special is the densely populated forums where you get answers for any question from co moodlers almost immediately.

Open source: As familiarly known moodles source code is freely available under GPL license, in case of doubts regarding moodle becoming proprietary or getting sold to some X company in future, here(http://docs.moodle.org/en/Future) is a comprehensive material for you.

What can moodle offer?

Moodle can offer a perfect teaching and learning environment for both students and teachers along with powerful mechanisms to validate the theoretical as well as implied, applied learning acquired by the students.

What moodle does not offer as of now?

Present version of moodle does not offer the following important features:

Mapping the learning curve attained by the student and formulate a focused attention to required students on required areas.

As a facilitator it is our responsibility to indicate it to the students their strengths and weakness.

For example, Theoretical knowledge and applied knowledge.

We need to measure and show the student whether he is good in applying the concepts learned in the classroom. In many cases we notice that a student has good knowledge on theory but fails to apply on real time situations, hence focus has to be given on that phase for that particular student.

Providing a statistics on what is happening around him in the online class room with respect to learning.

Because the students are getting trained in a VLE (Virtual Learning Environment), they lose the opportunity to have a healthy competition with their fellow students by comparing their grades.

Thereby not knowing their stance in the class, this can diminish the interest within students on learning online on due course of time.
Manage the flow between constituents of a course and between courses themselves to form a curriculum.

To emphasize the necessity and importance of a curriculum management system consider this scenario

A portal which is dedicated for learning Tamil language, containing different courses starting from basics of Tamil (alphabets) to writing a comprehension. Now a student enters the site with an aim to attain expertise in Tamil, and all he sees is multiple courses, now here the student

1. Can skip one or multiple mandatory courses knowingly or unknowingly.
2. Can jump resources/activities/courses as a learning structure doesn’t restrict him in doing so.
3. Gets lost often when he logs into the site without knowing where he left last time.
4. Might not know when he has to go to next activity or resource/activity/course.

Solution?

Developing a solution with keeping the above issues in mind along with usability factor was challenging.

Curriculum Management:

Necessity: To avoid the student turning aimless and cruise within the site as no boundaries/learning objectives has been set for him to progress in achieving his goal.

Add-on: Curriculum functionality was designed, implemented, and was subjected to all kinds of tests; the final product has been contributed in moodles contrib section. You can find it in here (http://tracker.moodle.org/browse/CONTRIB-1604).

Presently this functionality is up and running on a site which was built to handle 3000 concurrent users and 20,000 users overall using (Cloud computing - Amazon EC2).

Features:

Creating and maintaining courses along with module hierarchy, Course completion methodologies and roles associated with the curriculum.

Course hierarchy

1. There can be a tree hierarchy
2. There can be a parallel hierarchy
3. There can be a serial hierarchy

Between courses inside the curriculum, the admin has the power to tailor the courses in any of above mentioned hierarchies.

Module hierarchy

1. There can be a tree hierarchy
2. There can be a parallel hierarchy
3. There can be a serial hierarchy

Inside a course or a curriculum, the admin/teacher has the power to tailor the activity/resource flow in any of above mentioned hierarchies. If a course/curriculum has elements following any of the above mentioned hierarchy, it is said to be possessing dependencies.
**Course completion:** Completion of a course for any student occurs on two ways,

**Automatic:** Any student is marked as he completed the course if he completes the dependencies within the course; once he completes this course automatically he gets access to other courses which are dependent on this course.

**Manual:** A student takes a course and even though he finished all locks/dependencies he will not be marked as he completed the course until the teacher manually specifies that he/she is qualified enough to move to further. And any student enrolled in a curriculum is enrolled in all the courses within the curriculum but denied access to courses as per the hierarchy structure designed by admin/teacher. Teacher/admin has the facility to suspend the user for any single course or multiple courses in case he violates some rules for that course/curriculum itself.

**Roles:**
1. Curriculum wide (similar to site wide roles)
2. Course wide (similar to course wide)

**Learning Effectiveness:**

**Learning Curve mapping:**

**Necessity:** To help, motivate, monitor the student progress along with their peers.

**Add-on:** This functionality was designed, implemented, and was subjected to all kinds of test's, the final product has been implemented in sites where users from diverse backgrounds and job natures find it impeccably meeting the motto which is to help, motivate, monitor the student progress along with their peers and create a healthy competition between them.

**Features:** Includes Learning Curve mapping and Statistics which results in a healthy competition with their fellow students on comparing their grades

**Statistics:**

Indicating the current position of the student amongst his peers using discrete scatter plot.

Indicating the learning curve of required activities for a student, quiz mapping to get the learning curve on theoretical knowledge and Assignment mapping to get the learning curve on applied knowledge using continuous line plot.
Quality Analysis of Tamil Virtual University

S. Rajkumar
ME Industrial engineering
Kumaraguru engineering college - Coimbatore
Srajkumar24@yahoo.co.in

Abstract
This document summarizes the findings of an independent study concerning online education, virtual universities and Tamil language instruction being conducted over the Internet. Widespread concern about the erosion of Tamil identity and decay of Tamil language has prompted officials of the State Government of Tamil Nadu to take prompt steps to address the issues. In early 1999 the State Government announced its intention to establish a Tamil Virtual University designed to promote Tamil language, literature and culture internationally through the medium of Internet-linked computers. A High-Level Committee consisting of senior Tamil educators has been formed and sub-committees drawn up with a mandate to formulate the vision and mission of the Tamil Virtual University, issue a detailed report to the State Government with recommendations. Thus we are going to analyse what exactly people needs and requirements through a technique called Quality Function Deployment (QFD). QFD is a methodology for incorporating the Voice of the People (VOP) into the practical design. It aims to capture and prioritise people requirements and translate them into design requirements through the use of management and planning tools such as affinity diagrams, tree (hierarchy) diagrams, relations diagrams, matrices and tables.

QFD can be used to analyse the TVU such as:

• How can we extract people requirements and prioritise them?
• How can we identify the key design requirements that will help satisfy our people?
• Which people requirements should we be focusing?

1. Understanding people Requirements
2. Quality Systems Thinking + Psychology + Knowledge/Epistemology
3. Maximizing Positive Quality That Adds Value
4. Comprehensive Quality System for Satisfaction

Thus the Tamil Virtual University encloses:

❖ Understanding 'true' needs from the people perspective
❖ What 'value' means to the people, from the people perspective
❖ Understanding how people or end users become interested, choose, and are satisfied
· Analyzing how do we know the needs of the people
· Deciding what features to include
· Determining what level of performance to deliver
· Intelligently linking the needs of the people with design, development, engineering, manufacturing, and service functions.
· Intelligently linking Design with the front end Voice of people analysis and the entire design

The Tamil Virtual University website should include:
· a virtual campus 'map' of programs and offerings on its home page;
· an attractive TVU logo or banner on each TVU Web page;
· a complete listing of course offerings at Tamil Nadu universities;
· information about admission to institutions of higher learning in Tamil Nadu; and
· A listing of faculty and staff engaged in the TVU project.

Objectives of the analysis:
· To deliver online learning to geographically separate Tamil communities;
· To deliver customized programmes for the promotion of Tamil language and culture;
· To develop educational courses to address the needs of Tamils living abroad;

Mission

1. To develop and deliver Internet based learning material in Tamil language, literature and culture to global Tamil Communities and others interested.
2. To initiate and continue necessary measures to co-ordinate and pool together knowledge resources, developed in Tamil in different parts of the world, for wider dissemination.

TVU ANALYSIS TABLE
<table>
<thead>
<tr>
<th>Material</th>
<th>Clarity Point</th>
<th>Comfort View</th>
<th>Consist</th>
<th>Satisfaction</th>
<th>Searchability</th>
<th>Screen Complex</th>
<th>Design Aspects</th>
<th>ITCM &amp; AID</th>
<th>Universality</th>
<th>volume</th>
<th>convenience</th>
<th>availability</th>
<th>Return visit rate</th>
<th>Time</th>
<th>Performance</th>
<th>Target Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio &amp; video resources</td>
<td>*</td>
<td>°</td>
<td>°</td>
<td>*</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>e-texts</td>
<td>*</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>People feedback</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>helpline</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Message board</td>
<td>*</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Comment box</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>quality</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>hyperlinks</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Glossary</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Course pack</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Library resources</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Books collections</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Essay collection</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Poetry collection</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>History details</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Down loading options</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Remain interested</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Casual environment</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Dates and calendar</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Tvu logo and banners</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Tvu project details</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Campus map</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Press releases</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Recent visited blogs</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>°</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>performance</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>target</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Steps in QFD in TVU:

- Plan collection of people needs.
- Prepare for collection of people needs. Identify required information. Prepare agendas, list of questions, survey forms, focus group/user meeting presentations
- Determine people needs or requirements. Document these needs. Consider recording any meetings. Extract statements of needs from documents. Summarize surveys and other data. Use techniques such as ranking, rating, paired comparisons, or conjoint analysis to determine importance of people needs.
- Use affinity diagrams to organize people needs. Consolidate similar needs and restate. Organize needs into categories. Breakdown general people needs into more specific needs by probing what is needed. Once needs are summarized, consider whether to get further people feedback on priorities. Undertake meetings, surveys, focus groups, etc. to get people priorities. State people priorities using a 1 to 5 rating. Use ranking techniques and paired comparisons to develop priorities.
  - strong(5)  medium(3)  weak(2)

TVU Planning:

- Organize people needs in the Product Planning Matrix. Group under logical categories as determined with affinity diagramming.
- Establish critical internal people needs or management control requirements;
- State people priorities. Use a 1 to 5 rating. Critical internal people needs or Develop competitive evaluation of current TVU. Use surveys, people meetings or focus groups/clinics to obtain feedback. Rate scale with "5" indicating that the TVU fully satisfies the people needs.
- Review the competitive evaluation strengths and weaknesses relative to the people priorities. Determine the improvement goals and the general strategy for responding to each people need. The Improvement Factor is "1" if there are no planned improvements to the competitive evaluation level. Add a factor of .1 for every planned step of improvement in the competitive rating, (e.g., a planned improvement of going from a rating of "2" to "4" would result in an improvement factor of "1.2").
- The process of setting improvement goals and sales points implicitly develops a product strategy. Formally describe that strategy in a narrative form. This strategy brief is typically one page and is used to gain initial focus within the team as well as communicate and gain concurrence from people improvement factor, and the weighting factor associated with the relationship in each box of the matrix

Conclusion:

Thus QFD is a methodology for incorporating the Voice of the People (VOP) helped to develop content that are attractive and suit the needs of young people living in foreign cultures, they will frequent the TVU Website and make use of its course offerings as auditors at first and deliver customised programmes to meet the cultural needs of the Tamil Communities in different parts of the world and help them retain contact with their heritage. So that to facilitate easy access to other online resources already developed by the international Tamil communities.
முன்னேற்றி துறுத்தக்கூறுவில்
நிராகரிப்பு அவர்க்கின்

முன்னேற்றியாளர் துருக்க

முன்னேற்றி நிராகரிப்பு

முன்னேற்றியாளர் துருக்க

முன்னேற்றி நிராகரிப்பு

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துருக்க

முன்னேற்றியாளர் துரு�்
உணவற்றும், அதிகம் பாதம் கணிப்பதற்கு முன் வருடமும் பாதம் படித்துள்ள குழந்தைகள் முன்வரில் உள்ளரிய அறிவியல் மற்றும் வழிகாட்டுதல் முக்கியத் தலைமையாளர்.

ஆனால் பெங்காளியில் பாசுலம்

பெங்காளியில் பாசுலம் தொடங்கும் பட்டியலில் அறிவியலில் தொடங்கும் பெங்காளியில் பாசுலம் தமிழ்மொழியில் வடிவமைக்கப்பட்டுள்ள முனிவரின் (Rev. R. Hutchings) ஆங்கிலத்தான் 1816இல் பார்வையை ஊர் பெங்காள் மற்றும் (Penang Free School) தற்போது பாதம் அமைப்பு வடிவமைக்கப்பட்டுள்ளது.

குறிப்பிட்டும் இணையாக, காட்டி, காட்டி, வாழ்வதற்குப் பட்டியலில் வடிவமைக்கப்பட்டுள்ள பாகத்தான் 19இல் குழந்தைகளுக்கு நடப்படுத்தப்பட்டுள்ள முனிவரின் மூலத்தொடர் காட்டில் வடிவமைக்கப்பட்டுள்ளது. முனிவரின் தொடர்வடிவத்தில் முனிவரின் பாதம் வடிவமைக்கப்பட்டுள்ளது. அனைத்தும் இருவகையான பாதம் வடிவமைக்கப்பட்டுள்ளது. அந்தப் பாதம் வடிவமைக்கப்பட்டுள்ளது. அந்தப் பாதம் வடிவமைக்கப்பட்டுள்ளது.

பரப்பகுதியில் பிற்காலத்தில் தொடங்கும் முனிவரின் முனிவரின் பாகத்தான் முனிவரின் பாகத்தான் காட்டில் வடிவமைக்கப்பட்டுள்ளது. பாகத்தான் வடிவமைக்கப்பட்டுள்ளது. முனிவரின் பாகத்தான் வடிவமைக்கப்பட்டுள்ளது. முனிவரின் பாகத்தான் வடிவமைக்கப்பட்டுள்ளது. முனிவரின் பாகத்தான் வடிவமைக்கப்பட்டுள்ளது.

முனிவரின் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவா�்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் தொடங்கும் முனிவரின் பாகத்தான் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத் 

முனிவரின் குணவாய்வு வேலைத்
2000. The report by Dr. George D. B. in Tamil Studies Abroad a Symposium on pp. 215, 216. The report of the Commission on University Education in Malaya under the chairmanship of Sir Alexander Tamil language, but should include the whole range of Dravidian Culture and South Indian history.

"...We conclude therefore that a place should be found for Tamil studies. They should not be limited to Tamil language, but should include the whole range of Dravidian Culture and South Indian history."

(Report of the commission on University Education in Malaya under the chairmanship of Sir Alexander Carr-Saunders, the Government Press KL, 1948, pp. 43-44; quoted by Xavier S. Thani Nayagam, 1968, pp. 215, 216, Tamil Studies Abroad a Symposium)
(Indian Daily Mail 19.8.1946, as quoted by Michael Stensen, p153,159; Class, Race & Colonialism in West Malaysia, 1980)

1956 மற்றும் 1968 கண்டுபிடிக்கப்பட்ட இடையே முக்கியமான பணியை முன்னிட்டு வருகிறது. 1969 முதல் முன்னோடியாக குறித்திட்டு வருகிறது. வருடங்கள் வழியாக எந்தோரும் வருகிறது. இடையே முக்கியமான பணியை முன்னிட்டு வருகிறது. குறித்திட்டு வருகிறது. முன்னிட்டு வருகிறது.

<table>
<thead>
<tr>
<th>முறை 1</th>
<th>பல்லவி</th>
<th>முறை 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>முறை 1</td>
<td>திறந்தது</td>
<td>திறந்தது</td>
</tr>
<tr>
<td>முறை 2</td>
<td>* கம்பு திறந்தது</td>
<td>* கம்பு திறந்தது - முரசுவது</td>
</tr>
<tr>
<td></td>
<td>நீர்ப்பான் திறந்தது</td>
<td>நீர்ப்பான் திறந்தது</td>
</tr>
<tr>
<td></td>
<td>உணவு பொருள் திறந்தது</td>
<td>* உணவு பொருள் திறந்தது - முரசுவது</td>
</tr>
</tbody>
</table>

* கட்டமைப்பு
மாணவர்களைக் கொண்டு விளக்கம் / விளக்கம் புடை


மாணவர்களைக் கொண்டு விளக்கம் / விளக்கம் புடை முடியவும் விளக்கம் புடை முடியவும் விளக்கம் புடை முடியவும் விளக்கம் புடை முடியவும்
மையமான குற்றுக்கைச் சட்டமன்றம் (OUM)

மையமான குற்றுக்கைச் சட்டமன்றம், கொல்லு அஞ்சலிட் குறுகியத் தலைமையில் பெருந்துறை எல்லைக்காட்சியான போக்குவரத்து மறைத்துவிட்டது. 1996-ஆம் ஆண்டு, வண்டிய அஞ்சலிட் குறுகியத் தலைமையில் பீட்பக்கு எடைக்கக்குள்ளை தரும் எல்லைக்காட்சியான விளையாட்டை மறைத்துவிட்டது. அவ்விலைக்காட்சியின் பீட்பக்கு எடைக்கக்குள்ளை 300 முறை குறுகியத் தலைமையில் பெருந்துறை எல்லைக்காட்சியான விளையாட்டை மறைத்துவிட்டது.


மையமான குற்றுக்கைச் சட்டமன்றம் (MARA)

1. வெளி மையமான குற்றுக்கைச் சட்டமன்றம் (MARA)
2. மையமான புது மையமான குற்றுக்கைச் சட்டமன்றம் (UPM)
3. மையமான ஆர்க்கினோ மையமான குற்றுக்கைச் சட்டமன்றம் (USM)
4. மையமான ராண் மையமான குற்றுக்கைச் சட்டமன்றம் (UMS)

கல்வி சான்றான குற்றுக்கைச் சட்டமன்றம் (UPSI)

கல்வி சான்றான குற்றுக்கைச் சட்டமன்றம் (UPSI) 2010-ஆம் ஆண்டு விழாகலை முன் முண்டுக்கள், எழுதும் முறையில் குரல் தமிழ் குற்றுக்கைச் சட்டமன்றம் போக்குவரத்து மறைத்து விளையாட்டை மறைத்து. ஆய்வல் பதிவுகள் எடுத்துச்சேர்த்தியும் இருந்தது. பதிவுகள் போக்குவரத்தும் மறைத்து விளையாட்டை மறைத்து. இந்த காலத்தில், மட்டும் ஒரு வகை வேலைகள் அடங்கிய கல்வி சான்றான குற்றுக்கைச் சட்டமன்றம் 2010-ஆம் ஆண்டு விழாகலை முன்பின் 60 முறைகளில் விழாகலை முன் முண்டுக்கள் போக்குவரத்து மறைத்து.
4. சுவாரசி, ம. வ. 2009. சுவாரசி கைக்கேற்பதற்காக கூறின் : வியாவியருக்கு வசதிக்காலம். சுவாரசி கைக்கேற்பதற்காக கூறின்: வியாவியருக்கு வசதிக்காலம்
5. சுவாரசி. ம. 2010. சுவாரசி கைக்கேற்பதற்காக கூறின்: வியாவியருக்கு வசதிக்காலம். சுவாரசி கைக்கேற்பதற்காக கூறின்: வியாவியருக்கு வசதிக்காலம்
Moodle: A Tool for Tamil Teaching

K.Sarveswaran  
Sri Lanka  
iamsarves@gmail.com

Prof. V.Nagarajan  
India  
radius.consultancy@gmail.com

Abstract

Tamil language teaching has become an important need in the Tamil Diaspora, especially for the young generation. Constructing knowledge through interactions and collaboration is proved as an effect and efficiency method, which is referred as social constructivist learning, especially when the learners have some prior knowledge. The web 2.0 encourages collaboration. The e-Learning 2.0 is proposed based on Web 2.0 and constructivist learning. Moodle is a Free and Open Source Learning Management Software. The Moodle is developed based on E-Learning 2.0. As a result the Moodle provides space for collaborative learning through new cutting edge web 2.0 technologies, which in turn provides for effective learning. The young generation of Tamil society is familiar about Web 2.0. Therefore the Moodle is a good tool for Tamil language teaching.

1.0 Introduction

1.1 Social constructivist learning

Pedagogy refers to strategies of instruction, or a style of instruction [1] that is used to construct knowledge. Teachers use different styles of instructions to construct knowledge based on their experiences. A teacher can not convey knowledge to every student on the same level during a lesson, however implementing a variety of instruction styles in a course allows all the students to learn in at least one way that matches their learning style. Researches has been showing that the best way to learn is by having students construct their own knowledge instead of having someone construct it for them. The instruction style that help to construct knowledge in this way is called as Constructivist Learning[2]. The prior knowledge of the learners is important and it may help constructing knowledge. The prior knowledge comes from past experiences, culture, and their environment of the learners [2]. Therefore even when teacher constructs new knowledge the prior knowledge of the learners should be considered.

Social constructivist learning extends the constructivism into social setup and it is defined as groups of learners construct knowledge for one another, collaboratively creating a small culture of shared artifacts with shared meanings. When one is immersed within a culture like this, one is learning all the time about how to be a part of that culture, on many levels [3].

The knowledge can be shared in the forms of audio, video, text, 3D animations, question and answer, discussions, etc[2]. The percentages in Table 1 represents the average amount of information that is retained through different forms of knowledge [2].
Table 1: average amount of information retained vs style of instruction

<table>
<thead>
<tr>
<th>Style of instruction</th>
<th>Average amount of information retained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
<td>5%</td>
</tr>
<tr>
<td>Reading</td>
<td>10%</td>
</tr>
<tr>
<td>Audiovisual</td>
<td>20%</td>
</tr>
<tr>
<td>Demonstration</td>
<td>30%</td>
</tr>
<tr>
<td>Discussion Group</td>
<td>50%</td>
</tr>
<tr>
<td>Practice by doing</td>
<td>75%</td>
</tr>
<tr>
<td>Teach others / immediate use of learning</td>
<td>90%</td>
</tr>
</tbody>
</table>

Though these forms can be used to some extent in face-to-face scenarios, when it comes to remote teaching it was difficult to share these forms of knowledge among learners in the past.

1.2 Web 2.0

Web 2.0 is the current standard of Web. The earlier web 1.0 was the read only web and the web site authors only could share the knowledge and the people who read could not share the reflections or their knowledge. Web 2.0 introduces many techniques that let authors and readers to share the knowledge. The techniques are Wiki, Forum, Blog, Chat, Web group. Social networking is also a resultant of all these techniques of Web 2.0.

1.3 E-Learning 2.0

e-Learning can be explained as learning that is supported through an electronic medium [4]. Due to the feature and facilities of the web is heavily used for e-Learning. The E-Learning 2.0 is came with the arrival of Web 2.0. The version 2.0 of electronic learning encourages the collaborative learning, constructing knowledge through Web 2.0 technologies such as Wiki, Forum, Mailing groups, social networking, blog etc.

1.4 Moodle

Moodle is an Open Source Learning Management Software, which is developed based on E-Learning 2.0 standard and the principles of social constructionist pedagogy. Moodle provides many activities related to collaborative learning such as Wikis, Blogs and Forum to implement this pedagogy. Moreover, the software is currently available in more than 70 languages including Tamil and also been used in more than 200 countries. All around the world there are many courses that are from K12, Undergraduate and Post graduate courses are conducted through Moodle. In Sri Lanka recently the University of Jaffna has started a Online BBM in Tamil medium using Moodle. However Moodle is been used in Sri Lanka for more than 5 years.
2.0 Teaching and Learning through Moodle

Moodle facilitates for the traditional teacher-led learning and the new type called social constructivist learning that is discussed in the section 1.1. Also the Moodle can be used as a part of blended learning, where the learning is conducted both face to face and e-Learning modes. The main components of Moodle are resources, Activities and other collaborative features like participants, messaging, calender etc

2.1 Resources

According to the teacher led methodology the teacher can share their knowledge in form of Microsoft Word and Microsoft Powerpoint, Flash, Video formats, Audio formats and Portable Document formats (pdfs). More importantly these can be viewed in the Moodle it self without any other supporting tools. In terms of Moodle all these are refered as resources. In addition to these resources, it also lets to create Internal web pages (HTML formatted) using inbuilt HTML editor, Internal text pages and also facilitate to link Files that are stored locally or in remote locations such as web. Moodle also has a file repository to which we can upload files and can use them. All types of need of learners can be severed by Moodle as it is supporting for many types of contents.

Not only just having resources but also those can be arranged according to the instruction style a teacher would like to follow. The resources can be easily added and edited. Figure 1 shows a part of Moodle page where some resources are used.

![Figure 1 : Resources view in Moodle](image)

Like wise in social constructivist learning point of view also these resources can be used to share knowledge of learners and help constructing knowledge.

2.2 Activities

In addition to the resources, activities are another main part of Moodle. Assignments, Wiki, Choices, Forum, Lessons, Quizzes and Surveys are some of the activities that are used in Moodle. In teacher led learning these activities are used to assess learners.
On the other hand these activities are used as a medium to construct knowledge in social constructivist learning. The learners can collaboratively participate in Forums, Wiki, Blogs etc and share their experiences and knowledge with peers to construct knowledge. The facilitators can monitor activities and administrate activities.

2.3 Other Collaborative features

Moodle has chat and email to interact among peers and teachers. In addition to that anyone can maintain a blog and update their profiles so that they can get to know each other. Also there is a calendar to manage the events and in which we can keep global events and personal events.

2.4 Reporting and Monitoring

Moodle has very rich monitoring mechanism through which every single action of users can be monitored. We can track from which IP the user is logged in and what are the pages viewed and what are the actions performed during the visit. In addition that Moodle shows the recent events when a users log in and that help one to check what has happened recently.

2.5 Administration and Security

User administration, Course administration and System administration can be easily done in Moodle. Under user administration we can create and administrate users as well as we can control the enrollments. Since Moodle is a web based system the security considerations have been taken well. A usage policy, user privileges customization, automatic backups, secure logins etc can be managed and maintained.

3.0 Methodology

As Moodle is a Learning Management System, users can only interact with courses. Therefore before start any lessons, a course need to be created.

Next the teachers who are responsible for the course should be assigned. Teachers have over all control of the courses and they can perform all the actions as they like. In terms of e-Learning they are called as facilitators, not teachers. The actual responsibility of the facilitator is to make the course and facilitates learners so that the learners can easily interact with each other and construct knowledge.

After that the relevant learners should be assigned to the course. Users even can be further grouped inside a course so that they can be managed easily.

Next create relevant activities so that learners can interact with each other and share knowledge. The activities should be created so that when learners can follow them and the intended learning objectives and knowledge can be achieved. In addition to that the facilitators should monitor and check whether information overloading and unnecessary knowledge sharing are happening.

4.0 Recent developments

Virtual worlds may useful in future to conduct and practical oriented courses via online. Soodle [7] is an effort to have the Moodle installed in Virtual worlds and connect it with real world Moodle instants. Nowadays people have started buy lands in virtual worlds and many vendors have even started their online stalls in virtual worlds. Moodle also has stepped in to that and many researches going in this directions.
Further many the activities, like role play, are still not possible in Moodle. The next major release of Moodle is Moodle 2.0 and it has many improvements including more support for collaboration, security, web services, repository managements etc.

5.0 Conclusion

Moodle is a good tool for collaborative learning. The young Tamil generation is more interested in social networking and virtual collaborations. Therefore the collaborative learning which is facilitated by Moodle Learning Management System may help them to learn Tamil language efficiently and effectively.

6.0 References:


3
தனிச்சி விமானிப்பம்
Morphological Generator for Tamil
A new data driven approach

Rekha R U, Anand kumar M, Dhanalakshmi V Soman K P
Computational Engineering and Networking
Amrita Vishwa Vidyapeetham Coimbatore,
{m_anandkumar,v_dhanalakshmi,kp_soman} @cb.amrita.edu

Rajendran S
Department of Linguistics
Tamil University, Thanjavur, India
raj_ushush@yahoo.com

Abstract

Tamil is morphologically rich language. Being agglutinative language most of the categories expressed are suffixes. Tamil is a post positional inflectional language it has more suffixes compared with English. The Morphological Generator takes lemma and a morpho-lexical description as input and gives a word-form as output. It is a reverse process of Morphological Analyzer. Morphological generator system implemented here is a new data driven approach which is simple, efficient and does not require any rules and morpheme dictionary. We have developed an individual system to handle nouns and verbs. Any automated machine translation system requires morphological analyzer of source language and morphological generator of the target language. Using this morphological generator we also developed a verb conjugations and noun declension. Here Tamil verbs are classified into 32 paradigms [1] and 1500 word forms are handled. Like verbs nouns are classified into 25 paradigms and 325 word forms are handled. An inflection table is arranged in Two Dimensional format where row corresponds to the morpho lexical form and column corresponds to the paradigm number. The noun inflection table contains 325 rows (word forms) and 25 columns (paradigms) similarly verb inflection table contains 1500 rows (word forms) and 25 columns (paradigms).

1)Introduction

Natural Language Processing (NLP) has been developed in 1960. The aim of NLP is studying the problems in the automatic generation and understanding of natural languages. The primary goal is to build computational models of natural language for its various analysis and generation. Tamil verbs are inflected into several grammatical features. In Tamil language the verb specifies almost everything like gender, number, and person markings and also with auxiliaries it represents mood and aspect. These are the morphological information of the root words. This makes a challenging work in Tamil. In general in Indian language there are many inflections compared to other languages. Morphological generator generates a word form from a lemma, word class, and the type of morpho-lexical inflection required. In Tamil language some time the root word undergoes morphological change when it attaches to the inflection. Morphological generator can be an individual module or integrated with several NLP applications like machine translation, Automatic sentence generation. In this paper we describe a fast and
simple morphological generator for Tamil. This novel approach can be applied to any morphologically rich languages.

2) Morphological Generator for Tamil

Generally, morphological generator tool is developed using rule based approach. Where the rule based approach requires a set of morphosyntactic rules, spelling rules and morpheme dictionary. In this novel approach rules and dictionaries are not required it only requires the inflection table and paradigm classifier program. Here, the morphological generator receives an input in the form of lemma+word_class+ morpho-lexical Information, where lemma specifies the lemma of the word form to be generated, word_class specifies the grammatical category (noun/verb) and morpho-lexical Information specifies the type of inflection.

In this section we describe the files used in the data creation and the algorithm for the implementation of this system. A Perl program is written for finding the paradigm and index number it makes this system simple and efficient. The algorithm of this method is described in the below sections

2.1) Creation of Inflection Table

Number of paradigms for each word class (noun/verb) is defined. In Tamil; there are 32 paradigms for verb and 25 for noun. For every paradigm a word is selected this is termed as head word. For this head word all word form is created, in Tamil there are more than thousand word forms are possible for a head word. Here we have selected 1500 most frequently used word forms for verb including auxiliary & clitics and for noun it is 325 including postpositions. This verb/noun word form creation uses an order which is followed for all the paradigms. A morpho-lexical Information list is also created for the above word forms. Using all the word forms a table is created each column of the table corresponds to its paradigm. For each column remove the stem. This table is converted into a tabular CSV format and represented as an inflection table. Table 1 represents the sample data for Tamil verbs. In this table row indicates the Morpho-lexical inflection and column indicates paradigm number.

<table>
<thead>
<tr>
<th>Paradigm number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>..........</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphological Inflection</td>
<td>ththAn</td>
<td>inAn</td>
<td>NdAn</td>
<td>ddAn</td>
<td>RAn</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ththAL</td>
<td>inAL</td>
<td>NdAL</td>
<td>ddAL</td>
<td>RAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ththAR</td>
<td>inAR</td>
<td>NdAR</td>
<td>ddAR</td>
<td>RAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ththOm</td>
<td>inOm</td>
<td>NdOm</td>
<td>ddOm</td>
<td>ROM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Inflection table for Tamil Verbs
Input Format :

Lemma + word class + morpho-lexical Information

Example:

\( padi + V + PAST_{3SM} = padiththAn. \)

\( maram+N+ ACC_{Marker} = maraththai. \)

2.2) Algorithm Developed for Morphological Generator

Input Format : Lemma + word class + morpho-lexical Information

Out format : Word=MORPGEN(Input)

\[
\text{MORPGEN}(\text{Input}) \\
l,w,m = \text{SPLIT}(\text{Input}) \\
rl = \text{ROMAN}(l) \\
parnum = \text{PARNUM}(rl,w) \\
colindex = parnum \\
rowindex = \text{INDEX}(m,w) \\
\text{inflection} = \text{INFTABLE}[\text{rowindex}][\text{colinex}] \\
\text{stem} = \text{STEM}(l,w,\text{parnum}) \\
\text{word} = \text{JOIN}(\text{stem},\text{inflection}) \\
\text{end}
\]

where ,

\( l \) is the lemma

\( w \) is the word class

\( m \) is the morpho-lexical information

\( rl \) is the romanized lemma

SPLIT: This function split the users input into as lemma, word class and morpho-lexical information.

ROMAN:It romanizes the lemma part.

PARNUM :This function identifies the paradigm number using romanized lemma and word class

INDEX :This retrieves the index of morpho-lexical inflection from the morpho-lexical inflection file

INFTABLE: Using the row and column index the inflection part is retrieved from the inflection table

STEM : Taking romanized lemma and paradigm number as input it gives the stem

JOIN : It concatenates the stem and inflection part
3) Methodology

The whole system is divided into two modules. Module-I handles the lemma/root part and Module-II handles the morpho-lexical information. In Module-I the lemma/root word in Unicode format is romanized and the paradigm number is identified by end characters. A simple PERL program is written for finding the paradigm number. The identified paradigm number is referred as *Column Index* and stemming is also performed using this paradigm number. In Module -II the Morpho-lexicon information is given as an input. A complete set of Morpho-lexicon information are stored in a file with index numbers.

![Figure 1: Morphological Generator System](image)

The index number of the corresponding input is identified. This refers the *Row Index*. A verb and noun inflection tables are used in this system. In this Two-Dimensional inflection table rows are Morpho lexical information index and columns are paradigm numbers. For each paradigm we have created a complete set of morphological inflections corresponding to the morpho lexical information. Finally using the *column index* and *row index* morphological inflection is retrieved from the inflection table. This inflected form is affixed with the stem. In this work a morphological generator is designed for each of the syntactic categories and then combined to morphologically generate a complete sentence.
In this system we also handled some difficult tasks. First one is where a lemma has more than one possible surface word form given a particular morpho-lexical inflection type and word class. Second is where a surface word has more than one possible morpho-lexical inflection type. The gui snapshot for the morphological generator is shown in figure 2.

Examples:

\[
\nu\nu + V + \text{PAST}_3\text{SH} = \nu\nu \text{क} \text{रत} / \nu\nu \text{क} \text{रत}
\]

\[
\phi\phi + V + \text{PAST}_3\text{SF} = \phi\phi \text{क} \text{रत} \text{ोर}
\]

\[
\sigma\sigma(\text{\text{"}}) + V + \text{PAST}_3\text{SM} = \sigma\sigma \text{क} \text{रत} \text{ोर}
\]

\[
\nu\nu \text{+ N+PL} = \nu\nu \text{क} \text{रत} \text{ोर}
\]

\[
\nu\nu \text{+ N+ACC} = \nu\nu \text{क} \text{रत} \text{ोर}
\]

\[
\nu\nu \text{+ N+PL_DAT} = \nu\nu \text{क} \text{रत} \text{ोर}
\]

**Figure 2 : GUI of morphological generator**

**Conclusion and Future work**

Morphological generator which is explained here is a novel approach. It is developed using a very simple and efficient method. This is not a language specific method so this can be applicable for all the morphologically rich languages. Using this approach currently we are developing morphological generator for Malayalam and Telugu. This system provides a vast application it’s used in noun declension, verb conjugation, automatic sentence generator and also in featured. This system is unique handles auxiliaries and clitics, it does not require any spelling rules and this methodology can be implemented for any language. This work can be further used for implementing morphology based translation system between any language to Tamil.
References


Examining verbal forms inside the Tēvāram, in the light of the vocabulary found inside the 9th chapter of Cēntan Tivākaram

Dr Jean-Luc Chevillard
CNRS, University Paris-Diderot
Paris7, UMR7597 (HTL)
jean-luc.chevillard@univ-paris-diderot.fr

Introduction

The goal of the present communication is to examine a feature of a possible vocabulary data base, which could make more precise our knowledge of the history of Tamil vocabulary across the centuries, and especially that part of the vocabulary which consists of verbs. The intended final target of the article is to present information structured in such a way that it would be possible to make a comparison between two texts:

one of them, the Tivākaram, is the most ancient Tamil Thesaurus (or kōśa) and belongs to technical/scholarly literature

the other one, the Ėñ Tīrunūṭi, alias Cuntarar’s Tēvāram, belongs to devotional literature.

Due to space and time limitation, an effective comparison remains of course a distant goal. Moreover, I start by providing the reader with a relatively detailed presentation of the Tivākaram, a text probably not known well enough inside and outside Tamil Nadu, before examining the presence (or the absence) of the vocabulary of the Tivākaram inside Cuntarar’s Tēvāram and comparing this with what is found in other texts.

Structure and function of the Tivākaram

The Tivākaram, (also called Cēntan Tivākaram), is an object probably not very familiar to modern users of Tamil, unless they have received special training from a traditional Tamil scholar. However, some of them, having lived in the English-speaking world might in fact be familiar with some version of a standard reference work that has a lot in common with the Tivākaram, namely the “wordfinder” known

1. As regards the second one, there was no special reason for selecting it, except that it was available to me in digital form (See Subramanya Ayyar at alii [2007]) and that the two texts are probably not very distant in time.

2. Another possibility might be their having read some of the articles written by G. James, like for instance James[1989].
as Roget’s Thesaurus.3 The resemblance can be hinted at in the following way: whereas we can see inside
Roget’s Thesaurus an entry such as:

- (1) #360. Death. – N. death; decease, demise; dissolution, departure,[...] V. die, expire, perish;
meet one's death, meet one's end; pass away, be taken; yield one's breath, resign one's breath;
resign one's being, [...].4

we find, similarly, inside the Tivākaram, an entry which reads:

- (2) T1674 tuñcal, poŋ̥al, viṅtal, vītal, muṅcal, tuvaŋ̥al, muṭivu, ulattal, viṅtal, māytal, iṅtal, māṅl,
taputal, cātal emmac cāṟṟum pulavar.

Modern editors have added intermediate titles (or labels) to the Tivākaram and this cúttiram is labelled by
them “(cātalip peyar)”, which seems to make it clear that in their view, T1674 is an enumerations of 13
quasi-synonyms to cātal “to die”. I use the expression “quasi-synonyms” because just as English
speakers will recognize that “to expire” or “to perish”, appearing in 1, are not exact synonyms of “to
die”, similarly the users of Classical Tamil know that the 13 terms enumerated in 2 are not exact
equivalents of cātal, although the “nuances” are not easy to render into English.6

Additionally, it must be added that although this article concentrates on verbal forms, the
Tivākaram is basically an enumeration of nouns, many of which are proper names whereas others are
verbal nouns, common nouns, etc. And it seems clear that the practical synonymy problem is not the
same when it concerns the names of a God,7 the names of botanical species,8 and when it has to do with
what the title of the 9th chapter in Tivākaram describes as ceyal pariyya peyart tokuti, our main target here.

**Organization of the 9th chapter in Tivākaram**

I shall now try to give a concise characterization of that 9th chapter of Tivākaram (henceforth 9T), as it
appears in the 1993 critical edition (Madras University). The main facts concerning it which must be
mentioned in a preliminary abbreviated description are the following:

---

3 The most complete presentation of Roget’s Thesaurus which I know of is Hüllen [2004], *A History of Roget’s
Thesaurus*. It must also be noted that Knuth’s 1993 *Stanford GraphBase* contains a section devoted to Roget’s
Thesaurus.

4 This small citation is excerpted from the file available on the Project Gutenberg server (Edition 15a), itself
based on the 1911 edition.

5 This reference, like every other reference to the Tivākaram in this article, is to the 1990-1993 two-volume

6 The primary meaning of the first item (tuñcal) is “to fall asleep”; but readers of TC17c know that tuñćiḏūr
“they fell asleep” is a well-known euphemism (takuti) for cettār “they died” (see Chevillard [1996:65]).

7 The Tivākaram is divided into 12 chapters and the first chapter gives the names of Gods. For instance,
cúttiram 2 provides us with a list of 64 names for Civa: “caṅkaraṇ, iṅyioṇ, kāma takarṇaṇ, karṅkā
taraṇ, [...]”.

8 The 4th chapter of Tivākaram is called *Marappeyart Tokuti* As an example, its 78th cúttiram (alias T705)
provides us with a list of 5 alternate names for the plant generally called tāḷai: “kaṇṭai, kēṭakai, maṭi, kaitai, muṭarkal,
vanti iṅr tāḷai eṇa vakut taṇarē”.

148
1) 9T contains 213 cūttiram-s (numbered from T1558 to T1770 in the critical edition) and many of those entries belong to the type illustrated in 2. They are accompanied by a concluding verse (T1771) which is a praise of Cēntaṉ.

2) In the modern editions and in the first (1840) printed edition, each of the 213 cūttiram-s is preceded by a title, which has the form “X-ṉyeyar”, where X is generally a verbal noun or a verbal complex expression.

3) Item X, called by me in this article “primary entry”, is found again in the cūttiram itself, where it is accompanied by other items, which I shall call “secondary entries”. For instance, in T1674 (see 2), the primary entry is cātal and the secondary entries are the 13 items starting with tuñcal, poṉṟal, viṭṭal, etc. In what follows, locations of primary (resp. secondary) entries will be marked by “P” (resp. “S”).

4) The relationship between a primary entry and an associated secondary entry seems to be that the latter can be used (at least in certain contexts) as an approximate synonym of the former.

5) The total number of secondary entries associated with the 213 primary entries found in 9T is 673, which gives us the approximate figure that each primary entry is associated, on the average, with a little more than three secondary entries.

6) It must be noted that in the count of 673, some items are counted several times, notably because the same secondary item can be associated with several distinct primary items. More details will be given later on this feature.

Since this article is concerned with Tamil verbs in Classical Tamil, I do not take into account in this study the items which are not lexemes classifiable as verbs. I shall only consider, in this preliminary exploration of the verbal vocabulary contained inside the Tivākaram, the primary and secondary items in 9T which end with a few clearly recognizable verbal suffixes, namely:

7) 197 items ending in -tal or variants of -tal: akattītal [T1632], akāṟl [T1585], acaital [T1612, T1736]), ...

8) 157 items ending in -ttal: akaittal [T1575, T1672], acaittal [T1682], ...

9) 199 items ending in -al: aṅkai kuvittuk koṭṭal [T1765], aṭal [T1672], aṉṟkal [T1675], aṉṟl [T1607], ...

---


10 Exceptions are entries like 9T-1 (=T1558 “toliṉi peyar”), 9T-2 (=T1559 “toliṉpayilvi peyar”).

11 I have not verified whether those “titles” are found in manuscripts.

12 This means that the location of cātal (resp. poṉṟal) will be given as PT1674 (resp. MT1674).

13 I shall not try to demonstrate this fact and take as a working hypothesis that this is the “raison d’être” of a book like the Tivākaram (N.B. this assertion does not concern chapters 11 and 12).

14 That means that I shall not include in the preliminary data base primary items such as kōpan (T1581) “anger” or tuṉṟkai (T1762), which is the name of a dance.
In the remaining part of this article, I shall concentrate on that subset of 553 items, examining the way it has to be normalized, in order to become usable for a confrontation with a textual corpus such as the Tivākaram.

**Normalization of the 9th chapter in Tivākaram**

As already indicated, some items are found in several cūttiram-s of 9T. I include in this “multiple occurrence” phenomenon the cases where the same lexeme occurs, for metrical reasons, under distinct but equivalent forms such as "āptal and "āpttal." Additionally, it must be noted that a few cūttiram-s in 9T are preferably understood as explanatory definitions: such is the case for instance for T1762, in which tuṟṟkai is explained or T1766 which provides another designation for the action of “ārttu vāy puṭṭtal.”

If we set those cases aside, we are left with a core of 442 Tamil verbs, among which:

- 102 items occur in exactly 1 cūttiram, where they are the primary entry: they will be said to have profile “P”
- 2 items occur in exactly 2 cūttiram-s as the primary entry: they have profile “PP”.
- 260 items occur in exactly 2 cūttiram-s, as a secondary entry: they have profile “S”.
- 50 items occur in exactly 2 cūttiram-s, as a secondary entry: they have profile “SS”.

More complicated profiles are also found, because some items appear both as primary entry in one cūttiram and as secondary entry in another one. The exact figures are as follows:

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>PP</th>
<th>S</th>
<th>SS</th>
<th>SSS</th>
<th>PS</th>
<th>SP</th>
<th>PPS</th>
<th>SPP</th>
<th>SSP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>102</td>
<td>2</td>
<td>260</td>
<td>50</td>
<td>8</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>442</td>
<td></td>
</tr>
</tbody>
</table>

**Chart 1: Profiles of Verbal items in the 9th chapter of Tivākaram (9T)**

15 Ulterior versions on this exploration should of course take into account other verbal items such as: (1) 24 items ending in -uc: [āiyuu [T1619], ciāiyuu [T1710], kāṭyyu [T1587], ...]; (2) 8 items ending in -ppu: [iyāippu [T1710], uṭppu [T1615], ...]; (3) 7 items ending in -ci: [cāci [T1722], ścīci [T1618], ...].
16 Variants which I have noted are: əptal/əpttal, uṇapal/əṇapal, uṇakali/tulakali, ərttal/ərtatul, ərtatul/ərttal, ərttal/ərtatul, uṇapal/əṇapal, nārțtal/nărțatul, nīrktul/nīrktatul, nīrktatul/nīrktul, nīrtal/nīrta, uṇātul/əṇātul.
17 T1762 reads: “muṇṭkkai irukai paṭṭputtu orrti tūkkiya naṭṭiyatu tuṟṟkai ūkum” and T1766 reads: “āvalaṭi koṭṭal, ārttuvāy puṭṭtal.”
18 This list of 102 items starts with: akatttal (PT1632), aṭṭtal (PT1669), aṇakal (PT1607), avittal (PT1744), aṭṭtal (PT1652), aṭṭtal (PT1678), aṭṭtal (PT1580), aṇartal (PT1650), aṇaṭtal (PT1645), arṇyttal (PT1619), u[..] (PT1635), ...
19 Those two items are: kalatul (PT1643 & PT1723) and neruktal/naruṭkal (PT1591 & PT1671).
20 This list of 260 items starts with: akal (ST1585), acacitall (ST1682), aṇakal (ST1675), aṇaṭtal (ST1737), aṇaṟṟkal (ST1722), aṟṟtal (ST1567), amartal (ST1735), amakal (ST1591), amattal (ST1733), ayartal (ST1560), ayavarttal (ST1695), arital (ST1645), aruṛkal (ST1570), aklutal (ST1613), alarkarttal (ST1567), alartal (ST1731), avaittal (ST1728), aṭṭkal (ST1723), ...
21 This list of 50 items starts with: akaittal (ST1575 & ST1672), aṭṭtal/əṭṭtal (ST1672 & ST1675), aṇamattal (ST1607 & ST1722), aṇaṭtal (ST1569 & ST1570), aṭṭtal (ST1627 & ST1692), əṭṭal (ST1686 & ST1752), aṛttal (ST1682 & ST1686), aṛttal (ST1570 & ST1612).
22 Due to space limitations, the detailed content of each category will be made available in an appendix, contained in an online publication.
The Tivākaram seen as a semantic graph

Having thus provided some elementary statistics on the verbal forms which are found inside the 9th chapter of the Tivākaram,23 I shall now try to provide another perspective, concentrating on the cūttiram-s in which the verbal forms occur. As already said, the co-occurrence of two verbal nouns in the same cūttiram indicates an approximate synonymy: there are cases when the two verbs can be used to convey the same notional kernel, which is what the cūttiram itself shall be said here to represent. On the other hand, when a verbal noun is found in several cūttiram-s, it means that the verb is polysemic, and can be used (in different contexts) for expressing several different notions. This complexity can be represented by a graph G in the following way:

- each cūttiram of 9T in which there occurs one of the verbal nouns which we examine is said to be a vertex of graph G
- each time two cūttiram-s possess a verbal noun in common, we say that an edge of graph G connects the two corresponding vertices.

According to my calculations, the number of vertices to be considered here is 156 and the number of edges is 98. Additional statistics are as follows:

- The graph G contains 84 components, but 70 of those components contain only one element.
- Among the small components, 9 contain 2 elements; 2 have 3 elements and 2 have 4 elements
- There is a huge component containing 54 elements, and inside the huge component, seven 3-cliques are visible.

I have prepared a planar representation the graph G, but it is of course, due to space limitation, not possible to include it here. I shall content myself here with providing a small extract from the huge component, considering the fragment which contains the vertex T1575 (VAĻAITTAL) and the vertices which are connected to it through edges.

![Fig.1 A fragment from graph G](image-url)

23 Since this fact has not yet been stated, it should be said that verbal lexemes are also found in other chapters of the Tivākaram. See e.g. T1799 (aṭittaliṭṭeṭṭar). I have not however tried to count them.
In this fragment, the vertex T1575 represents the notion of “BENDING” which seems to be the common denominator, from a semantic point of view, between the various verbal nouns appearing inside cūttiram T1575, namely:

\[(3) \text{ ocittal, kōṭṭal, oṭṭal, akaittal, }\]
\[\text{ niemirtal, īpṭtal, vaṭṭtal ākum (T1575)}\]

Additionally, the element \(\text{ Đạickhōkāḥ} \) which appears on top of the edge connecting T1575 and T1672 (ALAIPPU) corresponds to the fact that cūttiram T1672 also contains the same element.

\[(4) \text{ ulaittal, oṭṭal, tīrttal, aṭṭal, migiceyal }\]
\[\text{ akaittal, vētaṇṭi, alaippu eṛu moḍha (T1672)}\]

**An elementary test and some questions**

After those premiminary explanations, we are now ready to present a few elements in order to give an idea of what could be attained by a systematic exploration of a corpus like the Tēvāram on the basis of the semantic graph which can be perceived inside the Tivākaram. Due to space limitation, and because of the huge nature of the task, I shall concentrate on one example, making use of the item examined above, namely VAŁAİTTAL (T1575). The basic idea of each elementary test is to verify whether all of the approximate synonyms are equally represented inside a given corpus of text. In the case of Cuntarar’s Tēvāram, on the basis of T1575 (with 1 primary entry and 6 secondary entries), the results are provided in column 2 of the following chart, where they can be compared with the results of the same test on the other parts of Tēvāram (attributed to Campantar and to Appar) and with the Puṇmāṉu as another element for comparison:

<table>
<thead>
<tr>
<th>Entry</th>
<th>Cuntarar (^\text{25})</th>
<th>Campantar</th>
<th>Appar</th>
<th>Puṇmāṉu (^\text{26})</th>
</tr>
</thead>
<tbody>
<tr>
<td>ocittal</td>
<td>NO</td>
<td>1-136_2: ocintu</td>
<td>4-39_9: ocitta</td>
<td>ociya “having broken” (80-8)</td>
</tr>
<tr>
<td>kōṭṭal</td>
<td>NO</td>
<td>?? 1-128 (kōṭṭīṉār)</td>
<td>4-10_(4): kōṭṭīṉār</td>
<td>cf. kōṭṭu (55-11)</td>
</tr>
<tr>
<td>oṭṭal</td>
<td>7-9_(10): ōṭkkum “it will destroy”</td>
<td>[2-24_(8): ōṭku “having been broken”]</td>
<td>NO</td>
<td>“to break”</td>
</tr>
<tr>
<td>akaittal</td>
<td>NO</td>
<td>NO</td>
<td>4-54_(1): akaitṭu</td>
<td>NO</td>
</tr>
<tr>
<td>niemirtal</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>īpṭtal</td>
<td>7-8_10: īpṭṭir</td>
<td>1-133_(7): īpṭṭavag</td>
<td>4-16_(10): īpṭṭār</td>
<td>“to break” (373-20)</td>
</tr>
<tr>
<td>vaṭṭtal</td>
<td>7-87_(5): vaṭṭita</td>
<td>3-106_(7): vaṭṭittu</td>
<td>6-42_(5): vaṭṭitu</td>
<td>158-1 (rare)</td>
</tr>
</tbody>
</table>

**Chart 2: one elementary test (T1575, VAŁAITAL “BENDING (or breaking?)”)**

\(^{24}\) The other VAŁAİTTAL (T1680) which appears on the graph is “SURROUNDING”.

\(^{25}\) The occurrences for the 3 Tēvāram authors were found through the concordance contained in the Digital Tēvāram CD-ROM (see bibliography).

\(^{26}\) The examples are based on V.I. Subramoniam [1962].
As should be clear from this chart, the result of this single test is of course highly inconclusive, but it is nevertheless an occasion for asking difficult and unavoidable questions. Based on more tests, which I cannot show here because of space constraints, my impression is that rare (or archaic?) synonyms are found more often in Campantar and Appar’s Tēvāram-s than in Cuntarar’s. That may be due to the fact that we have only 100 hymns by Cuntarar whereas we are having 385 hymns by Campantar and 312 by Appar, and that some rare words appear only when a corpus is large enough. This might however also be an indication of changes in the lexicon.

Additional questions concerns the meaning: How can we know what was the meaning intended for each word by the person who composed the Tivākaram, especially since it did not reach us with a commentary? And would those original intended meanings be relevant here, if we knew them for sure? In particular here, since some of the examples quoted here concern an action of bending to the point of breaking or crushing and since others concern an action of bending without breaking, my working hypothesis that all elements in T1575 (see 3) are approximate synonyms may in this case appear as partly problematic, but which other explanation could be given?

These questions being asked and being the basis for further exploration, I would like in conclusion to state that a parallel exploration of traditional lexicographical tools and literature can also contribute to a better understanding of the way a thesaurus like the Tivākaram was made (on the base of which compilation was it prepared?) and of the reasons why the need was felt later to create new tools, such as for instance the Piṅkalam, its immediate successor in time. The answer to such questions can only come from a joint examination of many pairs (Thesaurus/literary work), like the one (Tivākaram/Tēvāram) which has been sketched here. And I intend to continue my explorations in that direction. Finally, I also would like to point to the fact that the influence which traditional works such as those have had on literary composition is probably underestimated in the modern world. The following quotation, excepted from the preface to a 1968 edition of Piṅkalam, gives food for thought.

The examples examined concern what Civan did to Irāvaṇa, what Māl did to a wild lime tree, how a mountain was bent, etc.
Bibliography


8. Subramoniam, V.I., 1962, Index of Puranaanuuru, Department of Tamil, University of Kerala.


குண்டு துண்டபிள்ளிக் நூற்றாண்டு

தலைவர் க. பார்வாஜ்
குமரை பரந்திப்பு, பக்கமுதுகை தொலைக் குடியாளர் பல்கலைக் கழகம்
புதுச்சேரி - 625 021
k_umaraj@yahoo.com  94872 23316

திரு. அவ்விலங்கு
முன்னாள், தொலைக்கால வாழ்க்கையின் நிறுவனம்
6, புதுச்சேரி வட்டார, விளையாட்டு – 5
akilan.rp@gmail.com  99657 34497

குறிப்பிட்டு
இரண்டுமார் தமிழில் “ஏற்ற திகடை அடையும் புத்தம்” பண்டையத்காட்டில் இருந்து அனைத்து முறையிலும் பக்காக்கும். இவ்வாட்கள் விழாக்களை அவள் விளகிக் கிளர்ச்சியில் குறிப்பிட்டுகின்றது. இரண்டுமார் தமிழ்ப் புத்தகக் கூறுவிகளுக்கு பிரிவர்க்கும் போது A Concordance is an alphabetical list of the Principal words used in a book or body of work with their immediate contexts”. தமிழ், தொலைக் குடியாளர், பல்கலைக் கழகம் தமிழ் நூற்றாண்டுவர் இல்லாமல் காணப்படும். இந்திய தமிழைச் சாதாரணமாகிய பிரெச்சர் நவீன வர்த்தக தமிழில் எச்சாரணான் காணப்படுகிறது. இந்திய தமிழைச் சாதாரணமாகிய பிரெஸ்சர் நவீன வர்த்தக தமிழில் எச்சாரணான் காணப்படுகிறது.

சொற்கள்
இரண்டுமார் தமிழில் பார்வாஜான் தமிழ் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டுமார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டுமார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டுமார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டுமார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டுமார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டுமார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன். இரண்டு�ார் தமிழில் கூறுவிகளை விளக்கம் நூற்றாண்டில் உள்ளவுடன்.

தொடர்புக் குறியீடுகள்
குரு கோர்கினி முனை பல்கலைக் கழகம் மூன்றாம் விழாக்களின் நூற்றாண்டு அகட்சி அழக்குப் புத்தகம்
akilan.rp@gmail.com  99657 34497

139 
முற்பாட்டில் தகுதி வழங்கவும் அறுவத்தில் அவசம் நம்பும் ராணுவத்தில் அபுர்வத்தின் காதலுக்கு முன்னர் பிறந்து வருவதற்கான முறை புரோச்கி தொடர்பான பொருளை இயற்றாது. எனவே, கம்பிகையில் குறுகியது உள்ளது. பின்னர் வந்து வரவு எனும் பொருளை பார்க்கவும் வேண்டும். இந்த பொருள் இல்லாமல் ராணுவத்தில் உள்ளார் என்னும் நிறுவனத்தில் பொருளை வந்து வரவு என்பதை குறிப்பிடுகிறேன்.

<table>
<thead>
<tr>
<th>குறிப்பிட்டு முடிக்கும் விளக்கங்கள்</th>
<th>குறிப்பிட்டு முடிக்கும் விளக்கங்கள்</th>
</tr>
</thead>
</table>
| 1. புதியதும் வரலாற்றும் குடும்ப வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறையில் அப்படி வரும் வழியாகவும் அவசமாக வருவதாக முறை�ென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் வழியாக வருமென் அப்படி வரும் 

156
V. I. Subramaniam also accepts this view and quote as follows “In Some places segmentation of the head words was problematic”. The principles of segmentation adopted are: If the units A and B mean X and Y, A is listed with meaning x and B with y: If A and B mean only Z then A and B are listed as one word. In some cases A and B having the gloss Z may yield on etymological or metaphorical guess, an isolable gloss for A and B. For example in the form 'Vilaimakal' = 'prostitute', which is treated as a single word, it is possible to assign the meaning 'price' and 'women' to 'vilai' and 'makal', the prostitute is a woman who
sells her body. Such investigations are eliminated in plotting the meaning. The conclusion is if the word is
lexicalized then it should not be splitted.
Abstract

Both computer scientists and linguists are working towards developing natural language tools, resources and real-world NLP applications. Both aspire to comprehend the aspects of the language that makes analyzing computationally tractable. Understanding the nature of language is as important as understanding the computing environment. Mutual understanding in this regard will lead us towards expected outcome.

This paper describes the nature of Tamil data in detail with a view to the requirements of computing in terms of...

1. Spoken vs. written nature of Tamil data
2. Sequential nature of Tamil data
3. Hierarchical nature of Tamil data
4. Multi dimensional nature of Tamil data
5. Integrated nature of Tamil data
6. Multi lingual nature of Tamil data

Dr. P. David Prabhakar
Associate professor of Tamil,
Madras Christian College, Chennai-600 059
tamilprofessor@gmail.com
பல்கண் மூன்று வகைகளும் பெறும்

அன்புகை அல்லது கேள்விப்படி தமது கேள்வியை உருவாக்கினால் பெறும் வகையில் நுண்ணற்றவன், சரியான நூற்றாண்டு பதவிக்கப்பட்டது உடை பெறாதும் தனித்தோன்றம், எண்ணற்ற ஆண்டு, வரும் நூற்றாண்டு பதவிக்கப்பட்ட நிலையில் தனித்தோன்றம்.

பெருக்க வகையில் நுண்ணற்றம் பெறும்

பெருக்க வகையில் பெறும் நிலையில் பொருட்களை எடுப்பதை உருவாக்கினால் பெறும் வகையில் நுண்ணற்றவன், சிக்கம் எடுப்பதை உருவாக்கினால் பெறும் வகையில் நுண்ணற்றவன்.

கேள்விப்படி வகையில் பெறும்

கேள்விப்படி வகையில் பெறும் வகையில் நுண்ணற்றவன், சிக்கம் எடுப்பதை உருவாக்கினால் பெறும் வகையில் நுண்ணற்றவன்.

அடிப்படையில் பல்கண் மூன்று வகைகளும் (Speech Recognition)

அடிப்படையில் பல்கண் மூன்று வகைகளும் (Speech Synthesis).

பல்கண் வகையில் அடிப்படையில் பல்கண் மூன்று வகைகளும் (Diglossia) இரண்டு வகைகளும். பம்பிருந்து வியத்தில் பல்கண் மூன்று வகைகளும் கேள்வியை உருவாக்கினால் பெறும் வகையில் நுண்ணற்றவன்.
கூறு பின்பு பொய்யூடாய், பிற்கு காண்க அதனருக்கு அச்சந்தை நின்று, ராணை, பிளுவாக், இன்று பல்வேறு வகைகளை கழிக்கிறது கல்லக வல்லாம்பதாரர்கள். அதில், சுள் தருணத்தை உரை நெறியாக குறிப்பிட்டு பல்வேறு கொள்கைகளை, பல்வேறு கற்போட்டிகளை குறிப்பிட்டு கூறு குறிப்பிட்டு குண்டிகள், பொழியையும் பின்னால், பல்வேறு திற்கு பொழியையும் பின் குறிப்பிட்டு குண்டிகள் தொடர்பட்டு அரியையும். நீங்கள் குறிப்பிட்டு தொடர் தொடரும் (Intonation) அழைத்த குறிப்பிட்டு (1987) 67 கல்லக மாசு குறிப்பிட்டு விளக்கும். இதிலிருந்து இடையில் பின்னால் வீடியோ (Aspirated) அச்சந்தை வீடியோ (Voiced) பொழியையும் வீடியோ கண்டுபிடிக்க அத்துடன் காண்பது கரியில் மற்றும் கரியில் தொடர் தொடரும் அச்சந்தை குறிப்பிட்டு.

குறியீடுகளின் புரிந்து (Sequential nature)

நீங்கள் தொடர்ந்து பொழியையும் கூறு வழிகாட்டும் அச்சந்தைகள். தொடர்ந்து பல்வேறு வகைகளை குறிப்பிட்டு நீங்கள் குறிப்பிட்டு நீங்கள் குறிப்பிட்டு நீங்கள் குறிப்பிட்டு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு. தொடர்ந்து குறிப்பிட்டு நீங்கள் குறிப்பிட்டு நீங்கள் குறிப்பிட்டு நீங்கள் குறிப்பிட்டு நீங்கள் குறிப்பிட்டு நீங்கள் குறிப்பிட்டு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு.

புதியான புரிந்து (Hierarchical nature)

நீங்கள் அதனருக்கு குறிப்பிட்டு வழிகாட்டும் குறியீடுகளின் புதியான புரிந்து நீங்கள் புதியான புரிந்து நீங்கள் புதியான புரிந்து நீங்கள் புதியான புரிந்து நீங்கள் புதியான புரிந்து அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு.

குறியீடுகள் பொழியையும் காண்பது காண்பது காண்பது காண்பது காண்பது காண்பது காண்பது காண்பது காண்பது காண்பது (Discourse) பொழியையும் அச்சந்தையும். இதற்கு என்று ஏனைய உறுமுறைகளும் குறிப்பிட்டு என்று.

குறியீடுகளில் இம்முறை நீங்கள் குறிப்பிட்டு பல்வேறு அச்சந்தைகளை குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு அதனருக்கு. இந்த குறிப்பிட்டு வழிகாட்டும் குறியீடுகள் முடித்து இந்த குறிப்பிட்டு வழிகாட்டும் குறியீடுகள் முடித்து இந்த குறிப்பிட்டு வழிகாட்டும் குறியீடுகள் முடித்து இந்த குறிப்பிட்டு வழிகாட்டும் குறியீடுகள். இந்த குறியீடுகள் முடித்து இந்த குறியீடுகள். (பிரிவுக்குச் செய்யப்பட்ட 2002)
Multi dimensional nature

பல்கரையுடன் (Multi dimensional nature)

வழிமாற்று விளக்கம் பல்கரையுடன் பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை. இந்த சான்றியங்கள் தொண்டு விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை. இந்த சான்றியங்கள் தொண்டு விளக்கத்தை பல்கரையுடன் விளக்கத்தை பல்கரையுடன் விளக்கத்தை.}

Integrated nature (Integrated nature)

பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரை�ுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பல்பரையுடன் பalarm.
Multi lingual nature

Alan Garnhan 1998 Artificial Intelligence: An Introduction, Rutledge, London,
Christopher, S Butler (Ed) 1992 Computer and Written Texts, Block Well, Oxford, UK
2009 ‘Computer Analysis of Tamil Verb forms, International Tamil Internet Conference, Germany
Helen Amistary Dry (Eds) London 1998
FaceWaves : A Tamil Text to Video Framework

Madhan Karky, T V Geetha & Ravi Varman
{madhankarky@gmail.com, tv_g@hotmail.com, rvmsdh@yahoo.co.in}
Department of Computer Science & Engineering
College of Engineering Guindy
Anna University

Abstract
This paper presents FaceWaves, a framework for interactive information interchange for Tamil Internet and mobile users. A text-to-video subsystem for generating faces and animating them purely based on textual descriptions is proposed. Using textual descriptions to create a face by describing the visual features of a person and animating their face to speak out given information, can provide an efficient means of storing/transferring an animated video. The Text-to video subsystem comprises of a morphological analyser, ontology of facial features and expressions, Tamil text to speech, lip synchroniser, and emotion handler. To best of our knowledge, this is the first framework proposed for interactive information interchange with sophisticated language technology tools such as text to speech, text to video and voice to text. This framework facilitates storing and transferring video files over a network as plain text and converting the text to a video with a lightweight local client. This paper describes various components of the framework and shows the results from text-to-face generation module. The paper concludes discussing the results, opening a new area for the Tamil computing research community.

Introduction
Increasing number of Internet users, exponentially growing content and limited available bandwidth [1] has always been a problem to the Internet community. The number of information sources in Tamil and the number of Tamil users who contribute and consume is increasing every hour with the advent of blogs, microblogs and social networks. Almost every Tamil newspapers around the world now have their own portals feeding news and articles in Tamil every hour. We see bots collecting news from multiple sources and summarising the news. Auto-Journalists are already collecting sports scores from a website and generating a full fledged human like report about the match discussing performance of players purely based on numbers. Tamil is very soon going to adapt to these new growing technologies.

We foresee the need for human like auto-reporters who can read out a given news article, or can summarize any information source. We foresee the need for virtual faces that can listen to user queries and respond with answers. We foresee a system that can create a set of characters, backdrops and animate them from plain Tamil textual descriptions. In this paper we propose FaceWaves, a framework that enables creating such videos that can be transferred over any network as plain text and converted to video using a lightweight local client. This framework provides an ideal platform for transferring videos across world wide web and also in mobile networks.
This paper is organised into five sections. The second section discusses the background and motivation of this paper with related literature. The FaceWaves framework and components are given in section three. Section four provides the results from our face generator module of the FaceWaves framework in detail with snapshots from our GUI. The fifth section concludes the paper with our work in progress and future research in this domain.

Background

Tat Seng et al., proposed a animation sequencing method from text which tries to build 3D animations from a manual [2]. The proposed system tries to capture the concepts from the manual and extracts the features to provide the user with the option of choosing between textual or graphical modes. Narichika et al., in 2009 proposes a similar system that generates a movie from a script to a TV like program [3]. In this work the authors describe a framework that creates agents and a prototype that demonstrates controlling the generated agents. We find this work closely related to our proposed FaceWaves framework with respect to creating agents from text. A few Tamil Text-to-speech systems exist [2,6,7] and FaceWaves uses an improved variation of a text to speech system presented in [2]. Tamil speech to text has various proposals such as [3] and a fully working system yet to be demonstrated. Detecting emotions from a given text has been carried out in different languages [9-11] and the FaceWaves framework proposes an emotion identifier and applying emotions on a computer-generated face.

FaceWaves Framework

Face Waves architecture comprises of three main components. Information System, Wave Processor and Language Tools. An Interface Manager will be developed as a web tool, integrating all services to be accessible by clients. The components of the system are described briefly in the following sections.

![FaceWaves Framework Diagram]

Fig 1: FaceWaves Framework
a. Information System

Information System is responsible for crawling textual information on various topics over a structured information architecture such as Wikipedia(Tamil). The crawled information is processed and conceptually indexed using CoRe, a concept and relation based indexing system for large text collections. One key advantage of such an method would be to have a language independent indexing. The Information Manager will be responsible for crawling, processing, enconverting Tamil text to Universal Networking Language graphs, indexing the concept graphs, ranking documents based on concepts, retrieving and formatting results as requested by the Interface Manager.

b. Language Tools

Language Tools subsystem offers Tamil language dependent services such as morphological analysis, generation, named entity recognition, word sense disambiguation and sentence parser. These tools will be integrated into a system as a service for Wave Processor and Information System.

c. Wave Processor

Wave processor comprises three major units. Text to Speech, Speech to Text and Text to Video. These units work independent of each other and output of one unit can be piped as input to the other. The key responsibility of this system is to provide interaction elements (speech, agents, and agent animations) for the information interchange service provided by system. The wave processor makes use of language tools such as Tamil Analyser and Generator for various applications. The Wave Processor will be designed as a separate sub-system that can be plugged into different standalone, web or mobile applications.

d. Interface Manager

The interface manager acts as the central control object for FaceWaves. It holds the responsibilities of receiving queries from clients to create agents, displaying agents, collecting information as voice queries and sending to Wave Processor and converting the wave to text and sending the text to Information system to retrieve information and resend the information to Wave Processor to convert them to speech and agent animation and finally to play the animation video on the client’s interface.

Text To Video Subsystem

The Text to Video subsystem of FaceWaves framework, as depicted in figure 2, comprises of a Document Processor, Face Generator, Backdrop Selector, Emotion Processor and a Movie Manager.

A plain text, semi-structured script file comprising of character descriptions, scene descriptions and dialogues is processed by Document processor to extract the corresponding sections and formats them structurally and routes the corresponding formatted information to Face generator, Backdrop selector and Emotion Processor. The face generator uses a face description ontology to describe dimensions of different parts of the human face for the given description. Backdrop selector uses a background library and analyses the scene description to rank and choose appropriate background for the given scene. Emotion Processor analyses the text and tags the dialogues with appropriate emotions using an Emotion ontology. The Face description, selected backdrop and annotated dialogues are then sent to the Movie manager which produces a plain text movie description file which can be sent back to the client. A lightweight movie player in the client on receiving the Movie Description, processes the description and
uses a Tamil Text to Speech module and synchronises the lip movements with the generated speech and face with appropriate emotion with the selected backdrop. The Movie Player generates a full length running animation movie with synchronised subtitles.

![Text to Video Subsystem](image)

**Fig 2 : Text to Video Subsystem**

**Results : Face Generator**

This section discusses the Face Generator module and how plain text Tamil descriptions are converted to a human face. The Face Generator module as explained in the previous section uses a face description ontology along with a Morphological Generator. A sample face description can be in a free flowing form such as

```
aadhi adarthiyaana puruvangaLum neeLamaana mookkum akaNda
kankaLum siRiya uthadugaLum koNdirunthaan
```

Aadhi has thick eyebrows, long nose, big wide eyes and small lips
The image generated for the face description provided above is provided in figure 3 as a snapshot from our text-to-video GUI.

Fig 3 : Face Generator Snapshot 1  Fig 4 : Face Generator Snapshot 2

The snapshot in Figure 4 was generated for the following face description.

Dhiyaa has a small nose, thin eyebrows, big eyes and small lips.

The face generator module was developed in Java. The generator gets the face description text as input and uses a morphological analyser, sentence parser and face description ontology to retrieve the facial features from the given descriptions. A neutral face is initially described as a collection of objects one for each part of the face. The member variables of each object, which define the dimension of the corresponding part of face, are modified based on the adjectives describing each part in the face description.

**Conclusion and Future Work**

The current face generator module does not take into account the descriptions for hair, skin colour and texture, moustache, special marks or any three dimensional descriptions. Including these features in the face generation module will be our future work along with integrating the face generation module with adding expressions to the face based on the emotion identified from the dialogue. This paper provides the overall FaceWaves framework and describes one particular subsystem, Tamil text-to-video, and one key module of the subsystem, Face Generator. We believe that this paper will open numerous research problems for the Tamil computing community.
References


Abstract

ThiruKural is a discourse on the art of living, a set of healthy principles of guidance for the variety of segments of the civilization for a pleasant-sounding combined living. Each episodes and couplet(kurals) in Thirukural are related to the real time world. In the Tamil language "Thiru" means "holy" or "sacred," and "Kural" means anything that is brief or short. This paper focuses on context based searching for Thirukural and describes the techniques of Natural Language Processing (NLP) which is to design and build software that will analyze and apply in both understanding and generating natural languages that humans use naturally. A fundamental phenomenon of natural language is the variability of semantic expression. The system should understand the short story of some sentences like paragraph and result kurals with given relative meaning. Based on the given context, it retrieves the information of couplets (kurals) with the ranking priority. The system identifies the relative terms and using the term frequency calculates the weightage for each terms.

Keywords: Thirukural, Natural Language Processing, Couplets, Semantic expression, Term frequency.

Introduction

In this information search is based on the Natural Language Processing. NLP is a field of computer science and linguistic concerned with the interactions between computers and human natural languages. Two fields of NLP are: Natural language generation, system converts information from computer database into readable language and Natural language understanding, system converts human language into computers known format. The system participating in this competition must do something more than the system from other NLP competitions: to prove capabilities of understanding how language works.

Same meaning can be expressed by, or inferred from different texts mapping between language expressions and meanings. Humans use different expressions to convey the same meaning. Therefore, numerous NLP applications, such as, Question Answering, Information Extraction. Summarization require computational models of language that recognize the semantic approach. Trying to capture the major semantic inferences needed to understand equivalent semantic expressions. There are several levels of the meaning of the texts, ranging from shallow level to deep one. But, it is still difficult to make a consensus on how to describe the deep meaning.

In Question and Answering, an indexing is considered as the predictive annotation. Text Meaning Representation forms the Tamil dictionary which contains extended relative meanings. Information Retrieval must be started when the queries enters for searching couplets.
Literature survey

In Natural Language Processing, denotes as "Understanding" language means, among other things, knowing what concepts a word or phrase stands for and knowing how to link those concepts together in a meaningful way [6]. Basically Communication between Human and the system starts from Question and answering [3] and the main goal of QA is, whenever the user types a question, system must produces the correct answer. The Challenge is to analyze the questions, gathers information and presents the answer.

Question and answering is also related to text summarization technology [4], it contains short questions and answers. Sometimes there will be no answer for questions or multiple answers for questions, which have no answer object in documents to a given question or there are many answer objects. When a new question comes up, the system compares it with all the questions in the library and finds the most detailed standard question to match it. Then the corresponding answer is returned to the user [2].

While in Information Retrieval, An efficient indexing mechanism is normally used to quickly retrieve the information. And [7] Question Answering and information Retrieval are the two candidates of Natural Language Processing. An Indexing is used for searching, or finding the respective information. And the meaning represents the TMR language [1], which is the knowledge representation system for representing text meaning. An Onomasticon, or lexicon is nothing but the collection of proper names and terms; a dictionary.

Thirukural plays an important role of our human life[8] and the Greatness of Thirukural is a precious gem among the classics, unique in the deliverance of code of conduct to the mankind to follow for all time to come. This poem consisting of 133 sections of 10 couplets each which was predictable as a masterpiece of ancient literature in Tamil in its own times, has stood the test of history and is established by posterity as a decisive work which has predisposed the thoughts of man throughout the centuries. It is not only of great artistic and stylistic literary value, but also a direct to the art of living with pieces of precious wisdom.

How the search engine works

The context based Information search can be obtained by identifying the relativity among terms and derive a context from the terms. The term ranking can be calculated using Term frequency in the given sentence or paragraph of words.

Users query could be answered in semantic approach and get the couplets in ranking order. The above architecture diagram clearly explains the flow of the paper.

The system gets the input text as sentences like paragraph for searching information. The given sentence or paragraph is tokenized and each token is passed to Tamil language analyzer. The analyzer identifies the part of the syntax such as noun, verb, adjective, pronoun, adverb and so on that the token belongs to. The analyzer removes the common Tamil morphological. The system uses the noun, verb, adjective and adverbs to find the relation among the terms.

The relativity among the terms can be derived using a Tamil dictionary. The term weightage is calculated for each relative term using the term weightage formula. The context is extracted from the relative terms and finally searches the couplets in a Thirukural database.
**Term Weightage Formula**

By using Term Weightage Formula, calculates the weightage of the words which could be in any form as related meaning, different expression and comparative meaning. In a paragraph some of the sentences have same meaning but in different expressions by using term weightage formula, a well formed questions term having the weightage.

\[
t_w = \frac{nt_r}{T}
\]

where as,

- \( t_w \) – term weightage
- \( nt_r \) – number of related terms.
- \( T \)– Total number of terms identified in the given sentence or paragraph.

Calculate the total number of terms and the related meaning and the terms which have relative meaning.

For example, the following sentence shown as,

In the given sentence, the terms are analyzed and grouped into as proper names of all kinds and the origins of names.
List of terms:

- நூறாக்: Noun
- செய்யு: Verb
- அவசுறுவா: Adverb
- நூறாக்: Noun
- நூறாக்: Noun
- செய்யு: Verb
- பார்பு: Noun
- கெடும: Verb
- நூறாக்: Noun
- நூறாக்: Adverb
- செய்யு: Verb

[[நூறாக், செய்யு] [பார்பு] [செய்யு] [கெடும] [பார்பு] [செய்யு] [செய்யு] [செய்யு] [செய்யு] [செய்யு]]

Each term have their respective weightage and the related terms have more weightage than the remaining terms.

[செய்யு, நூறாக்] is the relative meaning of the above example and this could be calculated as

\[ \frac{2}{11} = 0.222 \]

Result as,

1. கெடுமையை இறுதியை விளக்கும் வருடாக்கம் உடல்கை இறுதியை முற்புண்டியது.
2. முடன்மை கூர்ந்த நூறாக் புள்ளிக்கட்டி விளக்கும் கெடுமை குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு.
3. தொழில் வளர்ந்து வந்த பொழுதுயிர் எனவும் வெளர்ந்து வெளர்ந்து.
Conclusion

Thirukkural is a treatise par excellence on the art of living and the real greatness of Thirukkural is its survival, even after the onslaufths of many heterogenous creeds. Each and everything in our life is related to the theme of Thirukural. By giving a paragragh the search engine analyse the context of the sentences and groups the related terms to find out the equivalent meanings of couplets in ranking order with the help of term weightage. This search engine helps the users to easily gather the information of couplets for any paragragh in the context based.

References

5. Mitsuru Ishizuka “A Common Concept Description of Natural Language Texts as the Foundation of Semantic Computing on the Web” The University of Tokyo.
8. “Tamil Virtual University” enable Tamil Education easily and effectively.
அங்கில ஒரு ஬வும் பார்க்காமை

ஒருசர் மங்களாய் பதிப்பாய் வைத்து

A Context-free grammar and a method to parse verses in metrical text

sundarbese@yahoo.com, ishwar@gmail.com

வார்ப்புரை மதிப்பிக்கும் அறிவுடைய விளையாடுகள் மாற்றங்கள் மற்றும் விளக்கங்கள் பிரயோஜனத்திற்குப் பயன்படுத்தப்பட்டுள்ளன. இதில் பல்வேறு வகையான விளையாட்டுகள் அறிவுடைய விளையாடுகள் மற்றும் விளக்கங்களைக் கொண்டுள்ளன. இவைகள் உரை விளக்கத்திற்குப் பயன்படுத்தப்படுகின்றன.

முழுமத்தில் உள்ள நோய்களின் விளக்கம் மற்றும் அறிவியல் விளையாட்டுகள் பயன்படுத்தப்பட்டுள்ளன. இதில் பல்வேறு வகையான விளையாட்டுகள் அறிவுடைய விளையாடுகள் மற்றும் விளக்கங்களைக் கொண்டுள்ளன. இவைகள் உரை விளக்கத்திற்குப் பயன்படுத்தப்படுகின்றன.

முழுமத்தில் உள்ள நோய்களின் விளக்கம் மற்றும் அறிவியல் விளையாட்டுகள் பயன்படுத்தப்பட்டுள்ளன. இதில் பல்வேறு வகையான விளையாட்டுகள் அறிவுடைய விளையாடுகள் மற்றும் விளக்கங்களைக் கொண்டுள்ளன. இவைகள் உரை விளக்கத்திற்குப் பயன்படுத்தப்படுகின்றன.

முழுமத்தில் உள்ள நோய்களின் விளக்கம் மற்றும் அறிவியல் விளையாட்டுகள் பயன்படுத்தப்பட்டுள்ளன. இதில் பல்வேறு வகையான விளையாட்டுகள் அறிவுடைய விளையாடுகள் மற்றும் விளக்கங்களைக் கொண்டுள்ளன. இவைகள் உரை விளக்கத்திற்குப் பயன்படுத்தப்படுகின்றன.
Linguistics and Philosophy


WE USE TEI


http://ta.wiktionary.org/wiki/node

179
புதுப்பிக்கத் தெரியாதோ?

புதுப்பிக்கத் தெரியாதோ?

1. பொருள்தான் நோக்கி பொருள்தான் நோக்கி பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும் அவர் புலனானாலும், பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும் அவர் புலனானாலும் பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும்

2. பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும், பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும் பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும

3. பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும, பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும, பொருள்தான் நோக்கியிட்டதற்கு "நன்மையை" என்று எந்தவுரையிலேயே அவர் கூறினானால் ஆணானாலும், புலனானாலும

14 http://pages.cpsc.ucalgary.ca/~aycock/spark/
15 http://groups.google.com/group/tamil_wiktionary/msg/30af374714abb5ba?


4.

1.

2.

3.

4.

http://groups.google.com/group/anbudan/browse_thread/thread/fc8c416a2b42f73a/00efb259006416d1?pli=1


[16] http://groups.google.com/group/anbudan/browse_thread/thread/fc8c416a2b42f73a/00efb259006416d1?pli=1
Computational approaches for learning inflections in Tamil

K. Rajan, Dept of Comp. Sc. and Engineering

V. Ramalingam, Dept of Comp Sc. and Engineering

M. Ganesan, Centre of Advanced Studies in Linguistics

Annamalai University, Annamalainagar

Abstract

This paper proposes machine learning techniques for the study of inflections in Tamil. In recent years there has been growing interest in using machine learning techniques for natural language processing tasks. Recent computational research on natural language corpora has revealed that simple machine learning mechanisms could make an important contribution to certain aspects of language acquisition. There are regularities that can be captured mechanically, by exploiting distributional patterns found in the language data. Very large corpora will yield to mathematical model of language systems.

The Artificial neural network (ANN) is used to learn the regularities of Tamil morphology. In this paper, the methods of feature extraction and feature representation for Tamil verb morphology are discussed. The ANN network with three layers is trained with the features extracted from different verb conjugations generated by the morphological generator. More than 7800 word forms are used for training and testing this model. The network produces the grammatical values as output for the morphemes present in the given word. This model does not require morphological rules. The output is comparable with the results of rule based morphological analyser. The performance of this model on different verb types is presented.

Keywords: Machine learning, Tamil morphology, ANN Language Model

Introduction

In recent years there has been growing interest in using machine learning techniques for natural language processing tasks. In this paper, the Artificial neural network is trained to learn the relationship between the morphemes and their categories from the training set of inflected Tamil verb forms. Morphemes are portions of a word that recur in other words with the same meaning. They are minimal, that they cannot be broken into pieces. The grammatical value of the whole word is related to the component morphemes of the word. In highly-inflecting and compounding languages the number of possible word forms is very high. This poses special challenges to Natural Language Processing systems dealing with these languages.

The morphological tagging is an important problem in the area of computational linguistics, as it underlies other crucial tasks such as syntactic parsing and machine translation. Artificial neural network is a promising approach to this kind of problem, for which the exact algorithmic solution is unknown or
not efficient enough. In this paper we present the results obtained by the application of neural network model trained with back propagation algorithm for Tamil morphology.

Knowledge discovery in text is a non-trivial process of identifying valid, novel, potentially useful, and understandable patterns in unstructured text data. Learning algorithms are an integral part of knowledge discovery. Learning techniques may be supervised or unsupervised. Supervised learning techniques enjoy a better success rate as defined in terms of usefulness of discovered knowledge (Maciej Majewski, 2008). **Machine learning** is the capacity of a computer to learn from experience (i.e., data) and to extract knowledge from examples. A successful learner should be able to make general conclusions about the data it is trained on. This allows it to act appropriately in new situations.

**Related works**

The supervised learning of morphological rules is based on the annotated data as the alignment of orthographic words with their morpheme representations. In order to apply supervised learning methods, the data should further be extended with information about inflectional classes and features (part-of-speech tags), thus making the output of the training mechanism compatible with the input for morphological analysers.

There is unsupervised learning of morphological rules. The advantage of unsupervised morphological learning is that it requires only the set of orthographic words (completely “raw data”) without any stem/affix lists or grammatical annotation. However, most of the unsupervised methods put restrictions on the number of morphemes per word, on rule complexity etc. The goal of these methods is merely splitting a word into stem and suffix without the capability of assigning any grammatical information. A number of unsupervised learning techniques have been applied: genetic algorithms (Kazakov, 1997), a minimal description length approach based on spelling of words and the set of suffixes that appear with each stem (Goldsmith, 2001), and the quasi-roots algorithm (Sheremetyeva and Nirenburg, 1999).

Many researchers have been working on similar morphological learning systems for Indian languages. One of the earlier rule based morphological analyser was developed for Tamil at CIIL (Ganesan M, 1994), S.R.Kolhe and B.V.Pawar have investigated the inductive inference of grammar of subset of Marathi, Pradipta Ranjan Ray presented a computational model for Bengali Morphological analysis using Finite State methods. A generic architecture for morphological generators of agglutinative languages has been proposed (Uma Maheshwar Rao,G, 2006), in which corpus is used to extract fully inflected word forms. Stochastic taggers like HMM will not give a high accuracy for Tamil because the language is inflectionally rich and is relatively free-word order (Arulmozhi et.al 2006). In their paper, they have discussed the algorithm for developing a rule based tagger for Tamil. No work has been reported on machine learning techniques using Artificial neural network for morphology of Indian languages. ANN is applied for text reasoning (Maciej Majewski et.al, 2008), Natural language processing tasks (Joao Luis Garcia Rosa, 2002),(J.L Elman, 1990), (Ahmed, 2002),(Rajan K et.al, 2002),(Quing Ma, 2003) and categorization of Tamil documents (Rajan K. et al, 2009).
Tamil morphology

The Tamil morphology is characterised as agglutinative or concatenative. That is, morphs are agglutinated or concatenated in a sequence after a stem. Concatenative morphology in Tamil involves always suffixation (Thomas Lehman, 1993). This is represented as

\[ \text{Stem + (affix)}^n \]

where the superscript \( n \) means one or more occurrences of a suffix. Morphs are concatenated as suffixes at the right of a word stem, to produce inflected or derived forms of words. In Tamil, there is another morphological process which is reduplication. All lexical or root morphemes are grouped into four major types. They are verbal, nominal, adjectival and adverbial roots. All other types of words can be identified as an inflected or uninflected form of these stems. Nouns can be inflected for the case and number. An inflected noun form may be the realisation of three morphemes, as given in the following representation.

Noun stem + [Plural suffix ] + [Oblique] + [Case suffix]

Verbs can be inflected for tense, person, number, gender and others.

Verb stem + [Tense Marker] +[ Verbal Participle Suffix] + [Auxiliary verb] +[Tense Marker]+[Person, Number, Gender]

Postpositions and adjectives cannot be inflected. The derivation occurs only in nouns. When morphemes or words combine, certain morphophonemic changes occur. Nouns can be inflected for the case and number. Inflected verbs in Tamil are finite or non-finite. Finite verbs mark both tense and subject-verb agreement, non-finite verbs do not. Finite verbs occur only in restricted contexts in the structure of a sentence; they typically mark the end of a sentence.

The lexical knowledge is characterised with respect to verbs. All the grammatical information that is, about tense, number, gender, etc is carried by the verbs in a sentence. These are the reasons why the study of verbs has acquired immense importance in Linguistics. Modern Tamil has three tenses. Each tense morpheme is realised by a number of tense suffixes or allomorphs. Tense suffixes are listed in Table 1.

<table>
<thead>
<tr>
<th>Tense</th>
<th>Suffixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>kiR, kkiR, kinR,kkinR</td>
</tr>
<tr>
<td>Past</td>
<td>t,tt,ndt,in,inR</td>
</tr>
<tr>
<td>Future</td>
<td>p,v,pp</td>
</tr>
</tbody>
</table>

Table 1. Tamil tense markers
The second inflectional suffix after the stem is a PNG marker. The table 2 shows the person, number and gender suffixes.

<table>
<thead>
<tr>
<th>Op</th>
<th>Number</th>
<th>Gender</th>
<th>Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>Singular</td>
<td>Een</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>Oom</td>
<td></td>
</tr>
<tr>
<td>Second</td>
<td>Singular</td>
<td>Aay</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>irkal</td>
<td></td>
</tr>
<tr>
<td>Third</td>
<td>Singular</td>
<td>Masculine</td>
<td>aan</td>
</tr>
<tr>
<td></td>
<td>Feminine</td>
<td>aaL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neuter</td>
<td>Atu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plural</td>
<td>Masculine/Feminine</td>
<td>aarkaL</td>
</tr>
<tr>
<td></td>
<td>Neuter</td>
<td>Ana</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Tamil PNG markers

Tamil distinguishes between 4 types of non-finite verb forms. The non-finite verb forms, except the infinitive have both positive and negative forms. Only the adjectival participle distinguishes tense. All other non-finite verbs are tenseless. Each of the non-finite verb forms is marked with a non-finite verb suffix, which is added to the verb stem, or to the tense suffix. The infinitive is formed by the affixation of infinitive suffix to the verb stem. The conditional verb form occurs both in a positive and negative forms. This is formed by adding the phoneme cluster of the past tense allomorph to the verb stem and then affixing the conditional suffix(-aal/-aaviTTal). Adjectival participle is formed by adding past or present tense allomorphs to the verb stem and the adjectival suffix (-a).

<table>
<thead>
<tr>
<th>Non-Finite</th>
<th>Suffixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infinitive</td>
<td>~a</td>
</tr>
<tr>
<td>Adj.Participle</td>
<td>t,tt,ntd,nt,in, inR</td>
</tr>
<tr>
<td></td>
<td>kiR, kkiR,um</td>
</tr>
<tr>
<td>Verbal Participle</td>
<td>t,tt,ntd, in</td>
</tr>
<tr>
<td>Conditional</td>
<td>~aal, ~aaviTTal</td>
</tr>
</tbody>
</table>

Table 3. Non-Finite verb suffixes

The suffix ~um morph realised the future tense morpheme and the adjectival morpheme. It is an instance of homophony of morphs in Tamil. The list of morphemes is given in the Appendix A.
Neural Network

Artificial neural networks are computational models based on biological neural networks. They can be used to model complex relationship between inputs and outputs or to find pattern in data. Neural network based approaches learn the associations of word-to-tag mappings, from a training data set and also generalise to unseen examples. Neural networks have self learning capability, are fault tolerant and noise immune.

In this paper, a three layer feed forward neural network with hyperbolic tangent (tanh) function is used in hidden layers followed by a linear output layer. The neural network is trained using backpropagation algorithm. A momentum term is used to achieve a faster global convergence. A bias value is used to enable each neuron to fire hundred percent.

![Figure 1. Architecture of ANN model](image)

The input layer is used to represent the 38 morpheme features of the input word. The output layer represents the grammatical values of the cluster of morpheme features. The structure of the ANN model is $38 \text{ L} - 30 \text{ N} - 15 \text{ L}$, representing the number of neurons/units in the input layer, hidden layer and the output layer respectively. The letters L and N denote linear and non-linear activation functions. In this study, the network is trained with a learning rate of 0.01 and 2000 epochs.

Feature Extraction

The input and output vectors are prepared from the word-forms generated from 208 Tamil verb stems using the morphological generator. The verbs are inflected for tense, person, number and gender. They are also inflected for non-finite forms. During the process of generating the various conjugations, their corresponding grammatical categories are also collected as output. The stems are combined with different morphemes of various categories by the generator. The morpheme table has 38 features. The categories are marked for 15 grammatical values according to the morpheme components added with the stem. This method has been utilised because, the corpus may not have all the combinations. Totally 7800 different word forms have been generated and stored in the form of input and output vectors. Among the 7800
samples, 5000 samples are randomly selected for training. The network is tested by using the remaining list of 2800 samples.

The binary representation of input and output vectors are created for all word-forms by scanning through all morphemes. The morphemes are listed in the descending order of their length. If a particular morpheme is found within the word-form, it is represented by 1 in the input vector and also the corresponding category value is set to 1 for the output vector. During testing phase, only the input vectors are created from the given word-form. The feature vector for the word is given by equation 1.

\[
\mathbf{w} = \{ \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \ldots, \mathbf{m}_n \} \rightarrow 1
\]

The word \( w \) is a sequence of \( n \) morphemes

Where \( \mathbf{m}_i = 1 \) if \( i \) th morpheme is present in the word \( w \), otherwise 0.

The input feature extraction from word forms during the test phase is same as the one followed during the training.

**Results and Discussions**

The proposed model of supervised learning for Tamil morphology shows the ability of neural networks to learn the morpheme to category association.

<table>
<thead>
<tr>
<th>Number of test Samples</th>
<th>Numbers Correctly recognized</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total 2800</td>
<td>2694</td>
<td>96.21 %</td>
</tr>
<tr>
<td>Finite Verbs 1900</td>
<td>1865</td>
<td>98.15 %</td>
</tr>
<tr>
<td>Non-Finite Verbs 900</td>
<td>829</td>
<td>92.11 %</td>
</tr>
</tbody>
</table>

Table 4. Performance of the ANN Model
The system produces correct categories for the morpheme sequences of all the inflected forms of finite verbs that are not part of the training set. The non-finite forms show lower recognition rate because of the shorter morphemes than for the finite verb forms. The table 4 shows the performance of the model on these two verb types.

**Conclusion**

The proposed neural network model is a powerful supervised machine learning system that can be applied effectively for any inflectional morphological language. This approach is cost effective, because it does not require any explicit grammar rules. The experimental result shows its promising performance for morphologically rich languages like Tamil. This work can be extended to cover other grammatical categories of Tamil, by changing the input and output vectors of the neural network.

**References**

Introduction

Natural Language Parsing (NLP) is a process by which any natural language text is analyzed for its components in terms of sentences, clauses, phrases, and words using the grammatical rules and the instructions for operating those rules. The major objectives of the parsing are, first to identify the grammatical function of different components in a sentence and second, to understand the semantic roles of different grammatical components. Efforts are being made to achieve the goal through a number of methods. Some of them are grammatical rule based and some are statistical approach. A number of grammatical models like Phrase Structure Grammar, Transformational Generative Grammar, Generalized Phrase Structure Grammar, Head Driven Phrase Structure Grammar, Lexical Functional Grammar, Case Grammar, Minimalist Programme, etc. are being used to represent the grammatical knowledge to the system. The choice of model mainly based on the grammatical system of the language in question. As most of the time the rule-based approach does not handle some of the exceptional structures in a language, and takes more time to represent the grammatical knowledge to the system, some of the statistics-based models like probability model, Neural network model, Hidden Markov model, Support Vector model, etc. are used. Text processing, understanding, and generation are the basic needs to achieve the long term goals like Machine Translation, Man-machine communication through a natural language, Text to Speech and Speech to Text system generation, Information Retrieval, Information Extraction, Question Answering system, etc. In this paper I discuss the criteria that I evolved for identification of various phrases, clauses and sentences The syntactic parsers that I developed work fairly well on modern Tamil corpora. The parser takes the text, which are tagged at word level, as input. The POS tags, which I have used are the standard to this parser.

Morphological Analysis

Tamil is an agglutinative language. The Morphology of Tamil is more complex. For example, a regular verb can be conjugated to around 1600 word forms. Therefore the composition of a sentence can comfortably studied at two levels: morphological and syntactic level. In morphological Analysis the main tasks for the analyzer are 1) to identify the boundary of a word, 2) to identify various morphs that constitute a word, 3) to label the grammatical category to each morph, and 4) to mark the Parts of Speech to the word. In order to achieve all these a set of Machine Readable Dictionaries (MRD) for stems and suffixes, and the morphotactics (i.e. the arrangement of morphs in different word forms) are basically needed. They represent the morphological information to the system for analysis words (For more details see Ganesan, 1994). The output of the Morphological Analyzer will be a word-level tagged text.

The morph level tags are marked in between two underscores immediately after the morphs. And the word level tags are marked inside an angular bracket ‘< >’ immediately at the end of the word.
For example

\texttt{aval\_pro\_<PNh> paaTTu\_abn\_<NN> paaTinaal\_vb, pst, 3sf\_<NN>}

‘She sang a song’

\textbf{Syntactic Analysis}

Normally, the syntactic parser makes use of grammatical rules and a dictionary consisting of words and their grammatical information. In any parsing programme, the lexical items within a sentence are first marked with their grammatical value, making use of the dictionary. Then, using different grammatical models and grammatical rules, the system parses a sentence and marks the phrases and clauses in the sentence.

The tasks involved for a syntactic parser are:

(i) to identify a sentence in a text
(ii) to identify various phrases and marking them in a sentence
(iii) to identify the head of a clause
(iv) to segment a clause in a sentence and
(v) to mark a clause with the grammatical information

For the identification of both clauses and sentences, first the text have to be tagged at phrase level. The phrase level tagging includes all the phrases, such as, noun phrase, verb phrase, adjectival phrase, adverbial phrase, quantifier phrase and postpositional phrase. The criteria for identifying the phrases are explained in detail in the report (Ganesan, 2010). For example, there are 14 rules given for the identification of different noun phrases. With these rules the system identified 98% of noun phrase correctly in the given text. The system fails only when two <NN>s marked for Nominative case occur consecutively. To solve such problems the system needs extra linguistics knowledge, which is difficult to provide in the knowledge base. Statistics-based approach can be used to resolve such problem. In this paper the criteria for identification of clauses and sentences are only discussed. We mark broadly five types of clauses in a sentence. They are:

i) Se – Embedded sentence
ii) Sm – Matrix / Main sentence
iii) Si – Interrogative sentence
iv) EQS – Equational sentence
v) Sel – Elliptical sentence

\textbf{Basic Principles}

The parsing described here is built on the principles explained below for the analysis of a sentence. They are:

i) The head of a clause always finds its place in the right most position of a clause. In other words, clauses in Tamil are left-branching.
ii) In a complex sentence where we have more than one clause, the arguments for a predicate (verb) occur before the verb of that clause and never in another clause.
Approach

For syntactic analysis, a sentence is the minimum unit. Identification and marking of a sentence is not very simple. Most of the time, the sentences are identified using the punctuation markers like full stop (.), question mark (?), colon (:), semicolon (;), etc. Apart from these punctuation marks, in a quotation, the quoted sentence and the quoting constitute are two different sentences. In the present study, a set of algorithm is used in the parser to identify a sentence.

Once a sentence is identified, the parser looks at the sentence word by word. This can be done either from the starting of a sentence or form the end of a sentence. In other words, a sentence can be approached either from Left-to-Right (L to R) or Right-to-Left (R to L). In the morphological analysis the words are approached from R to L. But in Syntactic parsing the sentences are approached from L to R. The reasons are as follows. The difference between the formation of a word and a sentence is that in word the suffixes are added to the right of the stem, whereas in sentence the arguments are added to the left of the head of a phrase or clause. The suffixes to a stem and the arguments to a head of a phrase / clause are dependent, but the stem (root/base) of a word and the head of a phrase / clause are independent unit. The dependent unit indicates the parser to proceed till it gets a correct independent unit. Therefore it is always easy and convenient to start with a dependent unit.

Procedures

As the finite State Automata (FSA) and the Augmented Transition Network (ATN) advocate, the input strings are taken for analysis, one by one, keeping the starting node as the first word of a sentence. The parser proceeds further, making use of the grammatical information labeled to the words and traverses to the next state. In our approach based on the word level grammatical tags attached with the input text, the parser select different modules. For example, if the word has <NV> as its tag, the parser moves to non-finite verb modules and identifies it as a clause ‘Se’. In each module, the parser follows two types of procedures.

i) First, it looks for the information within the word to decide a clause.

ii) If the information is not sufficient or if the information indicates to look at the following word, then the parser takes the information from the next word.

Identification of clauses and Sentences

All the four non-finite form of verbs viz., infinitive, conditional, verbal participle and relative participle constitute a sub-ordinate clause. In addition to these affirmative constructions, the negative forms (except infinitive, which does not have negative form) also construct the sub-ordinate clauses. Since the relative participle has the adjectival function, it is not segmented as a clause. As mentioned in the previous section of this paper, there are 5 types of sentences, including the subordinate clause.

Embedded Sentence

In the present study, the subordinate clause is marked as ‘Se’ that is, “sentence embedded”. The criterion for identifying the sentence embedded is given below:

i) If a verb phrase (VP) is marked as conditional (VP-con), or verbal participle (VP-vp) or infinitive (VP-inf), then the text from the beginning of the sentence or after a <NV> verb, till the VP is as a sub-ordinate clause (sentence embedded)
For example

1. **raaNi nanRaakap paTittu veelaikkuc cenRaal**

   \[\text{[raaNi } \text{< NN: (ppn)>]} \text{ NP}_{\text{nom}} \text{[nanRaaka < AV: (adv)> [paTittu < NV: (vpm, pst)>]} \text{ VP}_{\text{vp}} \text{[veelaikku< NN: (ian, dat)>]} \text{ NP}_{\text{dat}} \text{ [cenRaal <FV: vb, pst, 3sf>] } \text{ VP.}\]

   ‘Rani studied well and went to a job’

Since a non-finite form, verbal participle (VP_vp) has occurred with verb phrase

\[\text{[[raaNi } \text{< NN: (ppn)>]} \text{ NP}_{\text{nom}} \text{[nanRaaka<AV : (adv)> [paTittu<NV: (vpm, pst)> VP_vp]} \text{ ]} \text{ Se ‘Rani having studied well’}

is the sub-ordinate clause.

**Main Sentence (Verbal)**

The main clause of a sentence is called main /matrix sentence and is marked as ‘Sm’. The criterion for identifying the main clause is:

ii) If a verb phrase (VP) is not marked for any of the non-finite verb as _inf_, _con_, _vp_, then mark it as main clause.

For example

2. **mallikaa neeRRu viiTukku poonaaL**

   \[\text{[mallikaa <NN: (ppn)>]} \text{NP}_{\text{nom}} \text{[viiTukku<NN: (ian, dat)>]} \text{NP}_{\text{dat}} \text{[neeRRu <AV: (adv)> poonaaL<FV: Vb, pst, 3sf>] } \text{VP.}\]

   ‘Malliga went home yesterday’

Since, the verb phrase is not marked for _inf_, or _con_, or _vp_, [[mallikaa <NN: (ppn)>]NP_{nom} [viiTukku<NN: (ian, dat)>]NP_{dat} [neeRRu < AV: (adv)> poonaaL<Fv: Vb, pst, 3sf>]VP ]Sm, ‘Malliga went home yesterday’ is the main sentence.

**Interrogative Sentence**

To identify the interrogative sentence, the following mechanism is adopted.

(iii) If an interrogative word <ITW) or an interrogative marker occurs in a sentence, treat that sentence as an interrogative sentence.

For example

3. **raaman een azutaan**

   \[\text{[raaman } \text{< NN: (ppn)>]} \text{NP}_{\text{nom}} \text{[een<ITW:]} [\text{azutaan < FV: (vb, pst, 3sm)>]} \text{VP.}\]

   ‘Why did Rama cry?’

Since, an interrogative word<ITW> has occurs, [[raaman < NN: (ppn)>] NP_{nom} [een < ITW: (itw)>] [azutaan < FV: (vb, pst, 3sm)>]} VP] Si ‘Why did Rama cry? is an interrogative sentence and is marked as ‘Si’. 
4. *ciitaa koovilukku cenRaaLaa?*

   \[\text{[ciitaa <NN: (ppn) >]} \text{NP\_nom} \text{[koovilukku <NN: (ian, dat) >]} \text{NP\_dat} \]

   \[\text{[cenRaaLaa? <FV: (vb, pst, 3sf, int)>]} \text{VP}.\]

   ‘Did Sita go to temple?’

   Here, an interrogative marker –aa has occurred with the finite verb, hence the sentence is marked for ‘Si’ and is given below:

   \[\text{[ciitaa <NN: (ppn) >]} \text{NP\_nom} \text{[koovilukku <NN: (ian, dat) >]} \text{NP\_dat} \]

   \[\text{[cenRaaLaa? <FV: (vb, pst, 3sm, int)>]} \text{VP}\text{Si}.\]

**Equational Sentence**

   The criterion for identifying the equational sentence (EQS) is given below:

   \(\text{(vi) If a sentence has only two noun phrases (NP) and are marked for nominative case (NP\_nom), mark the sentence as an equational sentence.}\)

   For example

   5. *atu cari*

      \[\text{[atu <PNn: (pro, dem) >]} \text{NP\_nom} \text{[cari <NN: abn)>]} \text{NP\_nom}\]

      ‘That is right’

**Elliptical Sentence**

   The mechanism adopted to identify the elliptical sentence ‘Sel’ is

   \(\text{(v) If one or more than one noun phrase (NP) is marked for any case or for _emp_, mark it as an elliptical sentence.}\)

   For example

   6. *aamaam cengkatir*

      \[\text{[aamaam <NN: (ind)> cengkatir<NN: (ppn)] \text{NP\_nom} \text{Sel}}\]

      ‘Yes Senkathir’

   The above sentence has only one noun phrase marked for nominative case and hence it is an elliptical sentence and is marked as ‘Sel’.

**Conclusion**

As the parser has already identified the phrases and they are marked properly, it is easy to mark the clause and sentence boundary. With the above listed rules the system is capable of identifying 100% of clauses and 97% of the sentences in the given sample. The study was carried out on a smaller corpus. When a large corpus is taken for analysis, few more rules may be needed to be incorporated in the system.
Reference

The Technology Development for Indian Languages (TDIL) 13

Resource Centre for Indian Language Technology Solutions (RCILTS) 19

Consortium of Organizations Working on Malayalam Computing (COWMAC) 21

Morphological Analyzer (Morphological Analyzer), Machine Translation (Machine Translation), Text to Speech (TTS), Speech to Text (STT), Summarizer (Summarizer), Information Retrieval (Information Retrieval), OCR (OCR), Search Engine in Regional Languages.
Computer Aided Language Learning பதவியான மற்றும் சித்திரான வழிபாட்டை பயிற்று நிறைந்த நிர்வாக்கள் என்றும் புதுதியாக்கப் போக்கும்.

**Script Processors**

வழியில் அவர்களுடைய தமிழ்க் கோட்டையில் பயிற்றுத் தயாரிக்கப்பட்ட செயற்கை தமிழ் தன்மைகள் மற்றும் பாதுகாக்கச் செய்யப்பட்டுள்ளது. பிரிவுகூறுகள் நடைமுறை மற்றும் சேவைகள் இல்லாதிட்டு, வேறு பிரிவுகூறாண்கள் தமிழ் தன்மைகள் மற்றும் பாதுகாப்புகள் மற்றும் பல மற்றும் பராமரிப்பு செயல்களை தொடர்புபடுத்தும். அத்துடன், மற்ற பிரிவுகூறு மற்றும் சேவைகள் வழியில் அவர்களுடைய தமிழ்க் கோட்டையில் பயிற்றுத் தயாரிக்கப்பட்டுள்ளது. புதுதியாக்க உறவுகளை தமிழ் தன்மைகள் மற்றும் பாதுகாப்புகள் மற்றும் பல மற்றும் பராமரிப்பு செயல்களை தொடர்புபடுத்தும்.

**Word Processors**

இரு தமிழ்க் கோட்டைகள் என்றும் வருவியது பெரும்பாலும் கண்டுபிடிக்கப்பட்டுள்ளது பாதுகாப்பு வருவியது இல்லாமல் நிறைந்தது. இதனாலால் புதுதியாக்க உறவுகளை தமிழ் தன்மைகள் மற்றும் பாதுகாப்புகள் மற்றும் பல மற்றும் பராமரிப்பு செயல்களை தொடர்புபடுத்தும்.

இரு தமிழ்க் கோட்டைகளின் வழியில் பயிற்றுத் தயாரிக்கப்பட்டுள்ளது பாதுகாப்பு வருவியது இல்லாமல் நிறைந்தது. அரசியல் படைப்புப் பலகைகள் மற்றும் பல மற்றும் பராமரிப்பு செயல்களை தொடர்புபடுத்தும்.

197
Global Writer ஓரடையும் ஒரு பதிப்பு வடிவத்தில். இது 100 மைக்கேடுகள் பூர் அமைப்பங்கள் ஒன்றிற்கொன்றாக PDTEXT வடிவத்தில் கொண்டுள்ள பதிப்பு. இது பல்லு, நூற்றுக்கணக்கான இந்தியத் தமிழ் பதிப்புகளை மாற்றத்திற்கான நோக்கையிட்டு கையாய்வு செய்துள்ளது.

Script Converter
தமிழ் பம்பிச்சர் ரைட்ட் முறைத்தழைகள் மற்றும் றகாமல் பழக்கம் மற்றும் தமிழர் ரைட்ட் வழிகாட்டுதல் பட்டியல்கள் இயல்பாக்கமாகப் பேணி குணமுனைக்கப்பட்டுள்ளன. அவற்றில் தமிழ் பட்டியல் இணைப்பு பல்கலை இன்று ஆக்கம் உடையது. தமிழ் பட்டியல்கள் (ISM Import) வடிவத்தில் கொரிய -அமைப்பு மற்றும் இணைந்த பட்டியல் இன்றைய ஆக்கமாகவும் உள்ளது.

Electronic Online Dictionaries
தமிழ் மொழியையும் ஐக்கிய நாடுகளின் (Digital Dictionaries of South Asia) வடிவிலும் இணைந்த அனைத்து தமிழ்ச் சொற்கள் பூர் பட்டியல் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. ஆனால் அரைமுறை இயல்புகள் ஒன்றில் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. Charles Phillips Brown Dictionary of Telugu - English Dictionary இழைத்து காணப்படும் பட்டியல் மற்றும் அன்று அரைமுறை இயல்புகள் ஒன்றில் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. தமிழ் மொழி சொற்கள் பட்டியல் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. இந்த பட்டியல் தமிழ் மொழியையும் ஐக்கிய நாடுகளின் தமிழ் மொழி சொற்கள் பட்டியல் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. ஆனால் அரைமுறை இயல்புகள் ஒன்றில் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. தமிழ் மொழி சொற்கள் பட்டியல் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. இந்த பட்டியல் தமிழ் மொழி சொற்கள் பட்டியல் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. ஆனால் அரைமுறை இயல்புகள் ஒன்றில் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது. தமிழ் மொழி சொற்கள் பட்டியல் இணைந்த தமிழர் இயல்புகள் கொண்டுள்ளது.
மென்பொருள் என்பது விளக்கக் காட்சியில் பிரதானமானது. பரமாங்கம் மென்பொருளில் தம் ரீதியாக பதிப்புகளை வெளிந்துள்ளது. பயிற்சிக்குறிப்பு மென்பொருளில் 1997-ஆம் ஆண்டு நிறுவனம் வேலைகொடுத்து பார்வை முறை. 

நடவடிக்கை அல்லது அறிவியல் பயிற்சிக்குறிப்பு விளக்கக் காட்சியில் பிரதானமானது. மென்பொருளில் பயிற்சி பயிற்சிக்குறிப்பின் தொடர்புடையது புறங்கேற்கவும்.

முன்னேற்றத்திற்கு அகாதமி கொண்ட இணைப்பு இல் இலக்கணப்பத்தில் அடியானது. அங்கு அரசின் செயலாண் அளவை மென்பொருளில் படத்திற்கும் பயன்பாடு.

மூலம் பட்டியல்

மென்பொருளாளர் நாடகாரின் குறிப்பிட்டிய விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. மென்பொருளாளர் நாடகாரின் குறிப்பிட்டிய விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. அந்த விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. 

மூலம் பட்டியல்

மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. 

மூலம் பட்டியல்

மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. 

மூலம் பட்டியல்

மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. 

மூலம் பட்டியல்

மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை. 

மூலம் பட்டியல்

மென்பொருளாளர் மேற்கொண்ட விளக்கக் காட்சியில் பயிற்சிக்குறிப்பின் தொடர்புடையது பார்வை முறை.
A Knowledge Engineering Approach is being used for the development of Anveshak. Grammars are constructed by hand, domain patterns are discovered by introspection and inspection of a corpus. Much laborious tuning and "hill climbing" statistical methods are used. OCR (Optical Character Recognition) is used in the process.
Computer aided language learning software is seen in the website languageshome.com.

LI LA

Learn Indian Language through Artificial Intelligence

Tamil Smart system

Penn Language Centre

Computer aided language learning software is seen in the website languageshome.com.
மேற்புறம்

தமிழில் பொறிக்கப்பட்ட அற்புத்தப் பாடல்கள் பட்டியலிடுகிறது. அதிபர் கிளையில் விளக்கம் வழங்குவதற்காக எழுதிக்கொள்ளப்பட்டுள்ளது. வேளவாய்வு ஒன்றிய தொழில் மற்றும் செல்வை தமிழ் மொழியில் பொறிக்கப்பட்டது. இது ஒன்றிய தொழில்களுக்காக எழுதிக்கொள்ளப்பட்டுள்ளது. இந்த பாடல்கள் அருளும் கீழ் காலக்கான விளக்கப்பட்டுள்ளது. மேற்புறத்தில் பொறிக்கப்பட்டுள்ளது. பாடல் வழங்கும் காலக்கான விளக்கப்பட்டுள்ளது.
Conceptual Lexicon for Knowledge Representation

S. Rajendran
Tamil University
Thanjavur
raj_ushush@yahoo.com

Introduction

Conceptual graphs emphasize semantics. The earliest forms, called existential graphs, were invented by the philosopher Charles Sander Peirce (1897) as a graphical notation for symbolic logic. Lucien Teniere (1959) used similar graphs for his dependency grammar. The earliest form implemented on a computer was the correlational nets by Silvio Cecato (1961), who used them as intermediate language for machine translation. There are philosophical and psychological evidence that conceptual graphs are mental representation unbounded by knowledge of a particular language.

The proposed conceptual lexicon has concepts as its entries which are independent of a specific language and the meanings of concepts are given in terms of conceptual graphs from which the surface representation of lexical items belonging to a particular language can be derived. The proposed lexicon can be manipulated to generate a text in the form of a target language. The theory propounded by Sowa (1984) has been exploited to suit our purpose. At the same time the four levels of representations proposed for the a generative lexicon (Pustejovsky, 1995) and the semantic representation in WordNet (Pike Vassion, 2000) are also kept in mind while writing the meaning of a lexical item by means of conceptual graph.

Why Conceptual Lexicon

Dictionary definitions are mostly inadequate representations of words or concepts. A real definition will become encyclopedic. Let us take the concepts “horse” and “book”. “Horse” may require at least the following representation:

```
  hoof
  rider
  jockey
  stable

  mane
  HORSE

  stallion
  mare
  neigh
  foal
```
Similarly “book” requires the following representation:

![Diagram of a conceptual graph for “book”]

For a universal representation of concepts across languages conceptual graphs may be used.

**Conceptual Graphs**

Conceptual graphs from a knowledge representation language based on linguistics, psychology and philosophy (Sowa, 1984: 69). Concepts are language independent ones derived from percepts. A conceptual graph is a finite, connected, bipartite graph. The two kinds of nodes of the bipartite graph are concepts and conceptual relations. Every conceptual relation has one or more arcs, each of which must be linked to some concept. If a relation has n arcs, it is said to be n-adic, and its arcs are labeled 1, 2,…n. The term monadic is synonymous with 1-adic, dyadic with 2-adic and triadic with 3-adic. A single concept by itself may form a conceptual graph, but every arc of every conceptual relation must be linked to some concept. Concepts are discrete units. Combinations of concepts are not diffuse mixtures, but ordered structures. Only discrete relations are recorded in concepts. Continuous forms must be approximated by patterns of discrete units. For example, a space between a brick and a brick can be represented as follows (Sowa, 1984:72):

![Diagram of a conceptual graph for a space between a brick and a brick]

In the graphs, concept nodes represent entities, attributes, states, and events, and relation nodes show how the concepts are interconnected. Distinctions can be made between simple and complex concepts. Simple concepts are basic concepts from which complex concepts can be derived.

**Semantic Net Work**

Although the concept types CAT and TOMATO map directly to percepts, other types like PRICE, FUNCTION and JUSTICE have no sensory correlates. Abstract concepts acquire their meanings not
through direct associations with percepts, but through a vast networks of relationship that ultimately links them to concrete concepts. A conceptual graph has no meaning in isolation. For example, the description of the concept, MAN is represented as follows:

\[ \text{MAN} \rightarrow (ISA) \rightarrow [\text{HUMAN BEING}] \rightarrow (ISA) \rightarrow [\text{ANIMAL}] \]

Abstraction and Definition

Definition can specify a type in two different ways: by stating necessary and sufficient conditions for the type, or by giving a few examples and saying that everything similar to these belongs to the type. The first method derives from Aristotle’s method of definition by genus and differentiae. And the second method is closer to Wittgenstein (1953). AI systems have supported both methods. Conceptual graphs support type definitions by genus and differentiae as well as schemata and prototypes.

Type definition for KISS (Sowa, 1984:106)

\[
\begin{align*}
\text{type KISS}(x) & \, \text{is} \\
[\text{PERSON}] & \leftarrow (\text{AGNT}) \leftarrow [\text{TOUCH: } *x] \rightarrow (\text{MANNER}) \rightarrow [\text{TENDER}] \\
& \downarrow \\
& \text{(INST)} \\
& \downarrow \\
& \text{(PART)} \quad [\text{LIPS}]
\end{align*}
\]

Schemata

The basic structure for representing background knowledge for human-like inference is called the schemata. It is a pattern derived from past experience that is used for interpreting, planning, and imagining other experiences. Schemata incorporate domain-specific knowledge about the typical constellations of entities, attributes and events in the real world. Schemata are similar in structure to type definition. Yet concept type may have at most one definition, but arbitrarily many schemata. Type definitions present the narrow notion of a concept, and schemata present the broad notion. Type definitions are obligatory conditions that state only the essential properties, but schemata are optional defaults that state the commonly associated accidental properties.

Schema for BUS (x) is (Sowa, 1983: 129)

\[
\begin{align*}
[\text{BUS}] & \leftarrow (\text{INST}) \leftarrow [\text{TAVEL}] \rightarrow (\text{RATE}) \rightarrow [\text{SPEED:< 60kmsph}] \\
& \uparrow \\
& \text{(OBJ)} \quad \text{(CON)} \\
& \uparrow \\
[\text{DRIVE}] & \leftarrow (\text{AGNT}) \quad [\text{PASSENGER: } *] \\
& \downarrow \\
& \text{(QTY)} \\
& \downarrow \\
[\text{DRIVER}] & \quad [\text{NUMBER= 50}]
\end{align*}
\]
Schemata show the typical ways in which a concept may be used, but they do not describe a typical instance of a concept.

**PROTOTYPE**

A prototype is a typical instance. Instead of describing a specific individual, it describes a typical of “average individual”. A Schema for ELEPHANT might specify a range of characteristics for elephants or a range of behaviours and habitats for elephants. A prototype ELEPHANT would combine and restrict such schema to describe a typical elephant.

Prototype for ELEPHANT (x) is (Sowa 1984: 136):

\[ \text{ELEPHANT:*x} \rightarrow \begin{align*}
\text{CHAC} &\rightarrow \text{HEIGHT: @ 3.3 m} \\
\text{CHAC} &\rightarrow \text{WEIGHT: @ 5400 kg} \\
\text{COLR} &\rightarrow \text{DARK-GREY} \\
\text{PART} &\rightarrow \text{NOSE}\rightarrow \text{PREHENSILE} \\
\text{PART} &\rightarrow \text{EAR \{\} \rightarrow \text{NUMBER:2}} \\
\text{PART} &\rightarrow \text{TUSK \{\} \rightarrow \text{NUMBER:2}} \\
\text{PART} &\rightarrow \text{LEG \{\} \rightarrow \text{NUMBER:4}} \\
\text{STAT} &\rightarrow \text{LIVE} \\
\text{LOC} &\rightarrow \text{CONTINENT: \{Africa | Asia\}} \\
\text{DUR} &\rightarrow \text{TIME: @ 50 YEARS} 
\end{align*} \]

**Conceptual Representation**

Some of the conceptual relations listed in Sowa (1984) are adopted to suit our purpose.

- **accompaniment.** (ACCM) links [ENTITY:*x] to [ENTITY:*y] where *y is accompanying *x.
- **agent.** [AGNT] links [ACT] to [ANIMATE], where ANIMATE concept represents the actor of the action.
- **attribute.** (ATTR) links [ENTITY:*x] to [ENTITY:*y] where *x has an attribute *y.
- **cause.** (CASE) links [STATE:*x] to [STATE:*y] where *x has a cause *y.
- **characteristic.** (CHRC) links [ENTITY:*x] to [ENTITY]
- **destination.** (DEST) links [ACT] to [ENTITY] towards which the action is directed.
- **experience.** (EXPR) links [STATE] to [ANIMATE], who is experiencing that state.
- **instrument.** (INST) links [ENTITY] to [ACT] in which the entity is causally involved.
The following relations are also taken into account to define a concept using some other concept.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernymy-Hyponym</td>
<td>ANIMAL → MAMMAL</td>
</tr>
<tr>
<td>Hyponymy-Hypernym</td>
<td>COW → MAMMAL</td>
</tr>
<tr>
<td>Holonymy-Meronymy</td>
<td>TABLE → LEG</td>
</tr>
<tr>
<td>&quot;</td>
<td>DEPARTMENT → PROFESSOR</td>
</tr>
<tr>
<td>Meronymy-Holonymy</td>
<td>WHEEL → CART</td>
</tr>
<tr>
<td>Troponymy</td>
<td>WALK → LIMP</td>
</tr>
<tr>
<td>Entailment</td>
<td>SNORE → SLEEP</td>
</tr>
</tbody>
</table>

Pustejovsky (2001:56) characterize a generative lexicon as a computational system involving at least the following levels of representation:

1. ARGUMENT STRUCTURE: Specification of number and type of logical arguments
2. EVENT STRUCTURE: Definition of the event type of an expression and its subeventual structure
3. QUALIA STRUCTURE: A structural differentiation of the predicative force for a lexical item
4. LEXICAL INHERITANCE STRUCTURE: Identification of how a lexical structure is related to other structures in this type of lattice

Pustejovsky (2001:56) assumes that word meaning is structured on the basis of four generative factors, or qualia roles, that capture how humans understand objects and relations in the world and provide the minimal explanation for the linguistic behaviour of lexical items.

CONSTITUTIVE: the relation between an object and its constituent parts

FORMAL: the basic category that distinguishes the object within a larger domain

TELIC: the object’s purpose and function

AGENTIVE: factors involved in the object’s origin or “coming into being.”

The qualia structure is the core of the generative properties of the lexicon, because it provides a general strategy for creating increasingly specific concepts with conjunctive properties. A simple schematic description of a lexical item, $\alpha$, using this representation is shown below:

$$
\begin{align*}
\alpha & \\
\text{ARGSTR} & = \text{ARG}1 = x \\
\text{...} & \\
\text{CONST} & = \text{what } x \text{ is made of} \\
\text{QUALIA} & = \text{FORMAL} = \text{what } x \text{ is} \\
\text{TELIC} & = \text{function of } x \\
\text{AGENTIVE} & = \text{how } x \text{ come into being}
\end{align*}
$$

207
The lexical structure for *book* as an object can then be represented as follows:

\[
\text{book}
\]

\[
\begin{align*}
\text{ARG1} &= y: \text{information} \\
\text{ARGSTR} &= \text{ARG2} = x: \text{phys_obj} \\
\text{FORM} &= \text{hends} (x,y) \\
\text{QUALIA} &= \text{TELIC} = \text{read} (e,w,x,y) \\
\text{AGENT} &= \text{write} (e', v, x, y)
\end{align*}
\]

The ideas propounded by Pustejovsky will also be taken into consideration while defining a concept. The following is the sample of the conceptual lexicon. Each item is a concept and the concepts will be mapped against the lexical items of a language.

\[
\begin{align*}
[\text{CAT}] &\rightarrow (\text{ISA}) \rightarrow [\text{ANIMAL}] \\
[\text{PENCIL}] &\rightarrow (\text{ISA}) \rightarrow [\text{INSTRUMENT}] \\
&\downarrow \\
&(\text{FUNCT}) \\
&\downarrow \\
&[\text{WRITING}]
\end{align*}
\]

The verbs are provided with argument structures. A frame of arguments will be given with their necessary relations. The verbal concept ACT represented in the following fashion.

\[
\begin{align*}
[\text{ACT}] &\rightarrow (\text{ISA}) \rightarrow [\text{EVENT}] - \\
&(\text{AGENT}) \rightarrow [\text{ANIMATE ENTITY}] \\
[\text{ARRIVE}] &\rightarrow (\text{ISA}) \rightarrow [\text{EVENT}] - \\
&(\text{AGENT}) \rightarrow [\text{MOBILE-ENTITY}] \\
&(\text{GOAL}) \rightarrow [\text{LOCATION}]
\end{align*}
\]

**Lexical and Conceptual Structures**

Each natural language has a well-organized lexical and syntactic system. Each domain of knowledge has a well-organized conceptual system. Complexities arise because each language tends to use and reuse the same words and lexical patterns in many different conceptual domains. The lexical structures are

- Relatively domain independent,
- Dependent on syntax and word forms,
- Highly language dependent.

And the conceptual structures are

- Highly domain dependent,
- Independent of syntax and word forms,
- Language independent, but possibly culture dependent.

When there are cross-linguistic similarities in lexical patterns, they usually result from underlying conceptual similarities. English verb *give*, for example, takes a subject, object, and indirect object. Other languages may have different cases marked by different prepositions, postpositions, inflections, and word order; but the verb that mean roughly the same as *give* also have three participants - a giver, a thing given, and a recipient. In all languages, the three participants in the conceptual pattern lead to three arguments in the lexical patterns.
The distinction between lexical structures and conceptual structures addresses the following things:

- **Lexical structures** are oriented towards language. The representation developed here is strongly influenced by linguistic theories of syntax and thematic roles.

- **Conceptual structures** are designed for representing knowledge about the world. They may grow too large to be expressed in a single sentence, and they may contain concepts types that cannot be expressed by a single word.

- Since they can be represented by similar structures, the same operations can be used on them. Furthermore, lexical structures can be converted to deeper conceptual structures by a step-by-step process, not by a translation between radically different forms.

- Finally, common structures facilitate language learning and conceptual creativity. In learning, a child generalizes conceptual structures learned from experience to form the initial lexical structures needed for language. Metaphor and conceptual refinement create new conceptual structures by adapting old lexical structures to novel situations.

**Conclusion**

The distinction between lexical structures and conceptual structures provides a principled basis for partitioning knowledge into the lexicon and the more detailed knowledge about the world. Conceptual graphs provide a formalism for representing both kinds of structures with a level of precision that allows deeper and more systematic analysis of the relationship between them. As a result, they can help to replace vague discussion with a precise methodology that has a greater chance of being computerized. Finally, the direct mapping between conceptual graph and natural language can simplify the task of knowledge acquisition: a knowledge base of conceptual graphs could be generated directly from natural language inputs. After being primed with a dictionary of lexical knowledge, the system could build up its own encyclopaedia of world with the aid of a tutor communicating in English, not a knowledge engineer coding in a specialized notation.

**Bibliography**

கல்விக்குறிப்பிட்டு நார்த்திக்காரன் அறிவாளர்
அற்றக்கற் பகிர்கள்

ம. நூற்றாண்டு கற்பனே

முருக்குகள் & கல்வீனங்கள்

எட்டுசர்வேஸ் பாலாஜாவுத்தான்

ndsundaram@hotmail.com

அற்றக்கற் பகிர்கள்

(Communication and Natural Languages)

மக்கள் ஆய்வுகளில் பாரத்து மத்தியா பெரும் புறம் மாவுதானையிய அற்றக்கற் பகிர்கள் காரணமாக மண்டு நடக்கும்போது முன்னேறி நடைபெறும் அற்றக்கற் பகிர்கள். அற்றக்கற் பகிர்கள் முறையே நிலையியல் அடையாளங்களால் நிர்வகிக்கப்பட்டுள்ளன. பாரத்து மத்தியா பெரும் புறம் மாவுதானையிய, பாரத்து மத்தியா பெரும் புறம் மாவுதானையிய அடையாளங்கள், உயிரியத்து அளவு செறியுள்ளது, சமயத்து முன்னேறிக்கும் அடையாளங்களில் நெற்பாட்டு நுட்பங்கள் எந்த வகையான ஆய்வுகளும் ஆவாதியலை பொறுபைப்படுவது (Linguistics). கண்டு பிடித்து அந்தானாக அந்தானாக பாரத்து மத்தியா பெரும் புறம் மாவுதானையிய ஆய்வுகள் (Generative Grammar) கண்டுபிடித்து வந்துள்ளது. பாரத்து மத்தியா பெரும் புறம் மாவுதானையிய நுட்பங்கள் செறியுள்ளது (Model of Electronic Representation) கண்டுபிடித்து வந்துள்ளது.

(Pragmatic and

(Science of Human Cognition) என் பாய்ந்து ஆமையாகும்.

(Principles and Models)

(Science of Human Cognition) என் பாய்ந்து ஆமையாகும்.
other World knowledge) in the context of discourse, computational linguistics, and artificial intelligence. In the context of discourse, computational linguistics, and artificial intelligence (discourse - computational linguistics - artificial intelligence), understanding and processing of language is crucial. This involves representation and processing of knowledge in computational linguistics. From a discourse perspective, understanding and processing of language involves logical knowledge and embedded knowledge. This involves understanding and processing of knowledge in the context of artificial intelligence (science of artificial intelligence), which is a fundamental aspect of natural language processing.

Computational Linguistics

Computational Linguistics (Natural Language Knowledge Representation) involves the study of natural language at various levels, including sentence and discourse level, and at different stages, such as analysis and synthesis. This involves understanding and processing of knowledge in computability theory. Computational Linguistics also involves understanding and processing of knowledge in the context of computer science, which is a fundamental aspect of natural language processing.

Language Technology

Language Technology involves the study of natural language at various levels, including sentence and discourse level, and at different stages, such as analysis and synthesis. This involves understanding and processing of knowledge in computability theory. Language Technology also involves understanding and processing of knowledge in the context of computer science, which is a fundamental aspect of natural language processing.
மொழிவளர்ப்பு (E-language Planning - ELP)

மொழிவளர்ப்பு, நோய்பார்ப்பு, போட்டிக்கில்குறிப்பிட்டு மொழிவளர்ப்பானது என்னும் ஒரு குறிப்பிட்டு மொழியியல் மற்றும் நுட்பத்திற்கு அதிகம் அளிக்கின்றது (Globalization). மொழி வளர்ப்பு மொழி வளர்ப்பு போட்டிக்கில்குறிப்பிட்டு மொழியியல் (Applied Linguistics) என்பதாக மொழிவளர்ப்பு என்று குறிப்பிட்டு மொழியியல் (Language Planning) என்று குறிப்பிட்டு மொழி வளர்ப்பு (Graphization), பண்பாடு மொழி வளர்ப்பு (Modernization), என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று (Electronic Language) ஆகும். தன்மை மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வள ர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு (Digital Divide) என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு என்று குறிப்பிட்டு மொழி வளர்ப்பு (Encoding, Font, Keyboard)

மொழியியல் மொழி வளர்ப்பு என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்குறிப்பிட்டு மொழியியல் என்றும் போட்டிக்கில்
(Diglossic Tamil and Computational Linguistics)

Sense Disambiguation

Syntactic Structural Ambiguity
The document contains multiple sections discussing various linguistic topics. Here is a translation of the content:

**Syntax**

- Syntactic Knowledge
- Pragmatic Knowledge

**Phonology**

- Phonological Ambiguity

**Morphology**

- Morphological Ambiguity

**Semantics**

- Sense Ambiguity

**Discourse**

- Disambiguation

**Grammar Engineering and Computational Grammar**

The document further details each of these topics in depth, providing insights and explanations.
Generative Lexicon, Grammar, Structure Grammar - GPSG), (Mathematical Grammar), Tree -Adjoining Grammar - TAG), (Hidden Markov Model), (N-gram Method), (Two-Level Morphology), ( Generative Grammar), (Lexical - Functional Grammar - LFG), (Formal Grammar) (Research in Tamil Computational Linguistics), (Mathematical Linguistics), (Semantic Parsing), (Syntactic Parsing), (Parsers), (Morphological Parser), (Word Class Tagger), (Semantic Analyzer), (Word Sense Disambiguator), (Language Resources), (Computational Lexicon), (Generative Lexicon) (Tree-Adjoining Grammar - TAG), (Generalised Phrase Structure Grammar - GPSG), (Lexical - Functional Grammar - LFG), (Formal Grammar) (Research in Tamil Computational Linguistics), (Mathematical Linguistics), (Semantic Parsing), (Syntactic Parsing), (Parsers), (Morphological Parser), (Word Class Tagger), (Semantic Analyzer), (Word Sense Disambiguator), (Language Resources), (Computational Lexicon), (Generative Lexicon) (Tree-Adjoining Grammar - TAG), (Generalised Phrase Structure Grammar - GPSG), (Lexical - Functional Grammar - LFG), (Formal Grammar)


Noun Phrase Chunker using Finite State Automata for an Agglutinative Language

Vijay Sundar Ram R and Sobha Lalitha Devi
AU-KBC Research Centre,
MIT Campus of Anna University,
Chromepet, Chennai –44
{sobha,sundar}@au-kbc.org

Abstract
This paper presents a system for noun phrase chunking for Tamil, an agglutinative language. The partial chunking of the text is done by a rule-based approach where the rules are embedded in a finite state automaton (FSA), which recognizes the chunks at a high accuracy rate and speed. The chunking of text being the pre-processing task needs to be of good performance. The evaluation of the system shows a recall of 93.7% and precision of 94.9%.

Introduction
We present a system for noun phrase chunking for an agglutinative language, Tamil. The noun phrase considered here has a head noun preceded by determiner, quantifier, classifier and adjective in sequence. Here the determiner, quantifier, classifier, and adjective are optional. The recursive noun phrase occurs when possessive case marker occurs. Often the noun phrase has the information about the happening. The event executor and event receiver of the event are usually noun phrases around the verb. The proper chunking of noun phrase helps in improving the efficiency of the information extraction, machine translation and in information retrieval system by improving the terms in the term vs document matrix. The system is built using rule based approach, where the rules are embedded in a finite state automaton (FSA), as the structures are recognized with high degree of accuracy.

Several methods have come up for this task. Church’s stochastic noun phrase tagging was one of the early attempts, where the corpus frequencies were used to determine the noun phrase boundaries (Church, 1988). Abney did partial parsing using finite state cascades, where the finite state cascade has sequence of levels. Phrase at one level is built on the phrase at the previous level without any recursion (Abney, 1996). Ramshaw and Marcus used Eric Brill’s transformation based learning for recognizing the noun chunks and other text chunking (Ramshaw, 1995). Dimitrios Kokkinakis and Sofie Johansson Kokkinakis did a cascaded Finite-State Parser for Syntactic Analysis of Swedish (Dimitrios, 1999). Chunk tagger using markov model technique for recognizing the internal structures and syntactic category of simple as well as complex structures was done by Wojciech (Wojciech 1998). Noun phrase chunking for German using probabilistic context-free parser for learning and tagging the most probable chunk sequence is done by Helmut (Helmut, 2000).
The rest of the paper is organised as follows. The section 2 starts with briefing about the agglutinative language with samples of noun phases. In section 3 techniques of finite state automata (FSA), different preprocessing works and the implementation of the Tamil NP chunker are explained. The evaluation and the discussions on result in section 4 and finally the paper ends with conclusion.

Description about the Language

The system is developed for an agglutinative language such as Tamil, Telugu, Malayalam, etc. Here the language under consideration is Tamil. Tamil belongs to the South Dravidian family of languages. It is a verb final language and allows scrambling. It has post-positions, the genitive precedes the head noun in the genitive phrase and the complementizer follows the embedded clause. Adjective, participial adjectives and free relatives precede the head noun. It is a nominative-accusative language like the other Dravidian languages. The subject of a Tamil sentence is mostly nominative, although there are constructions with certain verbs that require dative subjects. Tamil has png agreement.

Noun Phrases in Tamil

Tamil is a relatively free word order language, but when it comes to noun phrases and clausal constructions it behave as a fixed word order language. As in other languages, Tamil also has optional and obligatory parts in the noun phrase. Head noun is obligatory and all other constituents that precede the head noun are optional. In this section we discuss in detail about different noun phrases in Tamil.

Consider the following

1. periya viitu  
   big+ADJ house+N+NOM  
   (Big House)
2. oru azhakiya viitukku  
   one +Q beautiful+ADJ house+N+DAT  
   (For a beautiful house)
3. inta koyil  
   this+DET temple+N+NOM  
   (This temple)

In examples 1, 2 and 3 the structure of the noun phrase has a head noun, which may or may not be preceded by optional categories like determiner, quantifier, classifier, and adjective. This is the general structure of noun phrase. This structure is represented by the following Rule.

Rule a:

\[
\text{NP} = [\text{Determiner}] [\text{Quantifier}] [\text{classifier}] [\text{Adjective}] [\text{N}]
\]

4. avanutaiya braamaandamaana araiyin vizakukal  
   His+PN+POS grand+ADJ room+N+POS lamp+N+PL  
   (His grand room’s lamp)

Example 4 has a possessive noun. If the head noun has a possessive case marker and followed by a noun phrase, both the noun phrases are chunked into one noun phrase. This may go recursively. This is shown in Rule b.
Rule b:
possessive NP => [Determiner][Quantifier][classifier] [Adjective] {N+possessive case}
NP=> (possessiveNP)* Rule a
The rule a and b are combined into a single rule
Rule 1:
NP=> [possessive NP] +Rule a

5. ivviitu
   this+DET house+N+NOM
   (This house)
Here compounding of the determiner and the noun forms single word. The tags of this word also obey
the previous rule but the tags are in the same word. This type is handled separately to accept the tags
within the word.

6. powuc ceyalaalar
   (Secretary)
In example 6, here the two nouns, which are suppose to be a single noun is written separately and
connected by a sandhi. This is handled by the rule given below:
Rule 2:
NP=> {N+sandhi} {head N}

7. therku      maaligai
   south+N+NOM palace+N+NOM
   (Southern palace)
Example 7 has two nominative nouns, where the first denotes the direction. These two nouns go together
to form a noun phrase. The eight directions north, south, east, west, northeastern, northwestern,
southeastern and southwestern are considered. This is handled by the following rule
Rule 3:
NP => {direction denoting NP} {NP}

8. kadal       niir
   sea+N+NOM water+N+NOM
   (Seawater)
When two nominative nouns are adjacent to each other the two nominative nouns can be chunked into
noun phrase.
   NP => {N+NOM} {N+NOM}
Tamil NP Chunker

The noun phrase chunking is done by building finite state automata using the linguistic rules. By traversing the text through the FSA, the chunking of the text is done.

Finite State Automata

A finite state automaton is an abstract device used for recognizing simple syntactic structures or patterns in text strings. An automaton is normally depicted by a directed graph in a so-called state diagram. An FSA as a string-processing device accepts text strings as input and decides if the structure is correct, that is, it either accepts or rejects the string. From a mathematical point of view, it may therefore be regarded as a function, mapping a set of strings to the set {ACCEPT, REJECT}. Based on the transitions given by the FSA, they are classified as Non-Deterministic Finite State Automata (NFA) and Deterministic Finite State Automata (DFA).

The deterministic FSA is found to be unsuitable for parsing of the sentence. Since the grammar of the languages are context free and non-deterministic in nature. But for partial chunking such as noun phrase chunking the structure of the phrase are highly fixed. So usage of deterministic FSA helps in high degree of accurate chunking.

The Deterministic Finite-State Automaton (DFA) which is a special case of NFA has the following requirements:

1. There are no transitions involving $\epsilon$.
2. No state has two outgoing transitions based on the same symbol

The Deterministic FSA used in this task is shown in Figure 1.

![Deterministic FSA](Figure 1 Deterministic FSA)

In the present work the FSA is used in the form of a state table. The morphological tags of the words are the transition symbols.
Architecture of the System

The architecture of the system is shown in figure 3.

Figure 3 Architecture

![Architecture Diagram]

The Preprocessing Works

The input text to be chunked is initially fed to the morphological analyser (Viswanathan 2003), which gives multiple outputs for each word. The morphologically tagged output is sent to a hybrid approach based Part of Speech (POS) tagger (Arulmozhi 2006), where the ambiguities in the output of the morphological analyser are corrected using the POS tags.

On giving the preprocessed text, which is now a tagged text, state transition takes place in FSA based on the tags as the symbol to trigger the transition. The chunking of the phrase completely depends on the morphological tags.

The process in Finite State Automata

The FSA is built using the manually crafted linguistic rules.

When a tagged sentence is given, the current word’s tag is considered as the transition symbol to have transition to the next state, in the next state the next word’s tags is considered. Similarly the traversing in FSA happens till it reaches the end state. If it successfully reaches the end state this part of the text is chunked as noun phrase.

The nouns with all the case markers have the same noun phrase structure except the nouns with possessive case marker. So the nouns with all case markers except the possessive case marker are treated similarly. So the case marking tags in the nouns except the possessive case are not considered. This suppression of the sub-tags reduces the number of transition from 400 to 40 in the determinized FSA. This optimizes the FSA and improves the speed and the efficiency of the chunking process.

The noun phrase chunked text obtained after traversing through the FSA, is again processed using the heuristic rules.

Results and Discussion

The system is evaluated with the data taken from the CIIL corpus (Central Institute of Indian languages corpus). The numbers of sentences under consideration are 500, which contains 2180 noun phrases. The
The number of noun phrases recognized correctly by the system is 2043, which is at a precision of 94.9% and the recall of 93.7%. The results are tabulated as follows.

**Table 1 Performance of the Chunker**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Number of Sentences</th>
<th>Number of NPs present</th>
<th>Number of NPs chunked</th>
<th>Number of NPs correctly chunked</th>
<th>Recall %</th>
<th>Precision %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500</td>
<td>2180</td>
<td>2153</td>
<td>2043</td>
<td>93.7</td>
<td>94.9</td>
</tr>
</tbody>
</table>

The capitalization of the first letter gives an easy identification of proper nouns in English language. This is not there in Indian languages, which makes the task of recognizing the noun phrase complex. The challenging task is chunking the noun phrase formed by two nouns such as “kadal niir” (sea water), which should be recognized as one chunk. We cannot go blindly by chunking two nominative nouns into a chunk because doing that will make error in the following case.

```
raman        palam      saappittaan
raman+N+NOM  fruit+N+NOM  eat+V+PST
(raman ate fruit)
```

Here if we apply the rule that two adjacent nominative nouns will form a chunk, then the two nouns ‘raman’ and ‘palam’ will form as one chunk, “raman palam”. But this is an incorrect chunk since they are two different noun phrases where ‘raman’ is the subject and ‘palam’ is the object of the sentence. The accusative marker, the marker of direct object, in ‘palam’ is dropped which is a common phenomenon in Tamil. We are trying to overcome this by subcategorizing the verb to identify the subject and object of the verb. These types of correction have to be done after the chunking is over.

**Conclusion**

The paper presents a noun phrase chunking for an agglutinative language, Tamil. In developing the system we are using finite state automata technique, where the rules are embedded in it. The system performs with 94.9% precision and 93.7% recall. The system takes 2.5 sec to chunk 1000 sentences. In future, we will be handling the noun phrases with multiple nouns.

**Reference**


8. Viswanathan, S Ramesh Kumar, B Kumara Shanmugam, S Arulmozi 2003, A Tamil Morphological Analyser , International Conference On Natural Languageprocessing, Central Institute of Indian Languages, Mysore

Animated Sangathamizh Poems - E Learning

Arul Natarajan
dr.n.arul@gmail.com

Introduction

This paper on Animated Sangathamizh poems is an attempt to

• Understand the importance of Sangathamizh poems
• Revive the vibrant and dynamic values of Tamil culture and heritage
• Introduce it to the youth of modern world
• With the latest technology available at our hands.

Tamil

Tamil is a language with a literature that is classical — i.e., it is ancient, it has an independent tradition that arose mostly on its own, not as an offshoot of another tradition, and has a large and extremely rich body of ancient literature.

The quality of classical Tamil literature is such that it is fit to stand ahead of the great literatures of Sanskrit, Greek, Latin, Chinese, Persian and Arabic. The subtlety and profundity of its works, their varied scope, and their universality qualify Tamil to stand as one of the great classical traditions and literatures of the world. Tamil is one of the primary independent sources of modern Indian culture and tradition.

Sangathamizh Literature

Sangam literature refers to classical Tamil literature created between the years c. 600 BCE to 300 CE. This collection contains 2381 poems composed by 473 poets, some 102 of whom remain anonymous. Sangam literature is primarily secular dealing with everyday themes in a Tamil context. Much of the Tamil literature believed to have been composed in the Sangam period is lost to us, though detailed lists of works known to the 10th century compilers have survived.

The poems belonging to the Sangam literature were composed by Tamil poets, both men and women, from various professions and classes of society. These poems were later collected into various anthologies, edited, and with colophons added by anthologists and annotators around 1000 CE.

Sangam Poems falls into two categories: the 'inner field' (Agam), and the 'outer field'(Puram) as described even in the first available Tamil grammar, the Tolkappiyam.

The 'inner field' topics refer to personal or human aspects, such as love and sexual relationships, and are dealt with in a metaphorical and abstract manner.

The 'outer field' topics discuss all other aspects of human experience such as heroism, valour, ethics, benevolence, philanthropy, social life, and customs.

Sangathamizh illustrates the thematic classification scheme first described in the Tolkappiyam
The classification ties the emotions involved in agam poetry to a specific landscape. These landscapes are called *thinai*. These are: *kurinji* mountainous regions; *mullai* forests; *marutham* agricultural land; *neithal* coastal regions; *paalai* deserts. In addition to the landscape based *thinais*, *kaikkiLai* and *perunthinai* are used for unsolicited love and unsuited love respectively.

Similar *thinais* pertain to *puram* poems as well, though these categories are based on activity rather than landscape: *vetchi*, *karanthai*, *vanchi*, *kanchi*, *umignai*, *nochchi*, *thumbai*, *vaagai*, *paataan*, and *pothuviyal*.

**Revival**

Tamil, being one of the most ancient languages, yet vibrant and dynamic with its inexhaustible thoughts in every realm influencing humanity through its arts and literature has provided guidelines for happy living, just governance, noble behavior etc. It’s our duty to take this torch further in our time for the future by all the modern means.

The works of Sangam literature were lost and forgotten for several centuries. A revival took place from the late nineteenth century when works of religious and philosophical nature were written in a style that made it easier for the common people to enjoy. Several Tamil scholars such as S. V. Damodaram Pillai and U. V. Swaminatha Iyer. They painstakingly collected and catalogued numerous manuscripts in various stages of deterioration. They printed and published *Tholkappiyam*, *Nachinarkiniyar urai* (1895), *Tholkappiyam Senavariyar urai*, (1868), *Manimekalai* (1898), *Cilappatikaram* (1889), *Pattupattu* (1889), and *Purananuru* (1894), all with scholarly commentaries. Damodaram Pillai and Swaminatha Iyer published more 100 works in all, including minor poems. Nationalist poets like Bharathiyar and Bharathidasan began to utilize the power of poetry in influencing the masses.

With growth of literacy, Tamil prose began to blossom and mature. Short stories and novels began to appear. Many scholars including our Hon’ble CM have written books on sangathamizh poems. The popularity of Tamil Cinema has also provided opportunities for modern Tamil creative heads to take up sangathamizh in. Movies like *Thiruvilayadal*, *Mannadhimannan*, *Iruvar* and so on, has references to sangam poems.

**Why Sangathamizh Poems?**

Sangam Poems are the authority for moral Life. They fall into two categories; Agam and Puram, referring to personal or human aspects, such as love and sexual relationships in the former and all other aspects of human experience such as heroism, valor, ethics, benevolence, philanthropy, social life, and customs in the latter. This cultural treasure is known only to Tamil scholars and is still out of reach of common mass. This is because of lack of awareness of our heritage and culture and above all poor education. Though some of these poems are taught in schools and colleges, there is still no proper way to inspire the learner to have a further reading of the literatures.

**Animated Sangathamizh Poems**

Animated Sangathamizh Poems would be the adaptation of classical Tamil poems to modern music and animations without losing the values and traditional tastes. Poems from *Ettuthogai*, *Pathu Paattu*, *Padhinenkezhkanakku* etc., are full of interesting scenes paving way for good visualisation of the glorious living of our Tamils. These poems can be visualised in animated format which will attract not only every Tamil, but also every human being in the globe who try to attain the ethical values.
How to implement

Since the need of understanding cultural values in the society are increasing day by day, Animated Sangathamizh poems with musical rendering should reach the public through all the technological means available update. The challenge is quality, process and technology we are going to use so that it reaches the mass with a greater impact. The media in hand are Satellite channels, CDs, iphones and mobile phones. Since technology is improving day by day, High quality music, Stunning graphics and animations are possible to achieve the objective.

Music: Differing from the traditional way of composing music, the Animated Sangathamizh poems will have tunes to the taste of youth to attract them, at the same time without compromise in the relevance and values. Various music composers have proved that this is possible.

Animation: Stunning animations can be made for sangathamizh poems as they naturally have wonderful visuals and descriptive storyboards of past glory. Care will be taken to make the animations and graphic interesting, relevant to present day situations.

Meaning: The ultimate objective of the attempt is to reach this to the common mass. This can be achieved by using simple language for describing the meaning of the poems. People need not ne literate to understand what is being said.

Advantages

• **A great revolution:** This will be indeed a great revolution in revival of Sangam Literatures because through this attempt we can uplift the values of life which is found losing importance in the modern, fast world.

• **Multimedia:** Throughout history, man has shown his fascination towards visualizing his creative thoughts in various available resources. The technologies available with us can attract audience of all stages of the social pyramid.

• **Teaching/Learning aid:** This can further be an excellent resource for researchers, teachers and learners where hours of time spent on by hearting text form can be reduced to a minimal time with permanent registry in mind.

• **Future Generations:** This will certainly be a treasure for the generations to come as the world is becoming more and more modern and there is no time to turn back to even glance at the values which are missing.

Conclusion

In spite of being spoken by a whole race of people including the illiterate among us, Tamil has maintained such a continuity and uniformity that literature written 2000 years ago can be understood by the educated readers of the present. But we should not stop with this. Every common man, literate or illiterate should enjoy the essence of Tamil and we believe that Sangathamizh poems - being the most ancient literature, yet with inexhaustible thoughts in every realm influencing humanity through its words has provided guidelines for happy living, just governance, noble behaviour etc. It’s our duty to take this further in our time for the future and one such way is Animated Sangathamizh Poems.
Representation of Kinship in WordNet

S. Arulmozi
Dravidian University

Abstract

WordNet (Fellbaum, 1998) is one of the most resourceful semantic lexicons in English. Its main advantage is that it is hand-crafted, so data stored within its semantic network are of high quality. It is used in most of the NLP applications, particularly in sense disambiguation tasks. WordNets are already available in most of the languages of the world including Hindi. Efforts are underway in Indian languages using the expansion approach with Hindi WordNet as the base. Kinship presents a tough problem in the construction of WordNet, especially in Tamil and other major Dravidian languages. This paper presents the lexicographical issues involved in the construction of synsets (synonym sets) in general and kinship hierarchy in particular. A brief account on the representation of kinship hierarchy in WordNet will also be provided.

Introduction

The Princeton English WordNet (Fellbaum, 1998) is one of the most resourceful semantic lexical database in English. Its main advantage is that it is hand-crafted, so data stored within its semantic network are of high quality. It is widely used as a resource in many NLP applications such as Information Retrieval, Word Sense Disambiguation, etc.

The continuous expansion of the multilingual information society with a growing number of new languages present on the Web has led in recent years to a pressing demand for multilingual applications. To support such applications, multilingual language resources are needed, which however require a lot of human effort to be built. For this reason, the development of language independent resources which factorize what is common to many languages, and are possibly linked to the language-specific resources, could bring great advantages to the development of the multilingual resources in Indian languages.

Princeton’s English WordNet inspired extensive development of WordNets in European languages, EuroWordNet (Vossen, ) and also in other languages across the globe including WordNets in Indian languages, IndoWordNet (Pushpak, ).

In this paper, a brief account on WordNet and construction of synsets is given. The paper is organized as follows: Section 2 details about WordNet and activities in Indian languages. Section 3 deals with the construction of synsets in general and pinpoints few problems faced during the construction of synsets. Section 4 briefly lists the ontology of kinship in English WordNet followed by the problems faced in creating synsets for kinship concepts in Tamil and Telugu. The last section summarizes the work.
WordNet

WordNet was originally conceived and developed as a lexical database for English on the basis of psycholinguistic properties. The major lexical categories such as nouns, verbs, adjectives and adverbs are organized in terms of sets of synonyms (synsets) each representing a lexical concept.

A synset is a set of synonyms (word forms that have the same or similar meaning) and two words are said to be synonymous if their mutual substitution does not alter the truth-value of a given sentence in which they occur, in a given context. For example, {computer, computing machine, computing device, data processor, electronic computer, information processing system} form a synset because they can be used to refer to the same concept. These synsets are interconnected by certain relations, lexical relations such as synonymy, antonymy and semantic relations such as hyponymy (between specific and more general concepts) and meronymy (between parts and wholes).

An example of a synset is reproduced (from WordNet 2.1) here for clarity:

The synset for {computer, computing machine, computing device, data processor, electronic computer, information processing system} is related to:
- more general concept or the hypernym synset {machine}
- more specific concepts or hyponym synsets {analog computer}; {digital computer}; {node, client, guest}; {number cruncher}; {pari-mutuel machine, totalizer, totaliser, totalizator, totalisator} and {server, host}
- parts it is composed of {busbar, bus}; {cathode ray tube, CRT}; {central processing unit, CPU, processor, mainframe};

Each of these synsets is again related to other synsets as is illustrated for {machine} that is related to {device}, and {CPU} that is related to other parts {mother board, CPU board}, {circuit, electrical circuit}.

WordNet (2.1 version) has approximately 120,000 lexical items (word forms) organized into 100,000 meanings (word meanings). For most of the synset, a brief definition (gloss) is provided.

The success of the English WordNet has paved way for the emergence of several projects with the aim constructing WordNets in various languages and developing multilingual WordNets. EuroWordNet, a conglomeration of WordNets in European languages is an important project that has come up with a multilingual WordNets. Similar efforts are underway in Indian languages. Hindi WordNet is leading the way for all Indian language WordNets under IndoWordNet.

WordNet building activities in Dravidian languages started with the work of Tamil WordNet28 at AU-KBC Research Centre using Rajendran’s (2001) ontological classification of Tamil vocabulary. Work on Dravidian WordNet (comprising WordNets in four major Dravidian languages, viz. Kannada, Malayalam, Tamil and Telugu) started during a Workshop29 held at Chennai in which synsets were built for Construction Domain. Currently Dravidian WordNet30 activity is being carried out for Kannada at University of Hyderabad, Malayalam at Amrita Vishwa Vidyapeetham, Tamil at Tamil University and Telugu at Dravidian University.

28 Project partially funded by Tamil Virtual University.
29 Workshop on WordNet for Dravidian Languages organized from 2-3 June 2003
30 Project funded by the Ministry of HRD, Govt. of India.
Construction of Synsets

Various approaches are followed in the construction of WordNets across the languages of the world. For the Indian languages, WordNets are constructed using the expansion approach. For the construction of WordNet in Tamil and Telugu, we also follow the expansion approach, i.e. Hindi WordNet synsets are taken as a starting point of departure. The concepts provided along with the Hindi synsets are first conceived and appropriate concepts in Telugu are manually provided by language experts. The Telugu synsets are then built based on the concepts created keeping in view the three principles, viz. Minimality, Coverage and Replaceability.

Below we present the challenges/problems faced in the construction of core synsets in general and kinship synsets in particular.

Lexicographical Concerns

As mentioned earlier, Hindi concepts are first conceived and appropriate concepts are provided along with synsets in target languages. But during this process, problems occurred when we are faced with concepts that have no appropriate equivalents. For example,

Concept (HWN ID 7531): चालीम मेर की एक तोला

cAlisa sera kI eka taula  ‘a measure of 40 kg’.

For this concept, there is no corresponding equivalent in Tamil or Telugu. But there are varying usages in different dialects. For instance, in Kuppam, the measure is equal to 10 kg whereas in Kadapa district of Andhra Pradesh it is 14 kg, whereas in Tamil Nadu it is 10kg. When it comes to providing equivalent synsets, Hindi and Telugu and Tamil uses the same, i.e. maNu.

Concept (HWN ID 24): मादा शेर

mAdA Sera ‘a female tiger’

For the above concept, there is no problem in assigning equivalent concepts and synsets, but when it comes to providing equivalent sentences (which we mostly translate for developing parallel corpora), we come across difficulty. In most cases, we do not use gender while providing translations. This is not the case with this concept alone, but in all the concepts involving gender. For example, female rat (HWN ID 335), female parrot (HWN ID 1278), etc.

Kinship in WordNet

In English WordNet, kinship is represented in the following way: Kinship is a kind of relatedness or connection by blood or marriage or adoption. It can be classified as: 1) **affinity or phylogenetic relation** which is in biology means a state of relationship between organisms or groups of organisms resulting in resemblance in structure or structural parts; 2) **descent, line of descent, lineage, filiation** which is the kinship relation between an individual and the individual's progenitors; 3) **affinity** which in anthropology is the kinship by marriage or adoption and not a blood relationship; 4) **consanguinity or blood kinship** which in anthropology is the kinship that is related by blood; 5) **parentage or birth** which is the kinship relation of an offspring to the parents; 6) **fatherhood** which is the kinship relation between
an offspring and the father; 7) motherhood which is the kinship relation between an offspring and the mother; 8) sisterhood which is the kinship relation between a female offspring and the siblings; 9) brotherhood which is the kinship relation between a male offspring and the siblings; and 10) marital relationship which is the relationship between wife and husband.

When we come to WordNet building activities in Indian languages, Hindi WordNet which is the pioneer is taken as the source language for building WordNets. That is, Hindi WordNet’s concepts are taken as a starting point and WordNets are getting built using the expansion approach. In the outset, this approach looks trivial and economical considering the interlinking of synsets of different languages. But when it comes to kinship relations, however, Hindi concepts create problems while assigning synsets in Dravidian languages, especially Tamil and Telugu. Let us examine a few in the following pages.

Problems in creating synsets involving kinship relations

In all, there are 54 concepts in Hindi WordNet which involve kinship relations. For the purpose of this paper, only the problematic concepts are taken into consideration which needs special attention.

1. HWN ID 7379: भाई का लड़का /bhai ka ladka/ ‘brother’s son’
   Synsets: “भाई, भाभ, भातूर, भातूर, भाती, अवंतंस, अवंतस्म”
   /BatIjA, BrAtRuja, BrAtAputra, BrAtRuputra, BatIja, avataMsa, avatansA/

   When it comes to Tamil, providing concept is a problem. Straightforward, one can assign பாய் /pāy/ பாய் /pāy/ பாய் பாய் /pāy/ பாய் /pāy/. But when assigning synsets, one comes across ambiguity in the concept. i.e. which brother’s son, whether one has to provide பாய் /pāy/ or பாய் /pāy/ or பாய் /pāy/ or both. The problem is the same in Telugu too.

2. HWN ID 1804. भाई की लड़की /bhai ki ladhki/ ‘brother’s daughter’
   Synsets: भाभी, भातूज /BatIjI, BrAtRujA/

   This is similar to the one given in 1 above; whether one has to provide பாய் /pāy/ or பாய் /pāy/ or பாய் /pāy/ or பாய் /pāy/ or both. The problem is the same in Telugu too.

3. HWN ID 683: मामा की लड़की /mAmA kI laDakI/ ‘uncle’s daughter’
   Synsets: मामी बहन, मामी बहन, मामी बहन, मामी बहन
   mamerI bahana,mamerI bahina,mAtuleyI,mamerI BaginI

   The concept in Telugu can be given as /mAmA kUturu/ But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in case of Telugu, if it is elder daughter then it is /vodhina/ and younger is /maradalu/.

4. HWN ID 2861: मामा का लड़का /mAmA kA laDakA/ ‘uncle’s son’
   Synsets: मामी भाई, मामी भाई, मामी भाई, मामी भाई
   mamerA BAI,mamiyAuta BAI,mAtuleya

   The concept in Telugu can be given as /mAmA koDukku/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in case of Telugu, if it is elder son then it is /bAva/ and younger is /bAvamaridi/.
5. HWN ID 685: పూర్ణ కా లడుకా /PUPA kA laDakA/ `aunt’s son’

Synsets: పూర్ణంబాలు,పూర్ణంబాలు,పూర్ణంబాలు,పూర్ణంబాలు,పూర్ణంబాలు,పూర్ణంబాలు,పూర్ణంబాలు

PuPerA BAI, PuPerA BaiyA, PuPerA BaiyA, pitRuShvasrAya

The concept in Telugu can be given as /atta koDukku/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in Telugu, if it is elder son then it is /bAva/ and younger is /bAvamaridi/.

6. HWN ID 686: పూర్ణ కా లడుకి /PUPA kl laDakI/ `aunt’s daughter’

Synsets: పూర్ణంబాల్లరి బాహినీ, పూర్ణంబాల్లరి బాహినీ, పూర్ణంబాల్లరి బాహినీ

PuPerI bahana, PuPerI bahina, PuPerI BaginI

The concept in Telugu can be given as /atta kUturu/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because in Telugu, if it is elder daughter then it is /vodhina/ and younger is /maradalu/.

7. HWN ID 9550: ब्राह्मण के पिता या पिता के भानोई /buA ke pati yA pitA ke bahanoI/

`the brother of your father or mother; the husband of your aunt’

Synsets: /PUPA/

This is an interesting concept in Hindi. In Hindi, the brother of one’s father or mother as well as the husband of one’s aunt is /PUPA/ but when it comes to Dravidian language they are to be given as different concepts. i.e. `the brother of one’s father’ as one concept; `the brother of one’s mother’ as one concept and `the husband of one’s aunt’ as separate concept. Let us detail these concepts.

a. `brother of one’s father’ is लिंग
b. `brother of one’s mother is लिंग
c. `the husband of one’s aunt’ is लिंग

8. HWN ID 4673: पति का भाई /patnI kA BAI/ `wife’s brother’

Synets: sAlA, sAra, nakalaparavAnA, SvaSurya

The concept in Telugu can be given as /bAriya sOdaruDu/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in Telugu, if it is elder brother then it is /bAva/ and younger is /bAvamaridi/.

9. HWN ID 6365: पति की बहन /patI kl bahana/ `wife’s sister’

Synsets: nanada, nanadI, naMdinI, nandinI, nanada, nanadI, naMdinI, nandan, nandan

The concept in Telugu can be given as /barta sOdari/. But, when it comes to providing synsets, one faces the problem in elder-younger distinction. Because, in Telugu, if it is elder sister then it is /vodhina/ and younger is /maradalu/.
10. HWN ID 7194: वह जो संबंध के विचार से किसी के बहन का पुत्र हो

/vaha jo saMbaMdha ke vicAra se kisI ke bahana kA putra ho/

`... sister’s son’

Synsets: भानजा, मांजा, भानजा, मासिनेय, बहनीता, बहनीता

BAnajA,BAMjA,BAnjA,BAgineya,bahanautA,bahanota

In the above example, the Hindi concept is `in relationship, it is anybody’s sister’s son’. But if one has to provide an appropriate concept in Telugu, then we have to split the concept into two. That is, if it is one’s (male) sister’s son, then the synset is /menalludu/ and if it is one’s (female) sister’s son, then /kodukku/. This is the same in Tamil, viz. கோப்பு and வல்லூர்.

11. HWN ID 7195: संबंध के विचार से किसी के बहन की पुत्री या पुत्र की पुत्री

saMbaMdha ke vicAra se kisI ke bahana kl putrI yA nanada kl putrI

Synsets: भानजी, मांजी, मांजी, बहनीती, मासिनेया

BAnajI,BAMjI,BA~MjI,bahanautI,BAgineyA

The Hindi concept is more general, i.e. it does not make a distinction between the speaker, whereas when it comes to Telugu and Tamil, one has to make the distinction. If the speaker is a male, then his sister’s daughter is கோப்பு, if it is a female, then her sister’s daughter is வல்லூர்.

In the above example, only few of the problems faced while constructing in Telugu and Tamil using Hindi concepts are presented. The reason for this is two-fold, viz. first, Hindi concepts are built based on most commonly used words, so one can see how shallow the synsets are and secondly, it is a problem of two different language families in the expansion approach. In Trauttman (1995) words, one needs a hierarchy of constructs, the genetic constructs — Dravidian and Indo-Aryan — and the synthesizing construct of Indian kinship to distinguish the Dravidian data from the non-Dravidian.

Summary

While WordNets are being developed for almost all the major Indian languages, one should give special attention while constructing synsets which is the core of WordNet. As seen from the examples above, it exemplifies that constructing WordNet based on Indo-Aryan languages into Dravidian languages is not a trivial task. It is also clear that kinship relation in Hindi WordNet is shallow and hence one has to take into account different culture specific constructs into consideration. This brings us to some interesting challenges in the construction of WordNets: How to integrate language independent constructs in IndoWordNet; How to handle problems such as male-female distinction in IndoWordNet and vice-versa; elder-younger, distinction which is prominent is Telugu but not in a cognate language such as Tamil? How to represent kinship hierarchy in WordNet? The only solution at this point of time is to build a Domain Ontology within the framework of WordNet Domains.

References


233
காசியின் பாண்டிய பாண்டியம்

7/1, மேலூர் புதுக்கோட்டை நினை
மலைப்பாண்டி, ஊராட்சி 600 004

திரு: 98844 81652 / 044 6499 3317
thiru2050@gmail.com / thiru2050@yahoo.co.in

காசியின் பாண்டிய பாண்டியம், தமிழகத்தில் காசியின் பாண்டியம் பாண்டியம், அதிகரியில் இருந்தது. அதிகரியுடன் காசியின் பாண்டியத்துடன் பாண்டியப் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம்

1. உதவியை மீண்டும் காசியின் பாண்டியம்
2. உதவியை மீண்டும் காசியின் பாண்டியம்
3. உதவியை மீண்டும் காசியின் பாண்டியம்
4. உதவியை மீண்டும் காசியின் பாண்டியம்
5. உதவியை மீண்டும் காசியின் பாண்டியம்
6. உதவியை மீண்டும் காசியின் பாண்டியம்
7. உதவியை மீண்டும் காசியின் பாண்டியம்
8. உதவியை மீண்டும் காசியின் பாண்டியம்
9. காசியின் பாண்டிய பாண்டியம்
10. காசியின் பாண்டிய பாண்டியம்

விளக்கம்: அதிகாரியுடன் பாண்டியம், தமிழகத்தில் காசியின் பாண்டியம் பாண்டியம் அதிகாரியுடன் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம்

முக்த உள்ளூரை காசியின் பாண்டிய பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் பாண்டியம் 

235
5.3.2.7.1.

1. உணவுகளிலை உணவுகள் என்றும் காணப்படுகிறது.

2. மேலும் இவைகள் உணவுகள் என்றும் காணப்படுகிறது.

3. காய பயிற்சிக்குத் திறன் பெறுவதற்கு உரியது என்றும் காணப்படுகிறது.

4. எச்சாரம் கணவிட்டு செய்யப்படும், நீண்டு இராச்சியத்திலை காணப்படுகிறது.

5. பயிற்சிக்குள் கணவிட்டு உரியது என்றும் காணப்படுகிறது.

6. ஏற்றுள் செய்யப்படும் காணப்படுகிறது.

7. சிறுவர்கள் உணவு பெறும்போது கனவிட்டு உரியது என்றும் காணப்படுகிறது. அவர்கள் தமிழில் பெறுவதற்கு எச்சாரிலை தமிழில் ாசிக்கிறது இருக்கும் என்றும் காணப்படுகிறது.

ஆங்கிலாக, மீச்சருளில் பிராம் கணவிட்டு உரியது என்றும் காணப்படுகிறது. கணவிட்டது செய்யப்படும் எச்சாரிலை தமிழில் காணப்படுகிறது. பெறும் பலுவைகளில் காணப்படுகிறது. பெரும்பாலை செய்யப்படும் எச்சாரிலை தமிழில் காணப்படுகிறது. இயற்கையான கணவிட்டு உரியது என்றும் காணப்படுகிறது. வண்ணம் பெறுவதற்கு எச்சாரிலை தமிழில் காணப்படுகிறது. 

ஆண்டு குழந்தைகள் பெறும் எச்சாரிலை தமிழில் காணப்படுகிறது. பெரும்பாலை செய்யப்படும் எச்சாரிலை தமிழில் காணப்படுகிறது. இயற்கையான கணவிட்டு உரியது என்றும் காணப்படுகிறது. 

சிறுவர்கள் உணவு பெறும்போது கனவிட்டு உரியது என்றும் காணப்படுகிறது. அவர்கள் தமிழில் பெறுவதற்கு எச்சாரிலை தமிழில் காணப்படுகிறது. கணவிட்டது செய்யப்படும் எச்சாரிலை தமிழில் காணப்படுகிறது.
சொல் கவர்களை பெருந்து டித்துக் குறிப்பிட்டோம் பதவியாளர் ஒருவர் கூறாத் தொடர்பில் உள்ளது பாதுகாப்பு நோக்கில் உலக அரசியல் கட்டுரையாளர் கூறுகிறார். குறியீடு தொடர்பிலும் பெரும் இருந்துள்ளன. இச்சொல்லை அறிக்கை என்று பயன்படுகின்றது போக்கில் இணைந்தவை இதுவாகும் நோக்கில்.
தொடர் விளக்கம் ஆலங்காரங்கள் பதிப்பிறப்பட்டுள்ளது. அதாவது இது எழுதியபெயரால் விளக்கம் செய்யப்பட்டுள்ளது. கவிதைகளும் மறு விளக்கம் செய்யப்பட்டுள்ளது. ஆனால் தமிழில் கவிதைகளை விளக்கப்பட்டது. நூற்றாண்டில் பதிப்பிறப்பட்டுள்ளது.

கைவழியாராம் குறிப்பிட்டு உள்ளூர் தொடர் விளக்கம். கைவழியாராம் குறிப்பிட்டு உள்ளூர் தொடர் விளக்கம். கவிதைகளும் விளக்கம் செய்யப்பட்டுள்ளது. ஊட்டம் விளக்கம் செய்யப்பட்டுள்ளது.

பிரிவுகள்

1. குறியீடு - பார்வை: பார்வைகள் வரும் நூற்றாண்டு பதிப்பிறப்பட்டுள்ளது (தமிழ் தமிழ் முன்னேற்றங்கள், வேலைவிளை)
2. நூற்றாண்டு: விளக்க வழக்கங்கள் - விளக்க வழக்கங்கள் (தமிழ் தமிழ் முன்னேற்றங்கள், வேலைவிளை)
3. கவிதைகள் வேலைவிளை: வேலைவிளை வேலைவிளை (தமிழ் தமிழ் முன்னேற்றங்கள், வேலைவிளை)
4. வேலைவிளை - வேலைவிளை: வேலைவிளை
5. ஆனால்: வேலைவிளை - வேலைவிளை
6. கவிதைகள் - வேலைவிளை: வேலைவிளை (தமிழ் தமிழ் முன்னேற்றங்கள், வேலைவிளை)
7. கவிதைகள் - வேலைவிளை (தமிழ் தமிழ் முன்னேற்றங்கள், வேலைவிளை)
8. Computer Dictionary (English - Tamil) - இழையார்; இழையார் குறிப்பிட்டு குறிப்பிட்டு
10. பதிப்பிறப்பு: பதிப்பு - பதிப்பு: பதிப்பு
11. கவிதைகள்: கவிதைகள் - கவிதைகள்: கவிதைகள்
12. பதிப்பு - பதிப்பு: பதிப்பு - பதிப்பு
Role of Regular expression (RE) in Morphological Analysis

R.Shanmugam
Madras University
shanfaace@gmail.com

Abstract
The aim of this article is to analyze the role of Regular Expression in Tamil Morphological Analysis. Morphological Analysis is essential for Natural Language Processing (NLP) and Machine Translation (MT). In morphological analysis, what we do is to parse the inflected word into root and affixes and then to tag them for grammatical categories. To build a Syntactic Parser, we need a Morphological Parser with POS (Parts-of-speech) tagger. The input for the Syntactic Parser is the output of the Morphological Parser. There are many formalisms and tools are used in the field of Morphological Analysis. Regular expression is one among the best tools for Morphological analysis.

Introduction
Regular expression is the standard notation for characterizing strings (combination of characters). It is a formula in a special language for specifying simple classes of strings. Formally it is an algebraic notation for characterizing strings. Regular expression was introduced by Kleene (1956). A string is any sequence of characters like letters, numbers, spaces, tabs, punctuation. Space which is also a character because it has encoding value. Regular expression needs a pattern (search type) to search strings. The following table shows how the words are matched by regular expression.

<table>
<thead>
<tr>
<th>Regular Expression</th>
<th>Example pattern matched</th>
</tr>
</thead>
<tbody>
<tr>
<td>/putthakam/</td>
<td>avaN putthakam patiththaAN</td>
</tr>
<tr>
<td>/kalvi/</td>
<td>kalvi aNaivarukkum avaciyan</td>
</tr>
<tr>
<td>/niir/</td>
<td>kutikka niir veeNntum</td>
</tr>
</tbody>
</table>

So the regular expression /putthakam/ matches any string containing the substring putthakam like the above example. Here the slashes (around the pattern) are used to clarify what is regular expression and what is pattern. (This notation used in perl). The upcoming table shows some of the important Regular Expression symbols which are used in Tamil morphological parsing.

Patterns in Regular expression: (Based on Perl language)

<table>
<thead>
<tr>
<th>RE symbol , Name</th>
<th>Example pattern</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>~ , tint</td>
<td>marankkaL =~ /kaL/</td>
<td>Pattern matching symbol finds if ‘kaL’ is the substring of marankkaL</td>
</tr>
<tr>
<td>( ), Open and</td>
<td>/marank(kaL</td>
<td>ai)/</td>
</tr>
</tbody>
</table>
Levels in Morphological Analysis

There are three levels involved in Morphological Analysis such as 1. Root checking, 2. Affix Stripping, 3. Morphophonemic changes. We can use Regular Expression for the above three levels. It has much influence in Morphological Analysis. It is a very handy tool to make this kind of Morphological analysis. It can be implemented easily with Perl language.

The Role of Regular expression in Root checking

This is a basic task in Morphological Analysis. Under this task, the root word existence in the database would be found. The following code may useful to know the importance of RE in this task.

```perl
open(filehandle,"C:\\DataBase\\Noun.txt") || die;
while ($line=<filehandle>)
{
  if($line=~m/$input\b/) {
    $root word=$&;
    last;
  }
}
```

Here the line in bold is important. The character ‘m’ is stands for matching i.e., to verify whether the database word is same with the end user input. ‘\b’ is used to denote the word boundary. ‘$&’ is called register in RE it is used to store the matched word into the variable $root word.

The Role of Regular expression in Affix Stripping

Affix stripping is used to strip the affixes from the given input. The following sample coding explains the role of RE in Affix Stripping section.

```perl
Stripping tense suffix from the input stem “patithth”
if($input=~/(v|N|t|R)$|(pp?)$|(iN)$|(m)?(th)+)$|(kk?iN|R)$/) {
  $tense=$&;
  $root=substr($input,0,-length($tense));
}
```
The above code is having 14 tense forms for stripping purpose. It is a great strength of RE. We can implement this in one line and it gives the following output.

```
$input (Input variable) = patithth
$tense = thth
$root = pati
```

**The Role of Regular expression in Morphophonemic Section**

This section is playing the key role for this programming. This section includes three sub-functions such as addition, deletion, substitution. The three functions would be used to make the remaining stem into root. If we want to change the stem ‘marathth’ into root ‘maram’ the following code is needed.

```
$input = marathth
if(($input=~m/thth$/) and ($case ))
{
    $input=~s/thth/m/;
}
```

**Output:** $input = maram

Here the bold lines are showing the importance of RE. ‘if’ condition is used to identify the occurrence of ‘thth’ in input and substitution is used to change the stem ‘marathth’ into root ‘maram’.

**Conclusion**

Regular expression is a handy formalism to make Morphological Parser and it is common for languages we can adopt these ideas for other languages too, so that if we have a clear view on this we will make a good Morphological Analyzer for our language.
கணினியில் கண்டுபிடிப்பாற்றுக்கொள்ளும்
பதிப்புடன் 2 இடங்களம்

என். வரலாற்று, கிளீ. எந்த. போன்,
கேள்வி எழுதுவருபவர்,
தலைவர் முன்னாளார் கூட்டம், முந்தியிருப்பாளர்-02.
அல்மிருது:9842599311
charlesjc85@gmail.com

பிரிவு சாலை ஆசிய சுருக்கத்தில் நேர தொழில்முனை வருகையில் கல்வி குறுக்கு கல்வியறிஞர். என்பு கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். இது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். அது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். அது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். அது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். 

கல்வி சரக்கு மேல் செய்வதால் கிளைசளத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். என்பு கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். இது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். அது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர். அது கிளைசளத்தில் பாண்டு சுருக்கத்தில் நேர தொழில்முனை கல்வி குறுக்கு கல்வியறிஞர்.
பாதுகாப்பு (Terms)
244
Control unit - CPU - Memory Unit - CPU.

LAN - Local Area Network

WAN - Wide Area Network

CPU – Control Unit

Compact Disk (CD) - Computer

Hard Disk - Computer

Software - Computer

Hardware - Computer

Software & Hardware - Computer

Control unit - CPU - Memory Unit - CPU.

CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.

Control unit - CPU - Memory Unit - CPU.
Computer

verb

function

Application

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun

noun
கைவன் குறி தொடர்கள் ஏற்று, குறி சிற்றிலகினும் பட்டியல்கள் தம் பாகங்கள் அவர்கள் கைவன்
கலப்புற்றுறைச் செயல்களில் புதுக்குறியுடன் பார்க்கப்பட்டன ஆனால் கைவன் குறியீடுகளுக்குப் பதிலளித்த 
நேர்முறை பயன்படுத்தப் பட்டியல் முறை எனும்.

தலைமுறை தேவை

1. இயற்கையியல், கைவன் கலப்புற்றுறைகள், குறிப்பிட்டு பதிவு செய்த (மி) முதல், அம்பகுடன், நூற்றாண்டு-98, 2005.
2. இயற்கையியல், கலப்புற்றுறை, குறிப்பிட்டு பதிவு செய்த முதல், அம்பகுடன், நூற்றாண்டு-98, 2006.
3. மொழிபெயர்க்குறியுடன், பார்க்க, குறிப்பிட்டு பதிவு செய்த, முதல், அம்பகுடன், நூற்றாண்டு 1957.
4. பார்க்க, பார்க்க, குறிப்பிட்டு பதிவு செய்த, பார்க்க, அம்பகுடன், நூற்றாண்டு-14, 2009.
5. கைவன் பல்கலை, பார்க்க, பார்க்க, குறிப்பிட்டு பதிவு செய்த, கைவன் பதிவு, நூற்றாண்டு-14, 2003.
Transliteration Schemes for Tamil to Roman and Roman to Tamil characters

Dr. S. Srinivasan
Scientific Officer, Computer Division
Indira Gandhi Centre for Atomic Research
Kalpakkam-603102, Tamilnadu, India
srini@igcar.gov.in

Introduction

Here a machine transliteration scheme is proposed to map the random striking on a Tamil keyboard to Roman equivalent characters and vice versa. Many a language spoken in this world is associated with a native script. However a group of languages might also share a common script. For instance, the Roman script is shared by a number of European languages viz., English, French, German, Italian and Spanish. If there is a means to spell out a foreign language using ones native script it lessens the burden of mastering another script and speeds up the process of learning. In such a situation, the transliteration scheme comes handy and helps to overcome the difficulty of knowing yet another script.

The Tamil keyboard layout

The mechanical Tamil typewriter consists of 4-tier structure and encompasses 23 consonants, 12 vowels, about 12 vowel modifiers, about 20 ukara vowel-consonants, one medial and a conjunct (Sri). Besides these the keyboard also contains shift key, dead key, caps lock and space bar.

There are 4 schemes that are widely used for Tamil transliteration.

They are:

1. Madras University Tamil Lexicon Scheme- ISO 15919 Standard
   (Based on lower ASCII and a few diacritical markers)
2. Library of Congress scheme (based on lower ASCII)
   (a) It uses markers from upper ASCII block, i.e., characters 128-255.
   (b) Variant of Library of Congress scheme (uses special characters such as # $ _ )
3. University of Koeln-Institute of Indology and Tamil Studies scheme (based on lower ASCII, case sensitive and also uses digit 2)
4. ITRANS developed by Avinash Chopde (uses special characters such as ~ ^ besides other roman characters)

The objectives of the transliteration scheme are as follows:

1. The transliteration scheme could be case sensitive but shall not employ any special character or digit.
2. The transliteration must be unambiguous (loss less) and is primarily meant for machine automation.
3. The transliteration may not be ideally suited for human reading. Hence its user-friendly aspect is kept secondary.
**Tamil-to-Roman transliteration**

Imagine an ape to snatch away a Tamil keyboard from a human hand and whimsically attempts to tap the keyboard to generate a sequence of characters. The strings of characters that are generated in the process may not be meaningful and some of them even contain a number of vowel modifiers occurring in succession (e.g. kombu,kaal,kombu...). To map such a sequence of modifiers, an improved rigid transliteration scheme is proposed. It is basically a variant of the UKoeln scheme.

None of the existing transliteration schemes are 100% romanized. All these schemes use either diacritical marker or special character or digit in addition to Roman letters. The UKoeln-IITS scheme uses the digit 2 to map the consonant Rannagaram and not any other special characters. Also this scheme does not differentiate between the pure vowel and its modifier. Hence attempt to improve upon this scheme was pursued.

The mapping for 'Rannagaram' was changed from n2 to nx so that the scheme becomes fully romanized. The choice for the letter x in lieu of 2 is the following:

nx => n(ExCHANGED) => 'tannagaram' exchanged to 'Rannagaram'

As the last bogie of a passenger train is indicated by the symbol X so also one can construe x to be the last consonant of archaic Tamil(ý) and capital X to be the augmented last consonant of Grantha(Œ) to the Tamil character set.

The word குரான் (kurAn) was not properly transliterated from Tamil to Roman and then back to Tamil. Recasting the pure vowels from the vowel modifiers solved this problem. The vowel modifiers used in UKoeln-IITS scheme are:

a A i I u U e E ai o O au

To make the pure vowel representation unique the respective vowel modifiers were prefixed with capital Y and the result is,

Ya YA Yi YI Yu YU Ye YE Yai Yo YO Yau.

The choice of the letter Y springs from the fact that the pure-vowel sounds close to the yakara vowel-consonant(uyirmey).

e.g. ராஜ் => ராஜ்; ராஜ் => ராஜ்; ராஜ் => ராஜ்; ராஜ் => ராஜ்; ராஜ் => ராஜ்;
   ராஜ் => ராஜ்; ராஜ் => ராஜ்; (யும்படைப்பொருள், புர. 1955)
   ராஜ் => ராஜ்; ராஜ் => ராஜ்; ராஜ் => ராஜ்; ராஜ் => ராஜ்;
   Yield => ராஜ்; Yes => ராஜ்; Yellow => ராஜ்;

The prefixing of the letter Y for pure vowels may not appear weird if one compares the following transliterated words:

e.g. காட்டி => kaiTai ; காட்டி => kai Tai

Also in Tamil the pure vowels occur at the beginning of words. They seldom occur in the middle of words (aLapeTai is the exception). On the contrary the yakara vowel-consonants seldom occur at the
beginning of words (yA is the exception). Case markers and words beginning with pure vowels when combine with a preceding Tamil word they change to yakara vowel-consonant.

e.g.

Hence to differentiate between the pure vowel and its modifier the letter Y was chosen to be a part of the pure vowel. In English there are only two nasals (n and m). But in Tamil there are six nasals. Hence many of the nasals in Tamil need to be represented using more than one roman letter. Three of the nasal consonants (í ï ý) and a grantha consonant (%‰) have to be represented by two Roman letters. In such cases the second letter (trailing part) was chosen to be unique and did not figure in any of the other consonant representations. If any phonetic semblance required its alternate case was chosen. The scheme employs 14 upper case letters of which 5 are used to represent the long vowels, one as prefix to pure vowels, 7 for consonants and one for conjunct vowel-consonant.

A I U E O, Y, N T R L J H X, S

The proposed transliteration table is given below.

A few of the problematic words are transliterated herewith.

e.g.

A few of the problematic words are transliterated herewith.

(e.g.

Yaintu

YaYiYatimuka (a popular political party in Tamilnadu)

kurYAnx (from kriya Tamil dictionary)

kurAnx

YAkA ! => YAKA !

YAkAAA

(one finds such usage in modern Tamil short stories- akin to emoticon usage in SMS and emails)
Also the split vowel modifier of aukaaram(OCR) is construed as a combination of the modifiers kombu(\textbar) and kaaI(¡): \textbar + ¡ => eA; \textbar\¡ => eke [archaic Tamil usage]: \textbar\¡ => " (\textbar\¡, 1981)

Evidence to support this contemplation exists. (\textbar\¡, 1993)

The refinement of Tamil characters is also going on at slow pace. Reverend Veeramaa Munivar for the first time in modern era brought out orthographic refinement in Tamil. He devised a method to remove the ambiguity in the appearance of long and short vowels of ekaram and okaram about 250 years ago. The next refinement of Tamil characters initiated by the Tamilnadu Government and coinciding with the birth centenary year of Periyaar (E.V.Ramasamy) took place in the year 1979. To refine pure vowels with minimal change R.Krishnamurthy, the editor of Thinamalar Newspaper advocates the following modification for long vowels (\textbar, ¡, 1978).

Ya YaA Yi YiA Yu YuA Ye YeA Yo YoA

For instance Thinamalar newspaper recasts the Ukaarams œ, ß, à, á, æ, ê, ë to š¡, Ï¡, С, Ñ¡, Ö¡, Ú¡, Û¡ and implements them in its edition. The aforesaid roman transliteration scheme is robust enough to take care of any foreseeable refinements in Tamil characters as well. The following is a sample text taken from Bharathidasan’s work and whose transliteration is also given alongside.

\begin{tabular}{ll}
\textbullet & Yazupavanx kOzai \\
\textbullet & YAvinx pAl Yinxitu \\
\textbullet & Yiravinxil t\mbox{"Un}gku \\
\textbullet & Yl vatu makizcci \\
\textbullet & YuLLataip pEcu \\
\textbullet & Y\mbox{"U}maippOl YirAT\E \\
\textbullet & Yetaiyum YUnxRip pAr \\
\textbullet & YEcEl Yevaraiyum \\
\textbullet & YaintiR kalai payil \\
\textbullet & YoRRumai vellum \\
\textbullet & YOramp\O teruvi \\
\textbullet & Yauvai tamizttAy \\
\textbullet & kaNakkil tErccikoL \\
\textbullet & cariyAy Yezutu \\
\textbullet & tamiz Yun tAymozi
\end{tabular}
Roman-to-Tamil transliteration

A rigid transliteration scheme is contemplated for Roman-to-Tamil conversion for the benefit of Tamils who can read only Tamil characters but wish to read English text as well. A parallel exists to this. Historically Tamils invented an alternate script to read Sanskrit text. It was called the Grantha. The Tamils preferred this script in lieu of the Devanagari script as it contained the Tamil alphabet too. In the same lines, an attempt has been made to map both the upper and lower case Roman letters (English characters) into Tamil.

Let us perform another thought type experiment. Imagine another ape belonging to North-Atlantic region to grab away a Roman keyboard from a human hand and strikes whimsically on the keyboard. Also assume that the digit and special character keys are disabled from the keyboard. In such a situation too, the string of characters that are generated may not appear meaningful and could even contain a large mix-up of both upper and lower case letters as nouns appear in German text. To map such a sequence of letters into Tamil a transliteration scheme is proposed. The contemplated case inclusive transliteration table is given below.

<table>
<thead>
<tr>
<th>Lower Case</th>
<th>Upper Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>a b c d e f g h i</td>
<td>A B C D E F G H I</td>
</tr>
<tr>
<td>சி சி சி சி சி சி சி சி சி</td>
<td>சி சி சி சி சி சி சி சி சி</td>
</tr>
<tr>
<td>j k l m n o p q r</td>
<td>J K L M N O P Q R</td>
</tr>
<tr>
<td>சி சி சி சி சி சி சி சி சி</td>
<td>சி சி சி சி சி சி சி சி சி</td>
</tr>
<tr>
<td>s t u v w x y z</td>
<td>S T U V W X Y Z</td>
</tr>
<tr>
<td>சி சி சி சி சி சி சி சி சி</td>
<td>சி சி சி சி சி சி சி சி சி</td>
</tr>
</tbody>
</table>

With the aid of the above table the following English words are transliterated into Tamil.

apple- அப்பல்
APPLE- அப்பல்
box- ஬ொக்
BOX- பொக்
cat- சேத்
cat- சேத்
dog- டொக்
dog- டொக்
The reading of this transliteration may seem difficult in the initial phase. This situation is akin to asking a German to read English text or an English man to read German text. But this difficulty can be overcome in due course with adequate practice. The poet Bharathidasan himself felt that the alphabetical system of Tamil could be simplified if all the vowel consonants were split into pure vowels and consonants (1948).

He tried even writing so; but discontinued this effort in due course.

In a nutshell the motive behind this scheme is the following:

1. To transliterate the lower case Roman letters into Tamil the following Tamil letters are used. They are short vowels, consonants, medial and a few ukara vowel-consonants.

2. To generate the mapping for the upper case Roman letters from the lower case Roman letters is simple. They consist of either long vowels or pure consonants.
Applications

A random sequence of characters that are generated from a keyboard can be faithfully transliterated to another script and can also be used as a substitute for a password in computer applications. A password of this kind is all the more difficult to crack by a hacker. Transliterating a text to a non-native script and then encrypting it would add to the level of data security. The mapping from Tamil-to-Roman involves variable length and dual case and hence once encrypted it is difficult to decrypt. This feature enhances the data security in transmission as well as storage.

Conclusion

A machine transliteration scheme is proposed to map all the random sequence of characters that could be generated from a Tamil or a Roman keyboard. The transliterated characters could further be encrypted so as to increase the level of security.

Acknowledgements

The author expresses thanks to Dr.Vasu Renganathan, University of Pennsylvania, Philadelphia, USA and Dr.K.Kalyanasundaram, Lausanne, Switzerland for their help in providing the various romanized transliteration standards available for Tamil and offering valuable comments on this work.

References

1. URL: http://homepage.ntlworld.com/stone-catend/translit.htm
2. URL: http://www.aczone.com/itrans/tamil/node5.html
9. பாலாராஜ், 1948, தமிழ் கிருஷ்ணேஸ்வர், புதுக்கோட்டை.
From Classical Tamil to Computational Tamil
A Perspective

Abstract

The area of Computational Linguistics deals with computational models that are employed for analysis, synthesis or transformation of content in natural languages. Many well known end user technologies, such as, language understanding, machine translation, monolingual and crosslingual information retrieval and extraction, etc., are based on such models. Given the exponential growth of content in the Internet and Social Media primarily in vast majority of languages of the world, it is highly imperative that tools and technologies be developed to process the natural language data effectively and efficiently. In this paper, we highlight the state-of-the-art approaches for Computational Linguistics that are primarily based on statistical and machine learning principles, and underscore the need for clean large annotated corpora and language resources for any and all types of Computational Linguistics research. In particular, we emphasize the need for corpora, basic tools and resources in Tamil, in order to ensure the development of technologies in the Tamil language. It is imperative that the community, academia, industry and the government come together to create a climate of consensus, coordination and collaboration to make sure that Tamil is taken successfully to the computational world.

Introduction & Motivation

The area of Computational Linguistics deals with computational models for analysis and synthesis of natural languages, and is a vital predecessor for many natural language processing tasks, such as language understanding, summarization, information retrieval and extraction, machine translation, etc. Given the exponential growth of the amount of available natural languages data due to the Internet and Social Media, it is highly imperative that tools and technologies be developed to process the data effectively and efficiently. More importantly, in countries like India where only about 5% of the people are English literate, the need for such technologies is even more important for including the majority into the Information Age.

Computational Linguists research pertains to development of such tools and technologies. Traditionally, Computational Linguistics research and systems relied on linguistics research resulting in rules that are distilled by experts in a language. For examples, rules that govern morphological variations of a word or formation of a sentence in a given language are devised by experts, and coded into practical tools and systems. However, given that the natural languages evolve, such systems become unmanageable as they are fraught with relatively large portion of exceptions for every rule. Further, such rule-based
approaches are expensive to create and maintain in terms of time and resources, as evidenced by decades of research put in the Western European (WE) and Chinese-Japanese-Korean (CJK) languages.

In the recent decade and a half, a host of newer approaches has been introduced in the Computational Linguistics Research, specifically, those based on Statistical and Machine Learning based methodologies. In these methodologies, specific tasks may be learnt automatically when provided with appropriate hand-crafted training data. These methodologies are broadly referred to as statistical learning or machine learning algorithms. For example, identification of names or places in a sentence may be learnt (with a certain level of accuracy) by programs that are trained on large hand-annotated corpora, so that they may be used subsequently for identification of names from sentences. While the quality of such depends on several factors: such as, the nature of the task, the algorithms used, features used for training, the quality and quantity of data used for training, etc., still such approaches had been proven to be very effective – as good or better than hand crafted systems for many of the natural language processing tasks. For example, all the state-of-the-art-translation systems in the world now are statistical learning systems.

In addition to being easier to develop, equally importantly, such methodologies are also largely language-independent, paving way for quick adaptation across languages. For example, a generic Statistical Machine Translation (SMT) system may be employed successfully to learn translations between any given pair of languages (with appropriate training data in those pair). Hence, such approaches exhibit a great advantage especially in countries like India, where a single system may be adopted for many languages, quickly and transparently. In essence, these methodologies rely on generic statistical and machine learning frameworks, trained on custom datasets.

Finally, given that the most popular medium for information, entertainment, commerce and governance - The Internet - is also turning multilingual31. The demographics of Internet users have changed from being predominantly English, to more than two-thirds that are non-native English speakers now. In addition, majority of the information available over the web is in a language other than English. Such shifts in demographics suggest that the technologies must be developed for supporting predominantly multilingual user population, pointing to the critical need for language neutral Computational Linguistics research to cater to wider audience, quicker. In countries such as India, we face additional challenges where the population is mostly English illiterate32, hence tools and technologies in local languages are even more important, in order to overcome the digital divide to include the common man.

In the subsequent sections, we specify the type of corpora needed for Computational Linguistics research, and appeal to the Tamil linguistics and computational linguistics community to work toward common sets of standards and corpora to make the research community vibrant and fruitful.

Linguistic Corpora to be Developed

In this section, we outline several types of linguistic standards and corpora that need to be developed for Tamil, to support robust computational linguistics research.

---

32 It is estimated that only about 5-7% of the Indian population is conversant in English.
National Efforts on Linguistic Corpora

National corpora are normally general reference corpora which are supposed to represent the national language of a country. They are collected by a consorted effort by the Government along with Academic and/or Industry players, in a focussed manner. These corpora are balanced with regard to genres and domains that typically represent the language under consideration, in that particular geographic or political domain. While many of the national corpora are available with parts of speech annotation, few of them have syntactic and semantic parses annotated.

The British National Corpus (BNC) is perhaps the first and best-known national corpus. It is designed to represent as wide a range of modern British English. This comprises approximately 100 million words of written texts (90%) and transcripts of speech (10%) in modern British English. In addition to part-of-speech (POS) information, the BNC is annotated with rich metadata (i.e. contextual information). The American National Corpus (ANC) project was initiated in 1998 with the aim of building a corpus comparable to the BNC. The first release of the corpus contains 11.5 million words of written and spoken data. When completed ANC will contain a corpus of (100M) words comparable to BNC corpus. The corpus is POS tagged using different tag-sets to suit the needs of different users. Similarly there are national corpora available in Polish (130.8M), Czech (100M), Russian (100M), Hellenic (32M), German (100M), and Chinese (700M characters) languages.

In India, Central Institute of Indian Languages (CIIL) collected corpus is available in most Indian languages. However, this corpus is a relatively small corpus (approximately 3-8M words per language) primarily a monolingual text collection in multiple languages, with no annotation. While this corpus may provide the seed for data creation, the volume and quality of such corpus needs to be enhanced significantly, to aid Computational Linguistics research in Indian languages.

Recently, Linguistic Data Consortium for Indian Languages (LDC-IL) has been initiated by the Ministry of Human Resource and Development under Government of India, to oversee the standardized collection of linguistic corpora in all Indian languages. Several academic and industrial partners are working together to get this collection created.

Monolingual Corpora

Monolingual corpora essentially refer to normal Tamil language text bulk from a standard source, such as, popular mass media, newspapers, television, etc. While it is good to have a wide variety of content; each genre, say, printed or spoken news, literary works, political speeches, religions writing, etc., each has its own characteristics, and best handled individually.

Ideally the text bulk should be in a standard encoding, such as Unicode, and annotated with some metadata, such as, source, author, date of publication, genre or category, etc. In addition to document level annotation, annotation of the content of the corpus itself could be extremely useful for many Computational Linguistics Tasks. For example, a corpus annotated with names (personal names, common names, places, dates, organizations, etc.), may be used for Named Entity identification tasks, and Information Extraction tasks.
Multilingual Corpora

Multilingual corpora refer to many types of corpora – parallel, comparable, etc. Parallel corpora are essentially sentence aligned corpus in multiple languages, where every aligned sentence pair contains the same semantic information in multiple languages. Such corpora may be readily used for developing Machine Translation systems.

In many practical situations, comparable corpora are more readily available than parallel corpora. Comparable corpora are defined as article aligned corpus in multiple languages, where the article generally is on the same topic, but may have different semantic content. Typical comparable corpora consist of news articles in multiple languages that cover the same news event; since each article may be written by different editor, it is likely to have similar, but not the same semantic content. Comparable corpora had been successfully employed in development of MT systems.

Annotated Corpora

Large annotated corpora are critically needed in any computational linguistics research. The annotation depends on the task at hand; for example, Part of Speech (POS) identification requires a rich annotation where every word in the text corpus is tagged, whereas Named Entity Recognition (NER) requires hand annotation of specific entities in the corpus.

Annotation Standards

Any type of annotated corpora underscores the need for annotation standards, in order to create standard annotated corpora that may be used by many tools and research groups consistently. It is imperative that standards be developed for annotation of the collected data such that the data created is rich enough to support many learning tasks that need to be based on the data, yet, flexible enough to be modified when the need arises. An example of such effort, is given in (Baskaran et al., 2008), where a Part of Speech annotation framework – called IL-POST – was designed collaboratively by a set of academic and industry partners and which is applicable to a variety of Indian languages. Standard frameworks exist, such as EAGLE’s (Leech et al., 1996) for linguistic tagging.

Linguistic Resources

Many resources, such as dictionaries, thesauri, Bilingual or multilingual dictionaries are necessary for obvious reasons, for a variety of tasks. A computational dictionary must be in a standard format (Unicode, XML tagged) and must be machine-readable with standard tags. In addition, all references linking various words (for example, in thesauri) must be navigable using unique identifiers. There are specific requirements for computational dictionaries, as against print dictionaries.

Standards Organizations

Over the last few decades there are many very successful initiatives between governments, industry and academia in developing standards and corpora according to those standards. An example of such an initiative is the Linguistic Data Consortium (LDC) in the University of Pennsylvania.

Creation of Data with Community-wide Participation

It is important to highlight the importance of crowd-sourcing as a methodology for creation of linguistic corpora, as many types of corpora does not need to be created by linguists or language experts, but easily
by the native speakers of a language. An initiative to generate parallel data is outlined in (Kumaran et al., 2009).

The Need for Linguistic Corpora

In this paper, we focused on mining NE pairs in two different languages, namely English and an Indian language, Tamil. While we adopted a methodology similar to that in [Klementiev and Roth, 2006], our focus was on mining parallel NE transliteration pairs, leveraging the availability of comparable corpora and a well-trained linear classifier to identify transliteration pairs. We profiled the performance of our mining framework on several parameters, and presented the results. While the results show the potential of our approach, we also uncovered several issues that need to be resolved, for effective mining of parallel NE transliteration pairs. Given that the NE pairs are an important resource for several NLP tasks, we hope that such a methodology to mine the comparable corpora may be fruitful, as comparable corpora may be available in perpetuity in several of the world’s languages.

References

Automated Identification of Grammatical Patterns For Tamil Poems

Sendhilkumar S., Mahalakshmi G.S., Prakash N

Department of Computer Science and Engineering,
Anna University, Chennai 600025,
thamaraikumar@cs.annauniv.edu,
mahalakshmi@cs.annauniv.edu,
yoursprakashever@gmail.com

Abstract

Tamil is one of the most classical world languages. But in modern teaching, people have started ignoring it. One main reason is that the grammar is comparatively tougher than that of other languages. The aim of this prototype is to build an application which makes teaching and learning Tamil grammar more easily. With rapid advancement in Tamil language processing packages, it is not very tough to build an interactive, user-friendly tool which will help in teaching-learning Tamil grammar in a better manner. The idea is to develop rule-base for calculating பாக்கிக்கர் and later, identification of grammatical patterns present in the Tamil poems as well as classification of Tamil poems based on the structure becomes more flexible.

Introduction

In recent days, a survey on natural language processing tools for Tamil results in a handful of research approaches and solutions, of which a few are officially recognized. There is enough research done to analyze the morphemes of a word or sentence (say, morphological analyzer, morphological generator, POS tagger, parser etc.), but usage of such NLP packages are yet to touch the real needy. The need for encouraging people to learn Tamil lies truly in the application of information and communication techniques to teach basic grammar in school curriculum, which is now a problem not yet addressed. This paper will be an initiative which will try to help school children a lot by encouraging them to use ICT for learning Tamil grammar in a more interesting manner.

Methodology

The three of five major divisions of Tamil grammar say, நூற்றாண்டுக்கர், புனிதான்கர் and பார்வான்கர் are considered for developing the application prototype. Rule base is constructed for calculating பாக்கிக்கர். The occurrences of நூற்றாண்டுக்கர், புனிதான்கர், பார்வான்கர், பார்வான்கர், பார்வான்கர், பார்வான்கர், பார்வான்கர் and பார்வான்கர் are identified and rules are generated. In addition, the classification of poems ( சிங்கத், சுற்றுத், சுற்றுத், சுற்றுத், சுற்றுத்) based on their structure is done. Adding to that, rule base is created to find the presence of various grammatical items including நூற்றாண்டு, புனிதான்கர், பார்வான்கர், பார்வான்கர், பார்வான்கர் and பார்வான்கர்.

The existing morphological analyzers [2] address பார்வான்கர் in a decent manner. Utilizing the morphological analyzer, in this prototype, we shall analyze the words in the given input. For our application, we can get either TAB or Unicode input with the help of NHM writer. For common processing of both inputs, the input string is converted into internal codes based on a coding system(shown in Table 1) which just contains codes for புனிதான்கர் and பார்வான்கர் and special
characters 6, 7, 8, 9, 0, 1. We do all the processing with these internal codes and while displaying the output, the results are converted back into corresponding TAB/Unicode encoding.

For example, the input string செய்தி செய்தி it will be converted into the byte array which contains the internal codes 1 23 23 22 22 2.

Table 1. Internal Encoding

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>11</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>12</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>8</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>9</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>40</td>
</tr>
</tbody>
</table>

1. உருமைக் கைலை


t is the time that should be taken to pronounce a word. இயை will have 1 உருமைக் கைலை, இயை will have 2 உருமைக் கைலை, இயை and இயை will have ½ உருமைக் கைலை. To calculate உருமைக் கைலை, we should check the byte array we have got one by one. For இயை இயை and இயை இயை, we can assign உருமைக் கைலை directly. In case of இயை, இயை இயை and இயை இயை, we should check for the next byte in the array before assigning உருமைக் கைலை, because the next byte only will decide whether the character is a இயை or இயை or இயை. If the next byte is இயை இயை, the character is இயை இயை. If the next byte is இயை இயை, the character is இயை இயை. If the next byte is not an இயை, the character is a இயை.

2. உருமைக் கைலை reduction

Due to the presence of some special characters like இயை இயை, இயை, இயை and இயை, இயை and இயை, இயை and இயை, there will be change (reduction) in the உருமைக் கைலை of the character in which it is present. As a result, the total உருமைக் கைலை of the word will also change. For இயை இயை, இயை and இயை, the உருமைக் கைலை of the character will get reduced by ½. For இயை இயை and இயை, the உருமைக் கைலை of the character will get reduced by ¼. For இயை இயை and இயை, there will not be any உருமைக் கைலை change. The rules for identifying the presence of these special characters in a word are mainly obtained from [10]. The other books referred are [6],[7],[8],[9]. The
rules obtained are programmed in java and put into rule base. Applying these rules, the reduction can be calculated.

We can give direct Tamil prose input or we can give already typed Tamil prose text file as input. In 

reduction can be calculated. If we click button, the above output will be obtained. It will display 
the of each character in a word and reduction of total of the word due to the presence 
of special characters like ஆண்டவை, எண்ணைப்பிள்ளை...... The sample output obtained for 

is shown in Figure 1.

Figure 1. தமிழ் பாணிக்கை

முடிவுக்காக

எண் refers to each individual character in the given input. The character may be எண்முறை, எண்முறையைக், எண்முறை, எண்முறையைக் or எண்முறையைக். Individual character or a group of characters together constitute an எண். The two types of எண் are எண் முறை and எண் முறை. One or 
more எண் are grouped together to form a எண். If one எண் forms a எண், it is called எண் முறை(தொடு, மூன்று, எண், பத்து). If two எண் form a எண், it is called எண் முறை or எண் முறை(தொடு, பத்து....). If three 
எண் form a எண், it is called எண் முறை. Two types of எண் முறை are எண் முறை (தொடு, 
எண் முறை,...) and எண் முறை (தொடு, எண் முறை,...). Similarly four எண் also can form a 
எண்(எண் முறை). By checking the எண் of the two consecutive words, the எண் between can be identified. 
The seven types of எண் are சிரைக்கைகள், இல்லைக்கைகள், இல்லைக்கைகள், இல்லைக்கைகள், இல்லைக்கைகள், இல்லைக்கைகள், இல்லைக்கைகள். Two or more எண்
together form an அக்கின் (line of a poem). The types of அக்கின் are தின்நாளுடைய (2 ரிசுக்), திகரமுயைந்த (3 ரிசுக்), அவர்களுடைய (4 ரிசுக்), நூற்றாண்டுகளுடைய (5 ரிசுக்), மூன்று நூற்றாண்டுகளுடைய (more than 5 ரிசுக்).

To make the poem sound good and meaningful, the characters and seers are arranged with care. One such attempt is named டைகேன். The two main types are டைகேன் and டைகேன். டைகேன் means the first character of the two words which are compared are same. The two types of டைகேன் are ரிசு டைகேன் and அக்கின் டைகேன். In ரிசு டைகேன், we compare the ரிசு கின் in an அக்கின். There are 7 types of ரிசு டைகேன் based on which ரிசு கின் in an அக்கின் we are comparing. In அக்கின் டைகேன், we compare the first ரிசு of two lines (அக்கின்).

டைகேன் means the second character of the two words which are compared are same, provided the முதல் கின் of their first characters are also equal. The two types of டைகேன் are ரிசு டைகேன் and அக்கின் டைகேன். In ரிசு டைகேன், we compare the ரிசு கின் in an அக்கின். There are 7 types of ரிசு டைகேன் based on which ரிசு கின் in an அக்கின் we are comparing. In அக்கின் டைகேன், we compare the first ரிசு of two lines (அக்கின்).

Finally, we can find the தமிழ் கின் of the input poem with the number of different types of அக்கின் and ரிசு கின் they are having and some special conditions. If the poem belongs to மூன்று நூற்றாண்டு, மூன்று நூற்றாண்டு கின், நான் என்று, அருணாரோ கையில், we can identify it. We can give direct Tamil poem input or we can give already typed Tamil poem text file as input. In பயிற்சிக்குள் தொடக்கம் tab if we press பயிற்சி button, it will display the type of each character in a word like மூன்று நூற்றாண்டு, மூன்று நூற்றாண்டு கின், நான், தமிழ் தானோ கையில், தமிழ் தானோ மரணம், மரணம். If we press வளாக button, it will split up of மூன்று and மூன்று கின் in it. If we press மூன்று button, it will categorize மூன்று (மூன்று, அருவு, கண்டி, கண்டி) based on தமிழ் கின் they have. If we press வளாக, it will categorize வளாக into தமிழ், பிறமுத்த, அருவு, எதிருமுத்த, எதிருமுத்த. If we press வளாக or வளாக கின் button, it will show if any வளாக or வளாக கின் present in it. If we press பயிற்சி கின் button, it will show whether the poem belongs to பயிற்சி, பயிற்சி கின், கண்டி, கண்டி கின் based on ரிசு கின் and தமிழ் கின் they have. The sample output obtained for பயிற்சிக்குள் தொடக்கம் is shown in Figure 2.

![Figure 2](image-url)
2.3 தமிழ் நெறி விளக்கம்
Tamil Morphological analyzer-atcharam [3], [4], [5] is used to get the output for தமிழ் நெறி விளக்கம். We can give direct Tamil prose input or we can give already typed Tamil prose text file as input. Various grammatical patterns like noun, verb, adjective, adverb, participle, clitics, singular, plural, accusative case, locative case, associative case, sandhi, infinitive suffix, permissive suffix, postposition, present tense maker, past tense maker, future tense maker, etc., can be identified by using this. The sample output obtained for தமிழ் நெறி விளக்கம் is shown in Figure 3.

Figure 3. தமிழ் நெறி விளக்கம்

Conclusion
While utmost care has been taken to develop the grammar prototype in Unicode, it is no doubt that automatic identification of grammatical patterns is going to be a worthy application with respect to school education. தமிழ் நெறி விளக்கம் and காலத்தில் விளக்கம் deal with the semantics of the language. With decent Tamil dictionaries like agaradhi [1], handling தமிழ் நெறி விளக்கம் is not an issue. However, காலத்தில் விளக்கம் requires enough research thought to be thrown before things could be materialized. But, times are not too far that in future, using Tamil NLP tools, one should be able to analyse the similes present in the input Tamil text.
References

3. A.G. Menon, S. Saravanan, R. Loganathan, Dr. K. Soman, “Amrita morph analyzer and generator for Tamil: A rule based approach”, in the proceedings of Tamil Internet Conference 2009, Koeln, Germany
5. P. Anandan, Dr. Ranjani Parthasarathy & Dr. T.V. Geetha, “Morphological generator for Tamil”, in the proceedings of Tamil Internet Conference 2001, KualaLampur, Malaysia,
6. இராபேண்டுஸ், சார்பனாராயண, “நித்திரைக்குறுக்கு இரவர்பாகம்”, முன்னெச்சரியா அறிமுகம் மைசூர், 2001
8. ப. குருத்தாரா, க. இராஜா, “இருவர்பாகமும் முன்னெச்சரியா இரவர்பாகமும்”, முன்னெச்சரியா அறிமுகம் மைசூர், 1999
10. பார்வதி ந. பார்வதி, “முன்னெச்சரியா இரவர்பாகம்”, மைசூர் முன்னெச்சரியா, February 2006
4

கலம் பறி முறிய செய்யி
நூற்றண்டுகள்
கருதியும் நிலையிலே நூற்றாண்டுகள்
இறுதியும் பட்டமிலித்த

B.E., M.A, M. A

வருந்தக் கொரியில் வருந்தக் கொரியில்

இறுதியும் பட்டமிலித்த

உரையின் முக்கிய மூலக்கூற்றுலகள்

1. தமிழ்வணிகத்தின் காலமான இயற்கையும் இருந்து நூற்றாண்டுகள்

2. வருந்தக் கொரியில் வருந்தக் கொரியில்

3. இறுதியும் பட்டமிலித்த

4. கருதியும் நிலையிலே நூற்றாண்டுகள்

5. தமிழ்வணிகத்தின் காலமான இயற்கையும் இருந்து நூற்றாண்டுகள்

6. வருந்தக் கொரியில் வருந்தக் கொரியில்

7. இறுதியும் பட்டமிலித்த
முற்பொறிக்கற் கிளற்றுமுறையில், என்று காண்பது ஒன்றில் தமிழில் பதிவு செய்யும் முறுக்கு குறிக்கள் முறையில் குறிக்கபட்டுள்ளது.

அத்தியாயம் - பொருளான்கள்

தமிழ் ஓர் பொருளான்கள் மொழி (Agglutinative language). தமிழ் ஓர் மொழி, திருநெல்வேலி, வெள்ளதென், வரலாறு, உருமை தமிழில் பொருளான்கள் மொழியில் பார்வைப்படுத்தும் குறிப்பிட்டிய அங்குலமினையான மொழியில் பொருளான்கள் (Lexical meaning) மற்றும் இன்றும் பொருளான்கள் மொழியில் பொருளான்கள் (Grammatical meaning) என நம்பப்படுகின்றன. உருக்கு பொருளான்களும் வண்ணம் பொருளான்களும் குறிப்பிட்டிய வழியில். புதிய பொருளான்கள் கூறப்படும் வழியில், வழி, பொருளான்களும் பொருளான்களும் குறிப்பிட்டிய வழியில்.

நேரலை பொருளான்கள் நிறுவனவிசையில் (Natural languages) இருண்டு கொண்டாற் தமிழ்ப்பொருளான்கள். தங்க, முன்னிலை பொருளான்கள் (Affixes) வரும் அவசை தமிழ்ப்பொருளான்கள். அறிவுக்குரியதாக, பொருளான்களில் பொருளான்கள் குறிப்பிடுகின்றன. பொருளான்களில் பொருளான்களின் பொருளான்களில் பொருளான்கள் குறிப்பிடுகின்றன. அவற்றை தந்திருந்து பொருளான்கள் குறிப்பிட்டிய வழியில். “Look - Looked”

பின்னர் பொருளான்கள் கொண்டு பொருளான்கள் வழியில் பார்வைக் (Preposition), குறிப்பிட்டிய பொருளான்கள் (Auxiliary Verbs) கொண்டு பொருளான்கள் வழியில் (Modal), திருநெல்வேலின் வழியில் (Aspectual) எங்கும் குறிப்பிட்டியதாக. இது மொழியில் பொருளான்களில் பொருளான்களைக் குறிப்பிட்டிய வழியில் பொருளான்களைள் அரங்க வழியில் அப்பொருளான்கள் குறிப்பிட்டிய வழியில் பொருளான்களைப் பொருளான்களையும் செயலுக்கு கொண்டாற் பொருளான்களையும். பொருளான்களை தந்திருந்து பொருளான்களையும்.

ஆனால், பொருளான்கள் கொண்டு பொருளான்களில் பொருளான்களும், பொருளான்களில் பொருளான்களும் குறிப்பிட்டியதாக. அரங்கங்களில் பொருளான்கள் குறிப்பிட்டியதாக. பொருளான்கள் குறிப்பிட்டியதாக, அதாவது வழியில் பொருளான்கள் குறிப்பிட்டியதாக, அதாவது வழியில் பொருளான்கள் குறிப்பிட்டியதாக. இதனால் எந்த அறிவுக்குரிய பொருளான்களில் பொருளான்கள் குறிப்பிட்டியதாக. இதனால் எந்த அறிவுக்குரிய பொருளான்களில் பொருளான்கள் குறிப்பிட்டியதாக. இதனால் எந்த அறிவுக்குரிய பொருளான்களில் பொருளான்கள் குறிப்பிட்டியதாக.

நேரலை பொருளான்கள் பார்வையாகும் வழியில் பொருளான்களில் பொருளான்களில் பொருளான்களை பார்வையாக ஆதாரவும் அளிப்பதற்கு மூலமாக (Morphological Analysis) எடுத்துக்கொள்ளப்படுகின்றது.

மொழிப் பொருள் ஐதல் (Natural Language Processing - NLP)

நேரலை பொருளான்கள் பார்வையாகும் (மொழிப் பொருள் ஐதல்) ஆக்கமல்லியா பொருளான்கள் பொருளான்கள் பொருளான்கள் (அளவு குறிப்பிட்டியதாக), பொருளான்களில் பொருளான்கள் பொருளான்கள் (Understanding and Generating) எடுத்துக்கொள்ளப்படுகின்றது.

நேரலை பொருளான்கள் குறிப்பிட்டியதாக பொருளான்கள் ஆக்கமல்லியா பொருளான்கள் பொருளான்கள் பொருளான்களில் பொருளான்களில் பொருளான்களை ஆக்கமல்லியா எடுத்து கொள்ளை பொருளான்கள்.
Linguistics)


1. Root Word Database)
2. Grammatical Affix Database)
3. Morphotactics)
4. Morphophonemics)

Language Faculty) (Knowledge of Language) requests for any crawlers.

Syntactic analysis), Semantic analysis) requests for any crawlers.

Grammaratical Category) requests for any crawlers.

Sub-Category) requests for any crawlers.
(Auto spell-checker)

"புகழ்வார்" மன்னர் உள்ளிட்டு வாழ்ந்தவருள் பெருமளவு கொண்ட மன்னர். அவர் என்பதாயினும் ஏற்றுவழியான புகழ்வாரமாக இருக்கிறார். அவரத் மன்னராக இருக்கும் வரை அவர் தனது புகழ்ந்தவர் பெயரும் மேம்படுத்தி கொண்டார். புகழ்வார் மன்னர் உள்ளிட்டு வாழ்ந்தவரின் புகழையை ஏற்றுவழியான புகழ்ந்தவர் பெயரும் மேம்படுத்தி கொண்டார். 

அவரத் பெயர் ஆண்டவர் என்பதால் அவர் தனது புகழ்ந்தவர் பெயரும் மேம்படுத்தி கொண்டார். அவரத் பெயர் பெயரும் மேம்படுத்தி கொண்டார். 

"புகழ்வார்" மன்னர் உள்ளிட்டு வாழ்ந்தவர்களுள் பெருமளவு கொண்ட மன்னர். அவர் என்பதாயினும் ஏற்றுவழியான புகழ்ந்தவர் பெயரும் மேம்படுத்தி கொண்டார். புகழ்வார் மன்னர் உள்ளிட்டு வாழ்ந்தவரின் புகழையை ஏற்றுவழியான புகழ்ந்தவர் பெயரும் மேம்படுத்தி கொண்டார்.

272
கனியான் கதிர்பில்லறிகள் எக்பாக்கத்தில்
நுள் பக்த்தில்லறி பாகத்தில்

R. புதமன், MCA, (Ph.D.), M. பரதன் B.E. M.A

மாணவியல் - குடியுரியல், கல்வி எழுத்துத் துறை,

சக்கரி: குளியத்தை பலகை செய்யப்பட்டு தனி கதிர்பில்லறிகள் (Computer Program) கனியமிரு கதிர்பில்லறிகள் (Auto Sandhi Checker). இது எக்பாக்கத்தில்

முழு பக்த்தில்லறி (Morphological Analyser) பாகத்தில் பயன்படுத்தும் எக்பாக்கத்தில் பாராட்டிகள். 

எக்பாக்கத்தில் பாராட்டிகள் (Morphophonemic Changes)

துணையில் குளியத்தை கதிர்பில்லறிகள் எக்பாக்கத்தில் பாராட்டிகள் பரிமாறி செய்யப்பட்டது. எக்பாக்கத்தில் குளியத்தை பாராட்டிகள் பாராட்டிகள் எக்பாக்கத்தில் பாராட்டிகள் பாராட்டிகள். 

எக்பாக்கத்தில் பாராட்டிகள் எக்பாக்கத்தில் பாராட்டிகள் எக்பாக்கத்தில் பாராட்டிகள். 

எக்பாக்கத்தில் பாராட்டிகள் (Lexical Ambiguity)

சக்கரி: குளியத்தை பாராட்டிகள் பாராட்டிகள் எக்பாக்கத்தில்

எக்பாக்கத்தில் பாராட்டிகள் பாராட்டிகள் எக்பாக்கத்தில்

எக்பாக்கத்தில் பாராட்டிகள் (Lexical Ambiguity)
சுருக்கிக் கலன்கள் அறிவு (Multilevel knowledge):

மென்செய்யப்பட்ட கிளை நாட்டுவகையில் தரும் ரூபக் கிளைகள் மொழிபெயர்வு கணினிச் செயற்பாடுகள், மொழி செயற்பாடு, குறிப்பிட்டிய ரூபங்கள் பிரிக்கப்பட்ட மொழி பயன்பாட்டை அறிவு பொருத்திய சிரீசெய்யும் முறையாகும்.

கிளை + கிளை = கிளை

சார்பில் இயக்கத்தில் இவ்வுரையார் (Front Vowel) உடுறுப்பு அல்லது புறக்கை வாழ்த்து கிளைப்பார் கம் மற்றும் (Glide) தெளிவுநிலையாகும். இரு கிளைகளும் (Phonology) அடர்ச்சியாக முறையிடப்படுகின்றன.

அகராதி நாட்டின் சில்லவியல் பொறியியல்

மென்செய்யக்கள் அறிவு, மொழி நாட்டு சார்ந்த சில்லவியல் பொறியியல் மொழித் தவறாகவும் கணினிச் செயற்பாடுகளை அறிவு பொருத்திய சிரீசெய்யும் முறையாகும் சில்லவியல் மொழி முறையிடும் சில்லவியலாகும். இரு சில்லவியலும் மொழி செயற்பாடு (Syntax) கிளையிடப்படுகின்றன.

பாதுகாப்பு கிளையிட்டியதில் பாதுகாப்பு முறையிடப்படும் பாதுகாப்பு நாட்டின் முறையிட்டியதில் பாதுகாப்பு முறையிடும் சில்லவியலாகும். இது இன்றுவையானமலை (Semantics) அடர்ச்சியாக முறையிடப்பட்ட கிளை முறையிடுப்படும்.

மொழியைப் பார்க்கும்

சுருக்கிக் கலன்கள் அறிவு (Grammatical knowledge) மொழியை முழுமையாக அறிவு (Pragmatic knowledge) முழுமையாக அறிந்து கொள்ள முடிகிறது. அதன் படிகள் ரூபீ எடுக்கப்படும். அவற்றை தொடர்ந்து கூற்றுத்துறை என்று பொருத்திய சிரீசெய்யும் முறையிடும் சில்லவியலாகும். அத்துடன் சில்லவியல் பொறியியல் அறிவு பொருத்திய சிரீசெய்யும் முறையிடும் சில்லவியலாகும். ஆனால், வெளியில் இருவர் அனைத்தும் இரும்புமற்றும் ஒன்றியக் கூற்றுத்துறையும் இன்றுவையானமலையும் அறிந்து கொள்ளப்படும். இரும்புமற்றும் ஒன்றியக் கூற்றுத்துறையும் குறிப்பிட்டியது கணினிச் செயற்பாடாக அறிந்து கொள்ளப்படும் குறிப்பிட்டியது கணினிச் செயற்பாடாக அறிந்து கொள்ளப்படும்.

குறிப்புக்கலா குறிப்பிட்டியதிலை (Auto Sandhi Checker)

செயற்பாடாக விளக்கும் சில்லவியல் பொறியியல் மொழிபெயர்வு மற்றும் கணினிச் செயற்பாடு மொழியானது எடுக்கப்படும் மொழி செயற்பாடுகள் மொழி செயற்பாடுகள் (Natural Language Processing) தொடர்ந்து இந்துக்கால மொழி செயற்பாடு (Data Structure), கணினிச் செயற்பாடு (Algorithm) மற்றும் இந்துக்கால கணினிச் செயற்பாடு (Computational Linguistics). இருந்துசுருக்கிக் கலன்கள் அறிவு பொருத்திய சிரீசெய்யும் முறையிடும் சில்லவியலாக முறையிடும் சில்லவியலாக முறையிடும்.
Language Technology.

( language technology).
விளக்கில் நம்பவர்களுக்கு முன்னர், தனியார் முக்கியத்துவம் கொண்டிருந்த விளக்கில் நேராகவும் இன்றைய தமிழ்நாட்டு குழுவுடன் நடவடிக்கை நடைப்படுத்தும் ஒப்பந்தம் எடுக்கவுகே என்று குறிப்பிட்டுள்ளன். நேராகவும் இணையாடப்பட்ட (Subcategory) பச்சையுடன் இணையத்தைக் கொண்டு விளக்கில் நேராகவும் இடைய்ச்சியில் அமைந்து விளக்கில் தமிழ்நாட்டு முதலியல் பெருமையை எழுதியுள்ளன்.
Spell Checker for Tamil using
Finite State Automata

Anitha.S Pillai
Hindustan University
anitha2293@gmail.com

Abstract
The problem of detecting and correcting misspelled words in text has received great attention due to its importance in several applications like text editing systems, optical character recognition systems, morphological analysis and tagging. Other applications like machine translation and information extraction, operate on text. There are possibilities that there may be errors in these text. The ability to automatically detect and correct spelling error should be of great help to those applications. The problem of detecting and correcting misspelled words in text is usually solved by checking whether a word already exists in the dictionary or not. If not, we try to extract words from the dictionary that are most similar to the word in question. This will not work for all languages where a large number of words can be derived from the root word and it is not feasible to store all the words in the dictionary. Hence a Morphological analyzer also plays an important role in these languages.

In this paper an approach for automatic correction of spelling mistakes in Tamil document using the Finite State Automata (FSA) is proposed.

Introduction to spell checking

This paper discusses the stages involved in the development of a Tamil Spell Checker. Since in Tamil, a large number of words can be derived from a root word, a purely dictionary based approach for Spell Checking is not practical. Hence a ‘Rule cum Dictionary’ based approach is followed. The lexicon (dictionary) is stored in the form of finite state automata.

The different modules in the Spell Checker Engine viz. Morphological Analyzer, Error detection and suggestion generation module are also explained.

A Spell checker is a tool that will check the spelling of words in a document, validate them and in case the spell checker finds error, list out the correct spelling in the form of suggestions. Any word processor should have a spellchecker associated with it as the user may commit mistakes during typing. According to Damerau (1964), 80% of all misspelled words in a sample of human keypunched text were caused by single-error misspellings. It could be either due to insertion, deletion, substitution or transposition. Kukich (1992) breaks down human typing errors into two classes. Typographic errors are generally related to keyboard. Cognitive errors are caused by writers who don’t know how to spell the word.

The spellchecker detects the mistakes and prompts the user with a set of suggestions, which will aid the correction of the misspelled word. The approach followed in the design of an Indian language Spell Checker has to be different from those for Roman scripts, because in Roman alphabet each letter is
complete by itself and represents a sound where as in Indian languages the characters are more syllabic in nature and most of the consonants will have a vowel sound added to it. Unlike other languages, development of Spellcheckers for Indian languages, especially Dravidian languages like Malayalam, Tamil, Kannada and Telugu is a bit complicated.

Here the suffixes, postpositions and case endings agglutinate with the verbs, nouns, adverbs or pronouns. Also one or more suffixes can combine with the base word. In other Indian languages, for eg. Hindi, case ending will not agglutinate with the base form. Also combinations like verb-verb, noun-noun, verb-noun etc. are not permitted in many other Indian languages.

Hence Morphological analysis of the input word is a must for Tamil Spell checker. Tamil is a morphologically rich language in which most of the morphemes coordinate with the root words in the form of suffixes. Person, gender and number markings combine with the root words. For error detection and suggestion an efficient morphological analyzer is required.

Spell checker application presents valid suggestions to the user based on each mistake they encounter in the user’s document. The user either selects from the suggestions or choose to ignore the suggestions and accept the current word as correct.

**Finite State Automaton**

Automaton is represented as a directed graph: a finite set of vertices (nodes), together with a set of directed links between pairs of vertices called arcs. Each node corresponds to a state. States are represented as circles with name tags in them. Arcs are represented by arrows going from one state to another state. The final states are represented by double circles.

The machine starts at the initial state, runs through a sequence of states by computing a morpheme in each transition, and ends in the final state. The path moves from the initial point on the left to the final point on the right, proceeding in the direction of arrows. Once the arrow moves one step, there is no backward movement (Of course, recursion of an item can be shown by using closed loops). The resulting Finite State Automata (FSA) is deterministic in the sense that given an input symbol and a current state, a unique next state is determined. A deterministic finite state automaton (DFA) is perhaps the simplest type of machine that is still interesting to study.

However, it is an enormously useful practical abstraction because DFAs still retain sufficient flexibility to perform interesting tasks, yet the hardware requirements for building them are relatively minimal. DFAs are widely used in text editors for pattern matching, in compilers for lexical analysis, in web browsers for html parsing, and in operating systems for graphical user interfaces. They also serve as the control unit in many physical systems including: vending machines, elevators, automatic traffic signals, and computer microprocessors. They also play a key role in natural language processing and machine learning.
Here S0 represents initial state, S1, S2, S3 and S4 represents intermediate state and S5 represents final state. The automaton has five states which are represented by nodes in the graph. State S0 is the start state which we represent by the incoming arrow. State S5 is the final state or accepting state which we represent by the double circle.

**Representation of words using Finite State Automaton**

Tamil lexicon is stored as a Finite state Automaton. There is an initial state and final state for each word. The set of valid characters that help to move from the initial state to final state is specified for all the Tamil root words.

Eg: initially the system is in state 1(initial state) . when a character ‘a’ is input it goes to state2. From state2 on receiving the input ‘m’ it goes to state 3. The list of characters that characters that help ‘a’ to move from the initial state to final state is given.

![Diagram of FSA for Tamil words](image)

**Representation of words amma and appa using FSA**

**Checking of Words**

It starts at the initial state (S0), checks the next morpheme of the input. If it matches the symbol on an arc leaving the current state, then it crosses that arc, and moves to the next state, and thus, advances one symbol in the input. Such a process gets iterated until the machine reaches the final state, successfully recognizing all the morphemes in the input string. But if the machine gets some input that does not match an arc, then it gets stuck there and never gets to the final state. This is considered as the FSA / machine rejecting or failing to accept an input.

![Diagram of FSA for Tamil words](image)

*The word ‘amma’ is accepted by the finite state automata*
Word not recognized by Finite state automaton

If a set of input characters do not help to move from the initial state to final state then the automata will not be recognized for the given word. This could happen due to 2 reasons:

1) The word is an inflected word i.e. suffixes are attached to the root word.
2) Spelling mistake in the word

Inflected Word: Root words to which suffixes are attached. These words will not be available in the lexicon. So Morphological analyzer is required to segment the word to root and suffix.

Morphological analysis: There will be certain suffixes along with the root word. This word can be called as an inflected word. The respective inflected word is stripped to its corresponding root word and suffix using backtracking algorithm such that the whole word can be written as root word+suffix. But in some cases when the suffix is stripped off the input string the root word will not be present in the root database. In such cases, we use orthographic rules which are solely used for certain words which are meant to be modified based on these rules.

For example:- pasangal. In this word the ‘gal’ denotes the plural form. After the plural suffix is removed the rest of the word is checked with the root database. If the word is not present in the root database then orthographic rules are applied such that the input string is identified as correct.

Sandhi checker

The Sandhi checker deals with the orthographic changes that occur in a root word when suffixes are added to it. When the input string has suffixes added to it and they are stripped off their respective suffixes, in some cases the root word may not be present in the root database. For example:- consider the word ‘pazhangal’. Once the suffix ‘gal’ is stripped off the rest of the word shows ‘pazhang’ which is not present in the root database. Here the sandhi checker helps in finding out whether the word is right or not. And finally the word is displayed as a correct word.

Suggestions

The minimum edit distance (or just edit distance) between two strings is the least number of elementary editing operations – insertions, deletions, and substitutions – that are needed in order to transform the first string into the second one.

Screen Layout of Spell Checker
For example, the edit distance between the strings summer and summary is 2: we substitute the e for a, and then we insert y at the end. No other shorter edit sequence exists. Therefore, the edit distance is 2. These metric measures how similar to each other two strings are. Edit distance is also symmetric: if $ed$ is a binary function mapping two strings to their edit distance, then, for every two strings $s$ and $t$, $ed(s, t)$ is always equal to $ed(t, s)$.

The whole program is developed using visual basic.net 2008 and MS-Access database. When the user types a word and if the word cannot be represented as a FSA an error is shown. This means that this particular word is not a correct one and so the FSA fails to move from the initial state to the final state. In such cases using minimum edit distance possible suggestions are displayed to the user to choose from. When the user types the word and press the space bar to type the next word, if there is an error then the word typed changes to blue. When the user right-click the mouse, suggestions are provided to correct the misspelled word. The suggestions are based on minimum edit distance.

Conclusions:

Users can type text in Tamil and whenever the application encounters a misspelled word it highlights the error by changing the font colour of the wrong word to yellow. When we right click on this wrong word a list of possible suggestions are displayed.

References:

2. Institute of Linguistics and Culture.
4. A Smart Spell checker system http://www.coe.neu.edu/
7. Yo-Sub Hana, Derick Woodb Obtaining shorter regular expressions from finite-state automata Theoretical Computer Science 2006
1.0  

Albert Einstein [Variable] ATM [ATM] E-Dictionary [E-Dictionary]; sbaskarantj@yahoo.com

Thesaurus [Thesaurus] Internet [Internet] Redefinition [Redefinition]

Albert Einstein [Albert Einstein] Alvin Toffler [Alvin Toffler]

Language Tools for Tamil: Needs for Computerization and Necessities for Standardization
2.0 கல்லச்செயல்பாடுகள் - வள கூறுகள்

அமைய்கள் பல்லவியல் துறவலால் கல்லச்செயல்பாடுகள் குறுக்குநிலையாக வருவதாய் இருக்கும். பல்லவியல் துறவல் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழிட்டு விளக்கத்தில், தொழிட்டு விளக்கத்தில் பல்லவியல் துறவல் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறவல்கள் பின்னர் பல்யாண் தொழில்நுட்பத் துறவல்கள் வருவது விளக்கம் பெற்றுள்ளது. தொழில்நுட்பத் துறав
4.0 **Corpus**

A **Corpus** is a collection of texts used for analysis. In this context, a **Corpus** refers to a set of parallel corpora, which are essentially bilingual corpora with aligned sentences. This allows for the comparison and analysis of similar or identical content in two different languages.

- **Parallel Corpora**: These are bilingual corpora that consist of two or more parallel texts. They are used to study the relationship between different languages and to identify patterns of usage.

4.1 **Standard Deviation**

The **Standard Deviation** is a measure of the amount of variation or dispersion of a set of values. It is calculated by taking the square root of the variance. The formula for standard deviation is:

\[ \sigma = \sqrt{\frac{1}{N-1} \sum (x_i - \mu)^2} \]

where \( \sigma \) is the standard deviation, \( N \) is the number of observations, \( x_i \) are the individual observations, and \( \mu \) is the mean of the observations.
1. **Zipf's Law**
2. **TTR** (Type-to-Token Ratio)

### Table

<table>
<thead>
<tr>
<th>Frequency(f)</th>
<th>[Rank (r)]</th>
<th>( \text{Frequency} \times \text{Rank} )</th>
</tr>
</thead>
</table>

**4.1.1 Zipf's Law**

The following formula is used to determine the probability distribution of word frequencies: \[ f(x) = \frac{1}{x^\theta} \]

Where: \( f(x) \) is the frequency of word \( x \), \( x \) is the rank of the word, and \( \theta \) is a constant.

**4.1.2 TTR**

**TTR** is the Type-to-Token Ratio, which is calculated as the number of distinct terms divided by the total number of terms.

**TTR** values range from 0 to 1, with higher values indicating a more diverse vocabulary.

**TTR** is calculated as:

\[ \text{TTR} = \frac{\text{Distinct Words}}{\text{Total Words}} \]

**Distinct Words** represents the number of unique words in the text, while **Total Words** is the total number of words in the text.

**TTR** values can be used to assess the diversity of a text, with higher values indicating a more diverse vocabulary.
4.1.3 பொருளாந்து பரிசுபாதம்

பாற்ற பொருளாந்து துணைக்குறிப்பிட்டு [Theory] பாதம் பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து 

4.1.4 பராமரிப்பு பொருளாந்து பகிர்வக்

ஜோட்ட பொருளாந்து குறிப்பிட்டு [Theory] பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து எளியாக பொருளாந்து 

286
5.0 G2C [Transliteration]

G2C represents Government to Citizen, a concept involving the interaction between government and citizens. This interaction can be facilitated by various means, including digital platforms. Transliteration is the process of converting text from one script to another, ensuring accessibility for non-native speakers. Google's translation tool is an example of this.

6.0 Bibliography

1. Baskaran S., Content Based E-Mail Classification System by Applying Conceptual Maps, 2009, IEEE.
2. Examples.
3. Examples.
Automated Processing of Census Forms in Tamil

Shashi Kiran¹, Rituraj¹, Suresh Sundaram¹, Swapnil Belhe², AG Ramakrishnan¹
MILE Lab, Dept of Electrical Engineering, Indian Institute of Science, Bangalore 560 012.
Center for Development of Advanced Computing, GIST Group, Pune
shashi.reach@gmail.com, sureshsundaram1980@gmail.com, swapnilb@cdac.in, ramkiag@ee.iisc.ernet.in

1. Abstract
This paper describes automatic form filling system for collecting census data based on online handwritten character recognition for Tamil. The aim is to facilitate easy digitization of Indian language ink-data gathered from field. The application interface is designed in such a way that the same application can be adopted for other Indian languages. It also describes the common interface framework required to facilitate this multiple language recognition engines. For Tamil or any other Indian scripts; inputting isolated symbols is not practical, hence the application uses non-isolated character recognition. The application incorporates simplified method of form design, layout analysis, engine error correction; engine level limited vocabulary based post-processor, validations etc. The performance evaluation of this application is also carried out against the traditional methods and promising results are obtained. The same system can be easily adopted for other types of forms required to be filled in Tamil e.g forms used by Government institutions, Banks etc.

2. Introduction
India is one of the very few countries in the world, which has the proud history of holding census every ten years uninterruptedly since 1872. The census provides information on size, distribution and socio-economic, demographic and other characteristics of the country’s population. The data collected through the census are used for administration, planning and policy making as well as management and evaluation of various programs by the government, NGOs, researchers, commercial and private enterprises, etc. [1]

Government agents visit each and every household in the country and collect complete data about the people and the condition of their houses. These agents gather this data by filling printed forms. The filled forms are then used to manually enter data in computers. This procedure takes long time and is prone to human errors while re-entering data manually. The data collection and the data storage are the two basic stages of this activity.

This census form processing system requires the members to carry the digital pads to the site, collect the information on this device, once the data collection is done plug in this digital pad to the computer and this application will generate the database.
Gathering information by using printed forms is predominantly used in governmental, educational, banking domain. Even on-field surveys are carried out using local languages. Traditionally, the major surveys like census are conducted on field by pen & paper wherein a printed form is filled by a surveyor. The surveyor collects information on the printed forms; and when the data collection is completed, forms are first scanned and then send for the verification and data entry. This method works well for small surveys requiring limited information. But when the survey requiring detailed information is to be carried out across cities, states and country; the traditional method becomes time consuming. The large amount of time is spent on scanning, verification and data entry.

The goal of this automated census form filling system is to provide simplified form filling process which works with very low cost digitizers and considers major Indian languages. Reduction in cost of digitization is also considered during design of this system.

In this system, the data collection process remains the same except an offline digitizing tablet is used for writing forms. The paper is kept on the tablet and the user writes on the paper by using special pen. The currently available offline digitizing tablets provide the advantage of retaining hard-copy of the filled forms for future use. These tablets with no display capture the ink data as X-Y co-ordinates and pen pressure. This collected ink data in X-Y co-ordinate format is then submitted to the form processing application which converts it to computer editable text.

3. Data Collection

The data collection phase is most crucial phase of the form processing application. The procedure for collecting data using digital tablet is similar to what currently being followed by data collectors. Instead of only-paper based forms the battery backed digital tablet is used along with digital ink based pen as shown below.

The data collected on such digital device can then be downloaded on the Computer for further processing. This is where Tamil online handwritten engine come into effect.

The Surveyor who is collecting the data is required to follow the manifest for collecting data which is similar to traditional data collection. We have collected data from ten writers of Tamil for testing by using G-Note 7000 digitizing tablet working at 160 points/sec.

4. Architecture of System

This system is primarily divided into four modules for simplicity during development, the modules are,

- User Interface
- Input Data Extraction module
- Recognition Engine Interface module
- Database Generation module
The application architecture is kept broad to easily plug-in different components. The application is currently designed to support two different types of digitizers namely Genius G-Note and iBall TakeNote but can easily be adopted to other devices. The Genius device produces the files with TOP extension while iBall produces files with DHW extension. As explained in the previous section the raw ink data (x, y co-ordinates) coming from the device is given to the application as input. Figure 1 explains the broad architecture of the application.

![Figure 1: Architecture of Tamil Form Processing Application](image)

The input data is displayed in full page format for verification. If the data contains any abnormal and misaligned ink strokes, they are removed by the operator. The verified data is then fed to the layout analysis module. This module aligns the input data to the pre-defined template using positional analysis. This way, the answers by the client are mapped to corresponding questions. Once the correct alignment is achieved the text/numeral fields are separated from ticks. The text and numerals are given to Indic engines for recognition. This is achieved by using common interface framework used to link the Tamil handwritten recognition engine to the application. The output of the engine contains the recognized text with confidence scores. The confidence score is used to provide the operator with choice to select the correct outcome. If the score is high enough then no selection choice is provided.

The operator is also provided with easy Tamil text editing functionality like virtual keyboard.

5. Common Interface Framework

Since this form processing system is designed to accommodate all Indian languages along with Tamil, there are many recognition engines to be interfaced to the system; hence there is a great need to have a standard communication protocol between different handwritten recognition engines and census application. A common Interface framework provides this vital link between the recognition engines and the application. This framework provides the flexibility to add/remove recognition engines. Also an entire new application can be built around this framework without the need of knowing the intricacies of online character recognition technologies. So this framework provides the scalability of adding new engines to the existing application. Simply put, it act as messenger between the application and the engine and thus helps to change any layer on the fly (i.e. we can change the application without disturbing the rest of setup or vice versa).
6. Pre-processing

6.1 Automated Segmentation

The segmentation of the validated data reduces the overhead of the Tamil recognition engine and helps in fast and accurate recognition of the handwritten text. All strokes in a page are given to segmentation module. In this module, all the strokes are arranged as per their position in a page irrespective of their order of occurrence. This module segments the lines based on horizontal projection of the strokes. Once the lines are segmented, the gaps between histogram (vertical projection) of each stroke on the line are clustered into within-word-gaps (WWG) and between-word-gaps (BWG) [2]. The words are separated by WWG.

6.2 Form Cleaning

The digital ink data collected by the surveyor often has noisy, unclean ink. Sometimes, the surveyor himself writes some notes, annotations, comments on the pages inside or outside the page boundaries. There also could be scrubbing which is large enough, crossing the full page or questions by digital ink. There could be overwriting, page misalignment or completely missing the pages of the form. This kind of data is required to be cleaned before proceeding for recognition. Otherwise the recognition accuracies could be very poor. Since it is difficult to clean all such noise automatically, the form cleaning is done semi-automatically. The cleaned, verified and segmented data is passed for recognition by the language specific engines.

7 Additional Features

7.1 Tick detection

The forms used for census survey contains various data fields like text, numerals, check boxes etc. The surveyor is supposed to tick into the boxes wherever necessary. First, the check boxes needs to be separated from other fields. This is achieved by positional analysis of the form. Once the relative position of the ticks is identified, it is required to separate actual tick marks from the scrubbing and other unwanted strokes. The tick detection is used for detecting the check mark on the form; this includes fields like radio buttons or check boxes. There are various ways of recognizing ticks in this implementation we have used Dynamic time warping (DTW). It is an elastic matching algorithm for matching the similarity between two given sequences. The similarity matching is based on the distance measure. The sequence with minimal distance is considered for optimal match. Since the distance based sequence matching is not feasible for every point in sequences we need to put two conditions i.e. boundary and continuity condition. The continuity condition decides how much the matching is allowed to differ from linear matching [3, 4]. The Boundary condition states that first and last points of sequences will be matched with each other. The continuity condition decides the measure of elasticity given by the formula:

$$\frac{N_2}{N_1}i - cN_2 \leq j \leq \frac{N_2}{N_1}i + cN_2$$
Where $c$ is the continuity constant, $N_1$ and $N_2$ are the number of points in first and second curve respectively. The points $i$ and $j$ of the first and second curve respectively can be matched only if above condition is satisfied. For $c=0$ the resulting match is same as linear matching.

![Figure 2 DTW matching of two ticks](image)

DTW is trained on data collected from 20 clients, for ten different styles of writing ticks. Certain conditions are applied to the data before passing the data to DTW such as checking for number of strokes.

### 7.2 Error Correction Measures

The recognition accuracy of the underlying Tamil handwritten recognizer is not 100% hence it becomes responsibility of the application developer to highlight those inaccuracies to the operator in a mild way. The ideal application for such kind of systems would be one which not only gives the operator the helping hand in case of a recognition error but also trains the engine over time by understanding corrections made by operator. It also allows editing the input file with stroke addition/deletion functionality and saving the same.

Significant percentages of operators are bound to get irritated in case of repeated errors on part of recognition (governed by recognition engines). An effective user interface in this case would be able to minimize the edits required by use of automatic focusing, multimodal inputting, easy suggestion list etc.

The application is built with stroke correction facility. If the ink data contains spelling mistakes or ambiguities then the operator can re-write the strokes and save it for future use. The operator is also given the facility to type using virtual keyboards. The complete application is localized for Tamil.

Since each field is separated by the values it may have, like text, numerals, numerals with symbols etc. The application performs the broad validity check on each output returned by the engine. This helps in narrowing down the errors for each field.

Based on the probability measure returned by the Tamil recognizer engine, the application takes the decision on whether the recognized text is suitable for displaying. If not, the field is marked with red color pointing to the operator about possible error that may require relook.

### 7.3 Standardized output format

The operator is given the flexibility to save the recognized data in Unicode based Comma Separated Variable (CSV) files or Extended Mark-up (XML) files. These files can be easily imported by any of the database engines like Oracle, SQL etc.
7.4 Wordlist for Post-Processing

The common interface framework includes the interface to send the domain specific Tamil wordlists to the handwritten recognizer. The application developer changes the wordlists as per different requirements of the applications and sends it to engines so that language models can be applied in order to improve the performance. The structure of the wordlist is defined by the framework.

8. Results and Discussions

We compared the performance of this automated census form filling application with the traditional ways of conducting census surveys. The performance of the application is highly dependent on the recognition accuracy of the Tamil Online handwritten recognizer. Application uses some domain specific knowhow of census form data to improve on the base recognizer performance. In the traditional surveys, the filled forms are scanned and then manual data entry is done. In this experiment the census form used for evaluation was very comprehensive spanning over 10 pages. Following table shows the composition of the form used for the survey.

<table>
<thead>
<tr>
<th>Field Type</th>
<th>Number of fields per form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text (with wordlist)</td>
<td>8</td>
</tr>
<tr>
<td>Text (without wordlist)</td>
<td>17</td>
</tr>
<tr>
<td>Numerals</td>
<td>22</td>
</tr>
<tr>
<td>Check boxes (multiple ticks)</td>
<td>10</td>
</tr>
<tr>
<td>Radio boxes (single ticks)</td>
<td>27</td>
</tr>
</tbody>
</table>

The text fields like clients education background, spoken language, city, district, state etc. were backed with limited dictionary while fields like name, last name, address etc. were without dictionary.

For evaluating performance of traditional census data entry, the forms were given to 3 data entry operators who were regulars in Tamil typing. Total data was collected from 10 writers; each form contained 10 pages. The same forms are processed through this application. Following table shows the comparative results of traditional manual data entry and online character recognition based application.

<table>
<thead>
<tr>
<th></th>
<th>Average Time (Time/Form) (10 pages per form)</th>
<th>Average Error in output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Data Entry</td>
<td>89.09 sec.</td>
<td>2.68 %</td>
</tr>
<tr>
<td>Automated Form processing + Verification</td>
<td>30.37 sec.</td>
<td>13.88 %</td>
</tr>
</tbody>
</table>
As seen from the above table, the manual data entry took more time but numbers of errors were less. Note that the page scanning time required for manual data entry is not considered in the above table.

9. Conclusion & Future Work

In our experiment, the online handwritten character recognition based form processing for Tamil clearly showed a promising area for further research especially where collecting huge quantities of data from field and converting it into editable text is concerned. In this application we retain the paper based form filling & editing thus allowing more natural text inputting and reduce cost incurred on display based inputting devices. In future, we would like to study the impact of the Indian language word models on user acceptance of online handwriting recognition.

10. Acknowledgment

The authors would like to thank Consortium for “Online Handwritten Character Recognition” and Technology Development for Indian Languages (TDIL), Department of Information Technology (DIT), Government of India for funding this consortium project. The authors would also like to thank all the consortium chief investigators and members consisting of IISc-Bangalore, ISI-Kolkata, IIT-Madras, IIIT-Hyderabad, CDAC-Pune for their valuable inputs.

References
1. Ashish Krishna, Girish Prabhu, Kalika Bali, Sriganesh Madhvanath, “Indic scripts based online form filling - A usability exploration”, 11th International Conference on Human-Computer Interaction (HCI), Las Vegas, 2005
Pattern based English-Tamil Machine Translation

S. Saravanan  
sarwanster@gmail.com

Dr. A.G. Menon  
menon.govindankutty@gmail.com

Dr. K. Soman  
kp_soman@amrita.edu

Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore

Introduction

The native languages all over the world are growing rapidly along with the growth of technology, in general, and information technology, in particular. On the one hand the world experiences a growth in the native language and on the other hand precious and nascent information come through foreign languages. Literacy in the mother tongue is no longer enough to follow the information supplied by the other languages. Because of this ever increasing gap and the speed with which information are supplied, it is necessary to bridge this gap with the help of modern technologies as early as possible. It is in this context that we are working on a Machine Translation (MT) system.

Even though, there are several approaches to develop a MT system, LTAG-based MT and SMT are very prominent. SMT is far away from tasting the success in case of agglutinative languages like Tamil. The only available MT system for English-Tamil is LTAG-based developed by AMRITA in collaboration with CDAC, funded by DIT. Transfer rules in LTAG-based MT are a pair of trees. Writing a new transfer rule is not very easy.

This paper proposes the use of pattern-based reordering rules in MT. Unlike the translation patterns in pattern-based CFG for MT (Koichi Takeda), the pattern based reordering rules are not lexicalized and the features and agreements are not handled in the rule. The source rules of these reordering rules are not used for parsing the source sentence. These rules are used to reorder the source phrasal structure to get the target phrasal structure. The English words are translated and lexicalized separately and this process is called lexicalization. The features and agreements are collected from the parse tree and used for synthesizing the words.

Parsing

One of the first steps is the identification of the structure with a parsing algorithm. Different parsing algorithms are available for the syntactic analysis of the source sentences. We are using the Stanford PCFG parser for the analysis of the source sentences in English. The sample parse output of the English sentence “Ram gave him a book” is shown in the figure-2.
Pattern based Reordering

A pattern is a pair of CFG rules (Takeda, Koichi: 1996). These rules give the equivalent structures in the source and target languages. On the basis of the patterns the reordering rules are formulated to facilitate the machine translation. They reflect the translation patterns of the source and target languages. For example, the following reordering rule is based on an English- Tamil pattern: \[ VP (VBD NP NP) \rightarrow VP (NP NP VBD) \] 1:2 2:3 3:1

‘VP (VBD NP NP)’ is the English CFG rule (‘VP’ is the root node and ‘VBD’, ‘NP’, ‘NP’ are the children nodes in the tree representation) called a source rule and ‘VP (NP NP VBD)’ is the Tamil CFG rule called a target rule. “1:2 2:3 3:1” is the transfer link. The tree representation of the above rule is shown in the figure-1 below.

The transfer link contains the order of the children nodes of the target rule. “1:2 2:3 3:1” says first child of the target rule is from second child of the source rule; second child of the target rule is from third child of the source rule, so on and so forth.

The parse tree of the source language is checked against the source rules. If any match is found in the parse tree, then the source rule is replaced with the corresponding target rule. For example, the parse tree of the English sentence, “Ram gave him a book” is (S (NP (NNP Ram)) (VP (VBD gave) (NP (PRP him)) (NP (DT a) (NN book)))). This English phrasal structure is checked with the available reordering rules for finding a match. The pattern in the source language such as ‘VP (VBD NP NP)’ is present in the English phrasal structure and it is eligible for undergoing the reordering rule. The pattern “VP (VBD NP NP)” of the source language is thus replaced with its counterpart “VP (NP NP VBD)” in the target language.

![Figure-1: Reordering Rule](image1)

![Figure-2: Transformation of Parse Tree to Reordered Tree](image2)
Tamil shows a very high degree of flexibility in ordering the words within a sentence. The position of the words can be easily transposed without much change in the meaning. For example, “Ram gave him a book” can be reordered in multiple ways in Tamil, and the most common ways are: Ram him a book gave, Ram a book him gave, Him a book Ram gave, etc.. The predicate verb takes mostly the last position. In our system, the reordering rules are strictly one to one map. Every source rule is mapped to one target rule. Based on the most common usage, the target rule is formulated. The Tamil clausal structure is more rigid and shows little flexibility. For example, “Ram, who is smart, gave him a book.” is reordered as “(smart Ram) (him) (a book) (gave)” Here the adjectival clause ‘who is smart’ has to be positioned before the noun ‘Ram’ in Tamil.

Reordering Algorithm

Let T be the parse tree with N number of nodes that we process for reordering; R be the number of reordering rules; S be the source rule tree; G be the target rule tree; Sc (n) be the sub tree of the node, n with the depth one. For example, the Sc (VP) in the parser tree is VP (VBD NP NP)

Re-order (T):

Visit node n
If n equals root (S)
For each St of reordering rule R
  If (Sc (n) equals S)
    Replace Sc (n) with target tree, G
  For each child c of n
    Re-order (sub tree(c))

The Re-order algorithm with the N number of nodes in the parse tree, T and R number of reordering rules has the complexity of O (N*R).

Lexicalization

The words in the terminal node of the reordered tree are lemmatized and the equivalent of that lemma in target language is replaced in the terminal node of the reordered tree. For example, lemma of the source word ‘gave’ is ‘give’ and its equivalent in target language is ‘koTu’. The target equivalents are found in the root word lexicon which contains the root forms of the source words and target words along with the POS category.
The major challenge in the lexicalization process is semantics. There would be multiple senses for the source word. In that case, the current system has the provision for multiple outputs. In the case of the functional words and auxiliary words, there are no direct equivalents in Tamil. For example, “The auditorium is decorated for the college day celebrations”, the words ‘for’ and ‘the’ do not have equivalent words in Tamil. ‘for’ comes as a post position (kkAka) in Tamil and is synthesized along with the Noun ‘college-day’. The auxiliary word ‘is’ has no direct equivalent in Tamil. The words ‘college day’ consists of two entities in source and the equivalents for these two words separately are ‘kallUri nAL’ which is semantically wrong. The compound nouns are to be considered as one entity. The equivalents of the words ‘college-day’ as single entity are ‘kallUri ANTu vizha’. The compound nouns are found and marked as single entity using the POS of the source sentence. The POS of this example sentence is, “The/DT auditorium/NN is/VBZ decorated/VBN for/IN the/DT college/NN day/NN celebrations/NNS”. Here the words ‘college’ and ‘day’ are annotated as ‘NN’. The consecutive words that annotated as ‘NN’ are marked as single entity in the pre-processing, and during lexicalization the equivalents of this entity are found and replaced in the terminal node.

Transliteration

Transliteration is an automatic method that converts words/characters in one alphabetical system to corresponding phonetically equivalent words/characters in another alphabetical system(Vijaya MS, 2009). The Named-Entities are transliterated from English to Tamil using the tool that based on ‘Sequence Labeling Approach’. 30k person names and place names (English-Tamil pairs) are used for training using SVM. The words that are not present in the root lexicon (Out Of Vocabulary words) are also transliterated in the same manner.

Synthesizer

The Morphological synthesizer glues the lemma and the morphemes to form a word using orthographic and morphophonemic rules. The lite version of Amrita Morphological Analyzer and Generator (AMAG) is used for synthesizing the words. The synthesizer requires the information along with lemma as an input. This information has to be gathered from the lexicalized target phrasal structure, the parse tree of the source and from the typed dependency information. The lexicalized tree has the lemma of the target language in the terminal nodes. Synthesizing these lemmas is very important in the process of translation. Synthesizing the noun gives the relationship between two nouns and in the case of verb it glues the TAM and gender information with the verb.
For example, the word ‘avan’ is given as input to the synthesizer with the morpheme information ‘DAT’ as ‘avan + DAT’ (dative). The first level of the synthesizer replaces the morpheme information with the correct morpheme that to be synthesized with the lemma as ‘avan + ku’. The second level of the synthesizer applies the spelling rule on ‘avan + ku’ and after the application of all possible rules; the second level outputs the synthesized word ‘avanukku’.

**Testing and results**

The current system is tested with the corpora of Tourism domain that we developed for DIT funded EILMT project. On testing with the corpora of the size of 2000 sentences, 60 % of the sentences are translated well and 70 % of the sentences are comprehensible. On the module wise testing, 80 % of the sentences are reordered perfectly, 60 % of the sentences are lexicalized properly with its correct target word, and the accuracy of the transliteration module is 93.3 % and more than 90 % of words that process through synthesizer module are synthesized properly, provided the information extracted and given as input to the synthesizer module is correct.

**Conclusion**

At present, the prototype version of the MT is developed with the 20k lexical entries and very few (40) reordering rules. In future, Word-Sense-Disambiguation-module can be plugged-in to the system to resolve semantics ambiguities. The system with all the necessary modules are in place, scalability is a key to improve its performance. Transliteration, Morph-synthesizer and extracting features are on its own a big task and these have to be enhanced as well to improve the overall performance of the system along with the root word lexicon and the reordering rules.

**Bibliography**

5
தலைசினில் தமிழ் எழுத்துக்கள்
செய்யப்பட்ட ஆய்வு
Bilingual TTS for Tamil and English

AG Ramakrishnan, Vikram LR, Abhinava, ShivaKumar HR
Medical Intelligence and Language Engineering (MILE) Laboratory,
Department of Electrical Engineering, Indian Institute of Science, Bangalore 560012.
agrkrish, vikram.ckm, abhinav.zozo, shivahr@gmail.com

Abstract
An unlimited vocabulary text-to-speech engine has been developed, which currently handles both Tamil and Kannada Unicode text. The input text is processed by a grapheme to phoneme converter module, which uses language specific pronunciation rules to convert the text into an unambiguous phonetic representation. This text is then parsed into demisyllable like basic units. The occurrence of these basic units are searched for, from the phonetically rich spoken database, which is segmented and annotated at the phone level. An unit selection algorithm then selects the best combination of the available speech units to be concatenated to synthesize the speech, which is then converted into .wav format.

Introduction
Text to Speech (TTS) synthesis is an automated encoding process, which converts the given text in a specific language into speech. Till date, only English and some European language TTS systems have gained commercial importance due to their quality output. This paper primarily deals with developing a modular, unit selection based TTS framework for Indian languages. Bilingual Tamil and English TTS is developed for this purpose. However, this framework can be easily modified for any other language. The TTS framework developed is concatenation based, with polyphone taken as the unit of concatenation. This framework is further optimized to suit embedded applications like mobiles and PDAs.

We designed and developed corpus-based concatenative Tamil speech synthesizer in Matlab and C. A concatenation based speech synthesizer requires a rich and large speech database with varied and natural distribution of prosodic and spectral characteristics of speech sounds. The sentences to be recorded need to be selected from a text corpus. We used CIIL (Central Institute of Indian Languages, Mysore) Tamil text corpus for our research. A greedy algorithm is used to select phonetically rich sentences from this huge corpus. This resulted in 1026 sentences, which were recorded from a professional, native Tamil speaker. These sentences are segmented offline and the database is organized in such a way that it facilitates faster search.

During synthesis, from the phonetic transcription of the sentence to be synthesized, specifications of the required target units are predicted. Units are then selected from the database that best match the target specification according to a distance metric and a concatenation quality metric. These units are then concatenated to produce synthetic speech. There may be audible glitches in the output after concatenation. This could be because of either poor segmentation of the speech database or improper selection of units by the TTS framework. In our case, we know that the segmentation is nearly error-free. Hence, post-processing is performed on the final set of units. This includes smoothing the pitch contour,
during concatenation, at junctions of units with unacceptable pitch discontinuity. Our experiments reveal that about 15-20% of the unit junctions require pitch smoothing. Optimal coupling technique is then used to concatenate these units at appropriate positions. This resulted in intelligible and reasonably natural synthetic speech.

Intelligibility of synthetic speech also depends on selecting the units that match the target phonetic contexts. At times, the required phonetic context may not be available in the database. In such case, we propose that similar phones that are perceptually indistinguishable may replace these phonetic contexts. The most confused pairs of Tamil phones, which can be replaced by each other in specific contexts at the time of synthesis (if they are not available in the corpus) are found. Explorative experiments to determine the applicability of incorporating these techniques resulted in high mean opinion scores for the synthesized output from the native Tamil evaluators. Hence, we consider that this possibility of replacing missing phonetic contexts can be used in practical TTS.

Finally, when any person speaks the same sentence repeatedly, the speech waveforms don’t have identical characteristics. With this motivation, the final portion of my research attempts to analyze the variability of characteristics of different instances of speech, when a speaker utters the same sentence multiple times, at different times. The idea is to look at the possibility of generating a slightly different synthetic speech each time the same text is synthesized, thus trying to make the TTS sound not monotonous and more human like.

Also, we observe that incorporating prosody and pause models for Indian language TTS would further enhance the synthetic speech quality output. These are some of the potential, unexplored areas ahead, for Indian speech synthesis.

**Motivation for bilingual TTS**

In the present scenario, usage of English in Tamil text has become common and inevitable. If such words are omitted in TTS, the TTS would be less effective. Hence we have developed a bilingual Tamil TTS for generating Tamil and English by using the same synthesis Tamil data for both the languages. The sparsely occurring English text is converted into phonemes using a separate grapheme to phoneme converter and the corresponding phonemes are obtained from the available Tamil database for concatenation. The initial results are encouraging and we are working on some more improvements for better sounding English.

Tamil synthesis database has 5 hours of Tamil sentences recorded by a male professional Tamil speaker. The recorded database is rich in phonetic context and phonetic variations. The TTS takes Tamil Unicode input, and performs equivalent phonetic translation. In Tamil, some letters and phonetic contexts influence the phoneme of a letter, such as “ka” can become “ga” in some phonetic contexts. The rules for these phonetic changes are coded as rules in the grapheme to phoneme converter. The phonemes are then again grouped into polyphonic cluster. The clusters are searched in database to find a best choice is selected for a given context. We have developed a prosody based unit selection algorithm to further enhance the best selection for a given text.

**Features/Specifications of the TTS**

1. **Unlimited Vocabulary**: Any sentence involving any combination of native words of Tamil (or Kannada) is handled.
2. Quality: The intelligibility of the TTS is quite high, and it is also acceptably natural.

3. Text Encoding: Only text entered in Unicode will be handled. Other proprietary or public font encodings are not handled, and will not be handled, even in the future.

4. Web Demo: The TTS has been made available as a web demo. Go to the link, http://mile.ee.iisc.ernet.in:8080/tts_demo. The TTS demo page can be seen and a box, where the Tamil or Kannada text in Unicode must be submitted.

5. Test Input to the TTS: If you want to submit your own custom text input, you can do so, by typing using our open source Multilingual Indic keyboard interface: மாண்டியா அம்மா or Vishwavaangmukha), which you can download from http://code.google.com/p/indic-keyboards


7. Testing: The web demo of Tamil TTS has now been tested by hundreds of people from around the world, and many GB's of synthesized .wav files have been downloaded by them.

Current Limitations

It cannot handle numerals, proper nouns and words originating outside the current language handled by the TTS, sentences needing intonations changes, such as interrogative and exclamatory sentences. There may also be some mispronunciations at times. In the case of long sentences, even the pauses (phrase breaks) may not be at the instances, where a human speaker will naturally pause for better intelligibility and clarity.

Future Enhancement

All the current limitations mentioned above are being addressed by us, and will be updated in time. The numeral handling facility will be added very soon. Some of the causes for mispronunciations have been identified, and will be corrected next. However, further enhancement of naturalness requires significant research in computational linguistics and prosody and hence, may take considerable time to reach good quality.

Conclusion

MILE Laboratory has teamed up with Bookshare.org, an International non-profit organization, to provide Tamil and Kannada digital books (copyright free or permitted by authors) online to print-disabled people (visually challenged, old people with vision disabilities and people with other disabilities that make it impossible for holding a book and turn pages of it). The MILE OCRs and TTS in the respective language (Thirukkural / Vak) will be used for this purpose, and thus the printed content can directly be heard as speech on a desktop computer or laptop.

Acknowledgment

The TTSs were developed as part of a Research Project, funded by Ministry of Social Justice and Empowerment (MSJE), Government of India. The author also thanks Karthigeyan, Jayavardhana Rama, Prathibha, Muralishankar, Lakshmish Koushik, Parthasarathy, Arun, Abhinav, Vikram, Shivakumar and others, who contributed to the TTS research and development over a period of time.
Voice based Tamizh Tutoring System for Primary School Children to Enhance their Tamizh Learning Interest

P. Vasuki and C. Aravindan
SSN College of Engineering
Chennai, TamilNadu, India
VasukiP@ssn.edu.in, 2AravindanC@ssn.edu.in

Abstract
In this paper, we focus on interactive tamizh learning system for children. Our system displays a picture and asks the child to recognize the picture and to speak out its name. Our speech engine captures and analyses the speech to give proper feedback. The speech engine has been trained with a corpus of words to obtain probability distributions (templates) of spectral coefficients of words. Spectral coefficients of the input speech are extracted and the word is recognized by matching the spectral coefficients with the templates. This word is then compared to the original text of the picture being displayed. By making many trials at different levels, the system induces the child to utter the object identity in tamizh. A scoring system runs at the background to calculate scores based on number of fouls and success. Providing picture along with the word helps the child to easily register the word in written form and meaning of it. The child who receives such training will acquire quick tamizh reading and speaking capacity. The conceptual mapping of images with words improves the declarative knowledge of a child.

Keywords – Speech Recognition, Computer Based Training, Tamizh Tutoring System, and Hidden Markov Model

1 Introduction
Pure class room teaching may be monotonous and computer based training acts as good supplement to it. Computer based training becomes more effective when it has a speech interface. Such systems are popular in foreign countries, but not available in tamizh, as there is less focus in tamizh speech processing. The system that we have designed provides a friendly environment for children and enables them to talk to the system in tamizh, while learning tamizh. Thus our system may help to motivate learning of tamizh & makes it joyful and interesting.

The Declarative knowledge is factual or experimental [7]. Displaying an object, pronouncing it, displaying its text and make a child to realize and later recognize improves declarative knowledge of the child and enables the child to map the word with real meaning of it. The child is active while learning by using her psychomotor skills, i.e. the child talks to the system using mike, uses keyboard and mouse for navigation and also using her cognitive skills to recognize the picture by looking it, recall, utter it correctly, learn to utter, learn to read.
The system gets speech input using omni directional mike and records the speech in wav format. The speech characteristics vary at each instant, each speaker even for same word. Thus we have gone for extensive training with ten speakers of various age groups and ten utterance of each word; i.e. every word is trained with hundred instances of utterance. Likewise the system is trained with hundred and twenty five basic tamizh words which are collected from first standard tamizh book.

HTK toolkit (Hidden Markov Model Toolkit) is used for building the speech engine[6]. The speech samples are parameterized in the form of Mel Frequency Cepstral Coefficients (MFCC) which reflects the characteristics of perception. In training, the word is divided into ‘n’ states. For each state an acoustic model is generated based on Gaussian Mixture Model (GMM). The probability of each state initialized and re-estimated based on its sequence using Forward Backward algorithm. HMM Model is generated for each word as initial state probability and sequence of states with transition probability.

In testing, MFCC features of testing word are generated and mapped against the model of each and every word in the vocabulary set, using Viterbi Decoding Techniques. The training wave gets the label of the word that closely matches with highest score. If the word uttered is not the original name of picture, it plays a sound of it with the help of look up table. A scoring system is running at the back to calculate the score and when the child quits, the score is displayed.

Section II talks about the system architecture in brief and section 2.1 briefs about speech Engine, 2.3 describes about tamizh tutoring system and section 3 briefs about scope and future enhancement.

2 Architecture
The project is of two phases (ii) tamizh Speech engine (ii) Tamizh Tutoring system.

![Fig.1. Architecture](image)
In first phase, Speech Engine has been developed with extended training with different speakers and various words. The speech Engine recognizes the word and converts into text. In second phase is tamizh tutoring system is implemented with speech interface on the top of speech engine. Fig. 1 shows the architecture of the complete system.

2.1 Speech Engine

Our Speech engine is an Isolated SpeechRecognizer (ISR) implemented using Hidden Markov model. It takes isolated word as a processing unit recognize it and produces text. The system’s word error rate is less when tested against system gets input as isolated word than connected word recognition. Our tamizh tutoring system needs only isolated speech recognition.

A. Data Preparation

i) Corpus Collection

The speech Corpus has been collected from 10 speakers of different ages( 6 year to 44 years) and different gender with sampling frequency of 16 khz and at 16 bit quantisation rate. They read 120 words in neutral form. Every speaker utters each word 10 times. The words are annotated manually. The signal is pre-processed before training.

Windowing is done to avoid the discontinuities of frames at stating and end [2]. The signal is divided into frames and windowed using hamming window of size 20 ms with the overlapping window size of 15 ms.

\[ S_n = 0.54 - 0.46 \cos(2\pi(n-1)/N-1)S_n \]  

ii) Feature Extraction

For each frame MFCC Mel frequency cepstral Coefficients are calculated. MFCC is a representation of the short-term power spectrum of a sound, based on linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency [4].Fig. 2. shows the processing steps of deriving MFCC from speech.

![Fig 2 : Derivation of MFCC From Speech samples](image)

MFCC -Mel frequency coefficients are working based on perception of word by human ear. Mel frequency Cepstral coefficients are 12 filters output that have been collected from the short time frame.

MFCC Feature vector of size 39 represents the signal of the frame.13 parameter value (12 MFFCC coefficients and absolute energy of the frame), 13 delta coefficients (variation of cepstral coefficients
between consecutive frames) and 13 acceleration coefficients are calculated for each frame. Fig 3 shows the screen shot of MFCC collected from different frames of a speech sample.

**B. Word Modelling**

The quantized MFCC parameters are used to construct a HMM model for each word.

HMM – Hidden Markov Model is as an efficient approach for short time stationary stochastic sequence [4]. Thus speech signal is modelled using Hidden Markov model[9].

In HMM training, HMM Model is created for each and every word in the training set. The Model is \( \lambda = (\Pi, A, B) \), where \( \Pi \) is the initial probability distribution, \( A \) is state transition probability and \( B \) is output probability. For a class whose data is considered to have multiple clusters, a Gaussian mixture model can be used to represent its probability distribution. The signal is represented by \( N(\mu, \sigma) \). The output probability is estimated using Gaussian Parameter \( (\mu, \sigma) \). The state transition probability is calculated based on state changes.

**C. HMM Parameter Initialization**

The word should be divided into number of sequential states. Number of states is fixed for each word. We use three utterances of a word to form a mixture. HMM parameter template is created with initial values of mean 0.0 and variance is 1.0 for the 39 feature vectors.

Gaussian Mixture model is used to find the probability distribution of the state of every mixture using expectation Maximisation algorithm. The signal is represented using Gaussian distribution using the formula [4]

\[
\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}
\]
**D. HMM Parameter Re Estimation**

The word is divided equally into number of states based on timing. The frame is expected to belong to a certain state. With that expectation the mean and variance of every state is calculated. The probability of each frame belonging to the corresponding state specified is calculated based on the mean, variance of that state and feature value of that frame. If the probability maximize the frame belongingness in that specified state the frame is retained there else moved to the other state where its probability of belongingness is more. Thus the HRest tool used to re estimate HMM parameters using Baum Welch Algorithm. Baum Welch algorithm estimates the probability in the forward backward sequence. i.e. the probability of n’th frame is calculated from the previous n-1 frame, in Backward algorithm the probability of n’th frame is calculated from the remaining n+1 to N frames. HMM Parameters are defined based on forward and backward probabilities. An acoustic model has been generated for every word as a sequence of states.

![Fig. 4. Re estimated values HMM Parameter after applying forward backward algorithm.](image)

**E. Word Net (Language Model) formation**

Word net describes about the different sequence about the word occurrence and the probability of the sequence [5]. The apriori knowledge about the sequence of states improves the performance of speech recognition system [10]. Dictionary defines the grammar about the sequence of word occurrence. Our system recognize only isolated words, thus dictionary is formed such as individual word as single entity.

![Fig 5 Word net – Network of words formed between 12 tamizh words.](image)
The word net is represented in SLF, ie Standard Lattice Format in terms of nodes and arcs. Node represents word and arch represents link between words. In the current word net we have 120 words (including NULL words) and arcs among the word. Fig 5 shows screen shot of sample file in Standard Lattice Format developed with 12 words network.

A statistical word net is formed by having words as nodes and the link –log transition probability as the link between arc.

F. Recognition

HMM recognizer recognizes the words based on the language model (SLF) and acoustic model created by training set. Initially forced alignment is used to align the state of the testing utterance [5]. The Viterbi decoding algorithm matches the state with the trained states. At every step, P suitable states are selected among the different states of M models. Then it produces the sequence of the next many possible states. Among those many possible states few highest probable states are chosen and so on. Likewise the lattice structure preceded and compared with the training pattern. For every possible state in every model the probability of observation is calculated. The overall probability of all possible states per model is multiplied. The model with the highest overall probability will be chosen as the resultant model (word) of utterance.

The word uttered is recognized based on Viterbi decoding techniques[8]. The HMM parameter of the state sequence of uttered word is generated and matched against the sequence of trained words as shown in the figure 6. Many similar words have the higher probability and the word is labeled as label of word which has the nearest matching or has highest probability [5].

![Image](image.png)


**Fig 7. HMM Model for Isolated word Recognizer – The observation vector applied to HMM models of different words.**

The probability of input belonging to different models are compared in Decision Logic. The model with highest score is the model of input observation.
2.2 Tamizh Tutoring system

The tamizh tutoring system is devised to enhance the tamizh learning interest amongst children. A Picture driven system along with audio attracts the children to learn tamizh language. It bridges the gap of lack of interest in learning with usual keyboard and mouse interface. Therefore there will definite impact in education system on this interactive tutoring.

This tutoring system has been developed with inputs from various theories of learning such as Bloom’s Taxonomy [6]. According to it, the learning consists of three domains. Cognitive, Affective, and Psychomotor skills. Cognitive learning is improving the mental ability (Informative, Recognizing and Recalling) - our system brings out cognitive skills by making the child to recognize an object, recall it. It addresses Affective skills (receiving, responding valuing, motivation and appreciation) by responding on her recognition. It also brings in Psychomotor skills (involve more physical activities) - by making the child to use microphone, mouse keyboard and speaker.

A picture of index one is displayed and a child is asked to name the picture. The child has to record the identity of object in his voice. The word is recorded and given as the input to speech engine. The speech engine produces the corresponding word in textual form. The word is mapped into index and matched against the index of the picture displayed. The matching module sends a signal to scoring system. The scoring system evaluates and pop out choice for next trial if needed. The system gives maximum three choices per word to a child, if the child recognition is wrong or uttered the word with different pronunciation, the word is displayed to child in textual form and audio corresponds to the word is playing. On success, the system just display the object identity in text then system iterates through next picture till end. At the end of picture gallery the system displays score with the word of appreciation and continue till the child wish to quit. This system not only trains the oral pronunciation of word but also trains the object in written form. Figure 8 shows the screen shot of VBaTTS system.

![Fig:8 VBaTTS - Screen Shot – Ready to record a child voice on “padivu sai” button and “aduthu” button helps to navigate on picture gallery.](image1)

![Fig 9. VBaTTS Evaluation sheet. It shows screen shot Final evaluation sheet which displays the evaluation result is not good and asks child wish to continue or not.](image2)
The scoring system adds score based on number of foul and negates score on failure. Child’s satisfaction or motivation or feeling of competence is also important. The child’s achievement level is meaningless if the child chooses not to exercise the skill after acquiring it. Thus the system operates in different level and the score is displayed in grade if it is Excellent or very good else the system just says the score is not good enough and asks the child to reattempt.

**Conclusions and scope for future work**

The speech engine is not producing 100% accuracy. To overcome this, system has to be enriched with speaker normalization and emotion normalization. The accuracy of the system also can be improved by adding some more parameter like speech rate and intensity etc to build the acoustic model of a word.

**Acknowledgment**

We acknowledge Dr. T. Nagarajan, Professor and Head, Department of IT, SSN College of Engineering, Chennai, for his valuable guidance in developing speech engine. We also thank AICTE for funding a project on development of tamizh speech engine.

**References**

Prediction of Pauses in TTS – Tamil

P. Arulmozhi
Project Associate, MILE Lab
Department of EE, IISc, Bangalore.
arulmozhi@ee.iisc.ernet.in

A G Ramakrishnan
Professor, Department of EE
IISc, Bangalore.
ramkiag@ee.iisc.ernet.in

Abstract

Text to Speech (TTS) involves the task of converting text typed in electronic format into a speech signal. In MILE lab, we are involved in making a TTS system for Tamil and Kannada. In this paper, the contribution of syntactic information such as Part of Speech (POS) Tags in enhancing the quality of a Text to Speech (TTS) synthesis system for Tamil is researched. The quality of a TTS system is measured by the intelligibility and naturalness of the synthesized speech. The NLP module of the TTS system (for example, text normalization) contributes to not only to its intelligibility, but also to its naturalness, by improving the prosody. The stress and pause modeling can be improved using the POS and other syntactic information. In a sentence, where there should be a pause, and where there should not be a pause has to be identified for the naturalness of the produced speech. Because, a sentence without pause or with identical pause intervals sounds robotic. Also, pause at a wrong place makes the sentence unnatural and there is a possibility of change of meaning. For example take the following sentence,

avarukku inRu <P> mAlai kitaittatu.
avarukku inRu mAlai <P> kitaittatu.

<P> here indicates that there should be a pause. The pause given in different places give different meanings. Syntactic information such as Parts of Speech (POS) can be used for identifying the rules for pause in a particular sentence. A rule based POS tagger is developed for this purpose without using a root word dictionary. Currently, manual evaluation shows an accuracy of approximately 74% using only the lexical rules. The performance is expected to improve after the context sensitive rules are applied. Rules are made for predicting the insertion of pause at the right place. The manual evaluation of pause insertion shows a significant improvement in the naturalness of the produced sentence.

1. Introduction

This paper presents a rule based Parts of Speech (POS) tagging method in the perspective of improving the naturalness in the synthesized speech. The quality of a TTS system is measured by the intelligibility and naturalness of the synthesized speech. There are two main modules in a TTS system. One is the Natural Language Processing (NLP) module, which takes care of the production of phonetic transcription, intonation and rhythm. Another is the Digital Signal Processing (DSP) module which takes care of the production of the speech wave-form of the given text. The stress and pause in the speech contribute majorly to the naturalness, which is controlled by the NLP module. Using this POS tagger, we try to find out the right place to introduce pause in the synthesized speech. Introducing a high degree of naturalness is theoretically possible, but the rules to do so are still to be discovered (Jonathan A. 1996...
Many linguistic aspects are analyzed (Thierry D) and syntactic information such as POS tagging is considered important to achieve a good TTS. In this paper, we present a POS tagger created for Tamil, which is a highly agglutinative and partially free word order language.

2. POS Taggers

The purpose of a POS tagger is to automatically find out the syntactic category of a word in a sentence. Different methods may be followed to do POS tagging. Most commonly used are rule based, statistical, and transformation based methods. Rule based taggers work on predefined linguistic rules for deciding the syntactic category of a word in a sentence. These rules may be lexical or context sensitive and they are language dependent. Statistical taggers work on the calculated probability, in which a POS tag for a particular word is decided based on the lexical and contextual probability. A training corpus is used to train the system and the input sentence is tagged based on the probability which is calculated using the training corpus. Transformation based taggers derive rules based on learning. Those rules are used to find the syntactic category. In English, Brill's tagger is the most commonly used, TBL based tagger. There are statistical, rule based and hybrid taggers being worked on for Tamil (Arulmozhi P, Sobha L, 2006).

3. MILE POS Tagger for Tamil

3.1 Purpose of MILE tagger

In natural speech, we introduce stress and pause at the right places, so that it is easily understandable. In a TTS system, the DSP component produces the speech wave-form and the NLP component is responsible for the naturalness of the produced speech. It should identify the right places to introduce pause. The amount of pause to be introduced has been identified. The pause is introduced using the category of the words in a sentence. Other syntactic information such as shallow parsing and clause boundary identification will be helpful not only in identifying pause, but also to find out the pitch and intonation.

There are a few POS taggers available for Tamil. They have been developed for the purpose of preprocessing for NLP activities such as machine translation and information extraction (Arulmozh P. et.al 2004). The probability based POS taggers need huge training corpus. They provide the POS tags according to the tagset used for training. In the case of a TTS, we do not need such detailed tagging.

3.2. Nature of Tamil Language

Tamil is a morphologically rich, agglutinative and partially free word order language. Compound words are common in this language, where two or more words are combined to form a single word. The case markers and tense markers appear as inflections of the root word itself. For example, taking the word ‘varukirAn’, the inflections and root word can be split as follows.

\[ vA + kir + An \]

\[ vA \] - root word

\[ kir \] - present tense marker

\[ An \] - 3rd person, Singular, Neuter gender.
Tamil is a partially free word order language because changing the word order to some extent does not affect the meaning of the sentence. However, this order change can not occur within a phrase. For the sentence,

1 ‘Aciriyar nanRAka paTitta mANavanukku paricae kotuttAr’

   Teacher thoroughly studied Student+Dat prize+Acc gave+Hon

   The teacher gave the prize to the student who studied thoroughly.

It can be written as

1 ‘nanRAka paTitta mANavanukku Aciriyar paricae kotuttAr’

2 ‘Aciriyar paricae nanRAka paTitta mANavanukku kotuttAr’

but,

1 * ‘Aciriyar paTitta mANavanukku paricae nanRAka kotuttAr’

changes the meaning. So, within a phrase, the word order does not change.

### 3.3. Tagset

A tagset is the set of all tags used by the POS tagger. There are two levels of tags - the main tags and sub tags. The main tags identify the main category of the word such as noun, verb or adjective. The subtags identify the category of the inflections such as person, number, gender, and tense. For the purpose of TTS, we do not need very detailed tags unlike other NLP activities. But using only the main tags would not give sufficient information. So we need some of the sub tags too. So, as a special case, we have developed a tagset for the purpose of inserting pause in a sentence. In our tagset, each tag is a combination of a main tag and one or more sub-tags. The nouns take the case and plural markers and the verbs take person, number, gender, and tense markings. Apart from this, pronouns have person, number, and gender. The clitics are suffixed to the root word to form adverbs and conjunctions. Then the dates, numbers, and punctuations are also tagged separately. English POS taggers and some of the Tamil taggers (Dhanalakshmi V. et. al. 2009) use monadic tags. Monadic tags do not give information on inflections, which is important for TTS. We use structured tags such as "NN+pl.acc" in which different pieces of information serve in different parts of the rules. This tag means, a noun in with the inflection for plural and accusative. We have 15 main tags and 30 subtags adding up to 45 tags.

### 3.4 MILE Tagger:

**POS Tagger.** This POS tagger is a rule based one. We do not use a root word dictionary. The tagger is based on a two stage architecture. The block diagram of the POS tagger is shown in Figure 1.

In the block diagram, each block explains its functionality. The first stage has the lexical rules and the second stage has the context sensitive rules. First, a sentence is taken as input and split into tokens. For each token, the suffixes are identified. Then, using the lexical rules, which work at the word level, each word is assigned a POS tag according to the suffixes identified. Then, this output is given as an input to the second stage, where the context sensitive rules change the tag if it is wrongly tagged by the lexical rules. Thus, the final tagged sentence is obtained.
Separate tables are created for programming purpose with the list of suffixes identified. A lexical rule looks like,

\[ 2^*1+1^*1, \text{NN+pl.acc} \]

This means, the suffixes indexed \( 2^*1 \) (Suffix Table 2 Column 1 - kaL) and \( 1^*1 \) (Suffix Table 1 Column 1 - ae) occur in a sequence, the word will be tagged as Noun+Plural+Accusative. Here 'kaL' is the plural marker and 'ae' is the accusative marker. There are 13 such tables which list 103 suffixes identified and put in. These suffixes are used by the lexical rules. The context sensitive rules are embedded in the system. For example the following can be considered as a context sensitive rule.

‘If a sentence starts with a verb, change it to noun’.

Since Tamil is a verb ending language, a sentence can-not start with a verb. So, if the first word of a sentence is wrongly tagged as a verb in the first level, it will be corrected in the second level. The combinations of lexical rules including the inflections are 533 and the number of context sensitive rules are 4.

For any POS tagger to work correctly, the sentence boundaries have to be identified. We use a sentence splitter for splitting paragraphs to sentences. Input to the sentence splitter is any Tamil text such as paragraphs. The output is an array of sentences. This process is also embedded in the POS tagger based on our need. We use a rule based sentence splitter and the rules are heuristic in nature.

4. Pause Model

Pause model. Insertion of the right amount of pauses at the right places adds to the naturalness of the synthesized speech. With a natural language text, native speakers introduce pauses with the knowledge of the language acquired. But in a TTS system, those pauses need to be inserted by the system at the right
places. For European languages such as Spanish, there are rule based pause models developed and experimented (Rafael M 2002). A wrong pause inserted between two words may make the synthesized speech unnatural. For simplicity, such an example sentence is illustrated in English. Here, the notation <np> denotes “no pause” between the words, whereas <p> refers to the required pause.

Example:

The<np>book<np>is<p>on<np>the<np>table.

Speech synthesized as per the tags for the first sentence will appear perceptually natural, whereas for the second sentence, a wrong pause between words ‘on’ and ‘the’ will appear perceptually unnatural. Hence POS information and pause are very important in the context of TTS. In this paper, we focus on pause insertion between successive words by pause prediction from the POS tags estimated from the input text. Presently we use key words and heuristic rules for inserting pauses. Rules for inserting pauses at the right places are created according to the POS tags.

At present, in the syntactic level, only POS tags are considered for identifying pauses. It is considered as a basic preprocessing needed. On top of this, the phrase chunks may be identified, and that will be useful in identifying the places where pause must be inserted. In that process, the phonological phrases must be identified finally for identifying pauses and intonation.

Six levels of pause have been identified, which determine the duration of the pause. In Chinese pause model, they use minor, major, and punctuation breaks (Fu-chiang et.al 1997). We have defined <P0>, <P1>, <P2>, <P3>, <P4>, and <PW>. P0: No pause, P1: Lowest pause, P2: Medium Pause (Ex: Pause after a comma), P3: Significant Pause (Ex: Pause after a semicolon), P4: Highest Pause (Ex. Pause between sentences). The pause levels P0 to P4 are derived from the existing synthesis database. This work has been carried out for Tamil and can be extended to all Dravidian languages with no major change. <PW> is the common pause between each word. Wherever <P0>... <P4> is not identified, <PW> is assumed.

At the initial level, <PW> is assumed for each word, and rules for converting it to <P0> to <P4> is made. The sample rules are given below.

1. There is no pause (or may be very minimal pause) between a number and certain words following, such as ‘mani’, ‘latcam’, ‘KOTI’. There is a list of words defined for this rule. Any noun in plural form, (NN+PL) after a number does not have a pause.

2. If the previous word has an accusative/dative marker, and the current word is a postposition, there is no pause between the current and the previous words.

   Ex : avanai <P0> pola,
   avanukku <P0> pin

3. Combine the words with POS tags Adjectival Participle (AJP) and Noun (NN - any number of them) occurring together. There is no pause between them.

4. There should be a pause before quantifier (Q).

   Ex : (azhakiya kiraamamum)<P3>(oru periya ooraaTciyum)
There are 15 such rules made for converting <PW> to <P0> - <P4>. In natural speech, we do not give any pause between most of the words. Taking that into consideration, we did another experiment. Instead of taking <PW> as default pause, we took <P0> as default i.e. there is no pause between any word initially, and then inserting pauses using rules. This reduced the rule set from 15 to 8, because we had more rules for <P0>. The output from both is obtained and given for evaluation.

5. Output and Evaluation

The output of the system contains the original sentence, the predicted POS tags and the predicted pause levels. All are displayed parallelly. An example output is shown in Figure 2.

<table>
<thead>
<tr>
<th>C.B.S.E</th>
<th>stream</th>
<th>loc</th>
<th>studying</th>
<th>students+pl+dat</th>
<th>next</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>NN+loc</td>
<td>VB+fut+3sn</td>
<td>NN+pl+dat</td>
<td>AJP</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>education</th>
<th>year</th>
<th>from</th>
<th>10th</th>
<th>class</th>
<th>exam</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>NN</td>
<td>PP</td>
<td>Q</td>
<td>NN</td>
<td>NN</td>
<td>VB+pst+3sn+neg</td>
</tr>
</tbody>
</table>

| <P0>      | <P0> | <P2> | <P0> | <P0> | <P0> | <P4> |

Figure 2. Example Tamil Sentence and Output

In the figure, the first line is the Tamil input; second line is the corresponding meaning in English; third line is the predicted POS tags and the fourth line is the pause levels identified.

This output is given to the DSP module and the wave form of the sentence obtained. The people who evaluated the outputs are native Tamil speakers, who did not have knowledge about the methods used for creating TTS outputs. 3 types of outputs are given to the evaluators and Mean Opinion Score (MOS) is obtained. 10 sentences are spoken by the TTS as follows:

- Without implementing the pause model
- After implementing the pause model with default as <PW>
- After implementing the pause model with default as <P0>

The scores are given from 1-5 according to the understandability and naturalness (1-worst, 5-best). The evaluation based on the mean opinion score gives encouraging results.

The rule based POS tagger is evaluated manually for the correctness of the tags. The evaluation is done manually. In a given tag, if the main tag is correct and the sub-tag is wrong, or vice versa, we take that as a wrong tag. The system gives 78% results without a root word dictionary. More context sensitive rules are to be added so that the accuracy of the POS tagger also can be improved.
References


Text Analysing and Retrieval System using Tamil Phonemes and Vector Space Model

Premalatha.R
rprema79@rediffamil.com
SCSVMV University
Kanchipuram, Tamilnadu, India

Srinivasan.S
sriniss@yahoo.com
Tamilnadu Open University
Chennai, Tamilnadu, India

Abstract

Intelligent information retrieval is one of the important topics in the 21st century. In Tamil documents, generally, morphology (separating noun and verb) concept is used to retrieve the text. In this system we use Tamil phonology so that we can widen our search criteria namely, Vowel – Short, Long; Consonant - Vallinam(Hard), Mellinam(Soft) and Idaiyanam(Medium). So the system would search quickly. Hence the performance of the system would increase. In the classical system, the user should give the exact word to retrieve the information. But in TAR system, as the system internally has its spell checker, the misspelled word can be corrected and also the information can be retrieved. This would be useful for Tamil literates, Tamil students, Tamil learners, etc.,

Introduction

For thousands of years people have realized the importance of archiving and finding information. With the advent of computers, it became possible to store large amount of information and finding useful information from such collections which became a necessity. The field of IR was born in the 1950s out of this necessity. Over the last forty years, the field has matured considerably. Several IR systems [9] are used on an everyday basis by a wide variety of users. The need for information retrieval [1] arises in the Tamil literature documents from the ancient era to the latest, which helps in sharing the data through the internet. In this paper, we are going to do text analyzing and retrieval system using Tamil phonemes and vector space model [8].

In this system, searches can be done in three ways namely (i) Main topics searches (ii) Subtitle searches and (iii) Keyword searches.

The input word would be given in Tamil using Tamil virtual keyboard.

In the classical system, the system [4] read the whole word and it identified whether the word is a noun or a verb. Finally it found the word from the database. In the past it took a lot of time for retrieval. Because in the older system once the entire word is entered, the system would do the searching process. But in the proposed system, the system would not wait for the entire word to enter; perhaps the searching process starts immediately after the first letter is entered. Because the system performs the
search processes for every letter of a word. This approach considerably increases the speed of the retrieval process.

**Tamil language:** Tamil is the oldest and truest of the Dravidian speeches. Tamil boasts its literary tradition for more than 2,200 years. Tamil is the most remarkable body of secular poetry extant in India (Tamil is a South Indian language spoken widely in TamilNadu in India). Tamil has the longest unbroken literary tradition amongst the Dravidian Languages. Tamil [1] has 12 vowels and 18 consonants. They combined with each other to yield 216 composite characters and 1 special character (aayutha ezhuthu). Summing up there are 247 letters in standard Tamil alphabet.

**Tamil literature:** It has a rich and long literary tradition spanning more than two thousand years. Contributors to the Tamil literature are mainly from Tamil people from Tamil Nadu, Sri Lankan Tamils from Sri Lanka and from Tamil Diasporas. A revival of Tamil literature took place from the late nineteenth century when works of religious and philosophical nature were written in a style that made it easier for the common people to enjoy. Nationalist poets began to utilize the power of poetry in influencing the masses. With the growth of literacy, Tamil prose began to blossom and mature.

**Literature Survey**

Language occupies an important position in the history of Indian cultural traditions. The central Government declared Tamil as one of the classical languages on September 17, 2004. Information retrieval [2] of Tamil literature is a difficult work to do because it was used in olden period Tamil format and it was on poetry format as well. Generally morphology approaches[3] are used for information retrieval of Tamil documents (Anand kumar M), 2009. An IR system returns a list of long documents to a user query. The construction and use of exploration models and search indices consumes processing time, memory, and disk space. Furthermore, in real systems any search and exploration methods must be computationally efficient. In particular, the delay perceived by the users is critical.

It is therefore important to develop methods that can speed up the search process while maintaining high perceived quality, particularly in the range of high precision and low recall which is most crucial in actual user settings. The TAR system is suitable for performing information retrieval using Tamil phonemes and vector space model that were organized for exploration of Tamil document collections

**Tamil Phonemes**

Native grammarians classify Tamil phonemes into vowels, consonants, and a "secondary character", the āytam.

**Vowels:** There are 12 vowels in Tamil, called uyireluttu (uyir – life, eluttu – letter). These vowels are classified into short (kuril) and long (Nedil), five of each type and two diphthongs, /ai/ and /au/, and three "shortened" (kurriyl) vowels. The long vowels are about twice as long as the short vowels. The diphthongs are usually pronounced about 1.5 times as long as the short vowels.

**Consonants:** Consonants are known as meyyeluttu (mey – body, eluttu – letters) in Tamil. It is classified into three categories with six in each category: vallinam (hard), mellinam (soft or Nasal) and itayinam (medium).
System Architecture

In the TAR system, 4 phases are used. They are (i) Classification (ii) Analyzer (iii) Retrieval and (iv) Spell Checker.

**Classification:** Most of the ancient Tamil literatures are rendered in the form of poetries. The critical edition of ancient Tamil works include 41 works namely 1) Thirukkural 2) Pura naanooru 3) Aga naanooru 4) Silapathigaram 5) Seevaga chinathamani 6) Manimegalai 7) Kundalakesi 8) Valayapathi 9) Padhinen Mel kanakku (18 Upper Classics) 10) Padhinen Keezh kanakku (18 Lower Classics) etc. Since most of the Ancient Tamil works are in poetry and in anthology forms, a Main class is derived with 41 categories and each category has various sub divisions. The sub divisions are classified as Main topics, Subtitles and Keywords.

**Analyzer:** Every letter has been analyzed by the analyzer using Tamil phonology[2]. If a key is pressed through the Tamil virtual keyboard, the analyzer would identify the letter through any one of the following – Vowels(V) or Consonants(C). Vowels are again classified into two types, Kuril(k) and Nedil(n). Similarly consonants are classified into three classes with 6 in each class and are called Vallinam(v), Idaiyinam(i) and Mellinam(m). Once the letter is identified, it locates the letter from the Database. The cycle would continue until all the letters in the word are processed. The index position should change for every letter until the complete word is entered.

\[
\text{Eg: } \bar{\ddot{\text{i}}} \rightarrow \ddot{\text{i}} + \ddot{\text{u}}, \quad \dddot{\text{a}} \rightarrow \ddot{\text{a}} + \ddot{\text{e}}, \quad \dddot{\text{y}} \rightarrow \ddot{\text{i}}
\]

\[(\text{C-v}) + (\text{V-k}) \quad (\text{C-v}) + (\text{V-k}) \quad (\text{C-m})\]

**Retrieval:** In the retrieval module [12], once the word is found in indices, it retrieves the poem and its explanations from the documents using vector space model [3]. The performance of the search are increased by splitting up the category by 3 ways namely (i) Main topics searches (ii) Subtitles searches
and (iii) Keyword searches. For example, if the input is given from “Main topics”, then the subtitle and keyword could be skipped.

**Spell Checker:** The spell checker basically checks the spelling of the given word. If the word is wrong or the word is not in the DB, then it would suggest some other word related to the input. Generally the user cannot enter the first letter wrongly, but he may do mid of the word for the letters like

1) ப 2) வ ம ர 3) ம ம ம

So the spell checker is designed in such a way that the first letter would be skipped and the rest of the letters in the word should undergo the spell-check. Basically it skips the first letter of the word and check the rest of the letters of the word. Finally it gets the word from the DB and replaces it. If there is more than one combination, it would list all possible combinations to select. Look at these examples below:

Case: 1 மமொர் → மமொர்; அமல் → அமல்
Case: 2 மமொர் , மமொர் → மமொர் மமொர்
Case: 3 வருங்காடு → வருங்காடு, வருங்காடு

**Vector Space Model**

In the vector space model [11], we represent documents as vectors. Term weighting is an important aspect of modern text retrieval systems. Terms are words, phrases, or any other indexing units used to identify the contents of a text. Since different terms have different importance in a text, an important indicator – the term weight[8]- is associated with every term. The retrieval performance of the information retrieval system is largely dependent on similarity measures. Furthermore, a term weighting scheme plays an important role for the similarity measure. There are three components in a weighting scheme is

\[ a_{ij} = g_i * t_{ij} * d_j \]

Where \( g_i \) is the global weight of the ith term, \( t_{ij} \) is the local weight of the ith term in the jth document, \( d_j \) is the normalization factor for the jth document. Usually the two main components that affect the importance of a term in a text are the term frequency factor[13] \( (tf) \), the inverse document frequency factor.

**TFIDF weighting:** TFIDF is the most common weighting method used to describe documents in the Vector Space Model[12], particularly in IR problems. we assign to each term in a document a weight for that term, that depends on the number of occurrences of the term in the document. In addition, IDF measures how infrequent a word is in the collection. This value is estimated using the whole training text collection at hand. The simplest approach is to assign the weight to be equal to the number of occurrences of term \( t \) in document \( d \).

\[ \text{weight}_{t,d} = \log(t_{f,t,d} + 1) \log \frac{n}{x_t} \]

where \( t_{f,t,d} \) is the frequency of word \( t \) in document \( d \), \( n \) is the number of documents in the text collection and \( x_t \) is the number of documents where word \( t \) occurs. Normalization to unit length is generally applied to the resulting vectors.
Evaluation

Objective evaluation of search effectiveness has been a cornerstone of IR. Progress in the field critically depends upon experimenting with new ideas and evaluating the effects of these ideas, especially given the experimental nature of the field. Since the early years, it was evident to researchers in the community that objective evaluation of search techniques would play a key role in the field. The two desired properties that have been accepted by the research community for measurements of search effectiveness are recall: [3] the proportion of relevant documents retrieved by the system; and precision: the proportion of retrieved documents those are relevant.

Experiments Results: Information retrieval systems comparison:

Input word is ու չոպ

<table>
<thead>
<tr>
<th>Title</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR using Morphology</td>
<td>0.41</td>
<td>0.44</td>
</tr>
<tr>
<td>IR using phonology</td>
<td>0.50</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Conclusion

The TAR system would be pretty much helpful for all Tamil literates and students to search and learn. The TAR system focuses on Tamil phonemes. Also performance tuning has also been done in the system. In addition with that, the TAR system has the spell checker concept which makes the system to search the data reasonably quick. This system currently supports 41 categories alone. In addition to this we can further add more documents in future. However this concept will be designed more useful to the users, in future.
References

7. Guo xian tan, Chsitian vivard, Alex.c, Information Retrieval Model for Online Handwritten Script Identification, International Conference on Document analysis and Recognition, Spain, 2009
Speaking tool in Tamil for vocally disabled

Shashikiran K¹, Abhinava¹, Swapnil Belhe², A G Ramakrishnan¹
¹MILE Lab, Dept of Electrical Engineering, Indian Institute of Science, Bangalore 560 012.
²Gist Group, CDAC Pune.
Shashikiran@mile.ee.iisc.ernet.in, abhinav@mile.ee.iisc.ernet.in,
ramkiage@ee.iisc.ernet.in, swapnilb@cdac.in

Introduction

It has always been a challenge to bridge the gap between vocally disabled and the masses. The development of sign language has only been partially successful in bridging the gap. It requires persons conversing to know the sign language. Our work is a conscious effort to overcoming this pitfall. The proposed methodology is a combination of two different entities namely (Online Handwriting Recognition) OHR and (Text to Speech) TTS. OHR deals with recognition of handwritten words online. In an on-line handwriting recognition system, machine recognizes the writing while the user writes. The OHR output is in Tamil Unicode & becomes the input to TTS. Unicode is a globally accepted encoding format which makes our application viable to be used in various circumstances. TTS is a system that takes Unicode text and produces natural and intelligible speech in that language. This enables the patients, who had laryngectomy and tracheotomy as well as the vocally challenged to communicate effectively. As vocal disability may be congenital or because of ailments like oral or throat cancer, our method serves equally to both.

The tool is based on a hand-held, Tablet PC based on Intel Atom processor. The user can write one sentence at a time on the screen using the stylus and then click the button “Speak”. The sentence is recognized and then converted into speech and spoken out. Thus, the patient can call the attention of the nurses or his relatives in another room easily.

Details of the OHR module:

In Online handwriting recognition, a machine recognizes, as a user writes on a pressure sensitive screen with a stylus. The stylus captures information about the position of the pen tip as a sequence of points in time. The sequence of point between a PEN DOWN and PEN UP signal defines a stroke. This spatio-temporal information of the character being traced is the only input available to the online recognition system. Also given a character, one can capture the different writing styles using the information from the stylus.

Given a Tamil word, we first run a segmentation algorithm to identify the individual symbols. This algorithm segments word level data into symbol level data, as the modeling of the data is done at the symbol level. The recognition is performed at this level and results are concatenated to form the words. The extracted symbols are subject to the following preprocessing modules: smoothing to remove noise, resampling to a fixed number of points for speed normalization and size normalization.
Once symbols are brought to a standardized form, a set of seven features namely,

| Pre-processed X & Y co-ordinates: | Preprocessed data points (x,y) are themselves good features |
| Pen direction angle: | At each sample point, the direction of pen tip movement from that point to the next point can be used as a feature |
| Normalized first derivatives of X & Y: | Derived at each sample point of the preprocessed Tamil symbol, are also used. |
| Normalized Second derivatives of X & Y: | Same as above. |

The preprocessing techniques & features are discussed elaborately in Rituraj et al [2]. These features are then fed to the SDTW classifier for recognition of the Tamil symbol.

Statistical Dynamic Time warping (SDTW): In SDTW, a reference character is represented by a sequence \( Q=(Q_1;Q_2;Q_3;Q_lq) \) of statistical quantities (states), as shown in Fig 1.

These statistical quantities include

1. Discrete probabilities that statistically mode the transitions between states. We have empirically used 20 states in our work.
2. A continuous probability density function that models the feature distribution at each state.

We have modeled this distribution as a multivariate Gaussian distribution for each of the 20 states.

![Fig 1. Transitions between states in SDTW](image)

While testing, the SDTW distance of test pattern to the reference model of each class is computed, and the test pattern is assigned the label of the class giving minimum SDTW distance. The definition of SDTW distance is different from that of DTW and is given in Claus et al [1]. Fig. 1 shows how the matching takes
place between the reference model and Test pattern. Matrix in Fig.1 shows the DTW path (path of best SDTW matching).

SDTW distance is the negative log state optimized likelihood of pattern T generated by the model Q, with optimal state sequence S, given by the Viterbi algorithm. So, models in SDTW frame work are similar to HMMs of particular type with state prior probabilities \( \alpha = (100...0)^T \) and are left to right models with step size of at most 1 and with null transitions (transitions that allow change in state without observation change i.e. transitions (0,1)). So the models in SDTW frame work can be trained by algorithms used for training HMMs. In our work, we used segmental K-means algorithm [3] for training SDTW model parameters.

Unicode Generation: Based on rules derived from the language, valid Unicode string is generated from the output labels of the classifier. This string is the input for the TTS module.

**Details of the TTS module**

The TTS is based on concatenative waveform synthesis [4]. There are 1026 phonetically rich sentences spoken by a professional Tamil speaker, which has been segmented and annotated at the phone level. The input text is passed through a grapheme to phoneme conversion module, after text normalization. Certain pause rules are added based on a preliminary POS tagger and several rules. The phonemic text is parsed into the basic units for concatenation and the units are searched for, from the synthesis speech database, based on context and prosodic parameters. The selected speech units are concatenated in the waveform domain using a pitch continuity metric [5] and a pitch synchronous concatenation methodology [6]. It takes directly Unicode Tamil text as the input and produces a .wav file as the output. This wav file is directly played by the tablet PC and hence, when the user writes a word, it will be read after a second by the system.

**References:**


2. Rituraj Kunwar et al. A HMM Based Online Tamil Word Recognizer, 8th Tamil Internet Conference, 2009, University of Cologne , Germany.


Standardisation of Modern Standard Tamil
Transcription for Comutational Tamil

Dr. Punal K Murugaiyan
Prof. Rtd. Annamalai University
Tamil Nadu, India.

SECTION 1 - Antiquity

Tamil is the senior most and uniquely distinguished member of the Dravidian family of languages, possessing maximum possible cognates and other common features of the group. Some scholars of academic reputation both native and foreign believe that it has at least sources to interpret, nucleus elements to share with quite a few languages of the world. Its phonology itself stands witness for its antiquity and primitiveness. Markedness features of the phonological pattern are the characteristic indices which could vouch for the earliest evolution i.e. the primitiveness of a language. Unmarked phonological features are the simple, earliest and most primitive ones. All the phonological elements i.e. phonemes of the Tamil language, without exception are surprisingly unmarked. Unmarked features, since simple and designating single articulation only, naturally the sounds unmarked must have developed at the initial stage of the evolution of speaking. The vowels and consonants of the Tamil language are unmarked for any of the phonological features of articulation, since the alphabets whether invented, adopted or adapted, as Alexander John Ellis (1842, p.2) points out might have been representing, at the time of introduction, only one individual sound rather only one allophone. Even now, though a lot of phonetic variants had emerged, the unmarked characteristics of obstruent and sonorant remain unchanged. This kind of unique unmarked phonological feature has not been so far noticed in any of the world languages except the one, Curier, the aboriginal Australian language, which again reinforces the theory of Lemurian vast land.

Standard writing

As Edward Jewitt Robinson (Tamil Wisdom, 1873, p.3) observes, the Tamil Sangams were like “……. the Royal Academy of Sciences founded by Louis XIV at Paris and made it a rule that every literary production should be submitted to their Senatus Academicus before it was allowed to circulate in the country, for the purpose of providing the purity and integrity of the language”. Even now the standard form of writing was strictly adhered while the pronunciation, the actual torch bearer, the communication base had/ has never been insisted, not even for name sake, cared for. The mushroom growth of the Tamil dictionaries, due to business motive, never spare even little time on the entry of pronunciation. Even now, the supposed edition of the great Tamil Lexicon of the University of Madras has ruled out the phonetic transcription, hammering it out mercilessly. It is very painful to notice that the scholarly comment of P.S. Subrahmanya Sastri on the entry of ūyam in the existing volume remains uncared (History of Grammatical Theories in Tamil, 1934, p.68).
The said entry reads, Sastri quotes translating, “......... the 13th letter of the Tamil alphabet occurring only after a short initial letter and before a hard consonant as alkam, and pronounced sometimes as vowel and sometimes as consonant”. But the medieval authors of grammatical treatises themselves are clearly stating that it is neither a vowel nor a consonant, and fix its pronunciation as guttural fricative which coincides with the current one. The irony is one of the advisers of the new edition, may be going to appear ingeniously discovered that őytam is a voiced double implosives!

Standard pronunciation

In the Sanskrit grammatical literatures there are separate treatises on phonetics like sikshas (general phonetics) and pratisakyas (phonetics of particular school of veda). Pronunciation is equally insisted along with the other features of grammatical principles. It is said that he who knows the distinction between the length and tone can secure a seat by the side of the master. Such an importance has been given for correct speaking. Due to the advent of linguistic science, pronunciation dictionaries even on specialized subjects appeared in European and a few other languages and phonetic transcription even in the regular dictionaries too, finds entry.

Importance can’t be ignored

It is very unfortunate that neither the grammarians nor the scholars of the other related fields of the Tamil language up till now have taken cognizance of this subject. But that can’t be no longer remain discarded as the demand is waiting right at their door steps due to unavoidable need of the monstrous advent of the computer pervasion in every walks of modern life. So no longer the Tamil savants ignore the importance of the pronunciation features of speech of the Tamil language as it had already established its supremacy in other languages. No one can think of an alternative for computer speech of Natural Language Processing

Standardization of transcription

Tamil since holding an extraordinary situation than the rest of the languages of the world i.e. being in current use, since the beginning of the historical period and preserving the ancient and primitive features of speaking due to its conservative elders, is now desperately in need of alteration in its alphabet giving space to new sounds which entered due to importation of knowledge and culture. Even the earliest extant grammar Tolkappiyam itself gives way to annexing new sounds, suggesting rules for accommodation. But in spite of the grammatical sanctioning the so called purists wage pseudo war for fancy sake.

New entries

As seen in the foregone paragraph one cannot go but accommodating new words into the corpora admitting new sounds. As it was already happened to be the practice of using Grantha letters for the Sanskrit and Prakrit borrowings and appended them to the existing alphabet, if any new sounds added may be represented by the Grantha letters which go in harmony with the Tamil scripts since having similar shape and structure. If one thinks of diacritic forms-allograph that spoil the whole beauty of the structure and system of the alphabet.
Standard variety

First of all, the standard variety of the speech, the pronunciation must be chosen as not only being scientific but also wildly to be acceptable and practicable. In this connection the elites’ and well accepted speech, mostly used in formal occasions of the educated people has to be taken as the data, ignoring all voice quality, intonation and idilolectic variations; mostly isolated words in all length are taken in to consideration.

Phonemes

There are 10 vowels (5 short and 5 long), 25 consonants (18 native: 6 plosives, 6 nasals and 6 approximants; and 7 marginal: 4 voiced plosives and 3 fricatives) – they are in the native and Grantha forms of script.

Ex: அ, ஆ; இ, ஈ; உ, ஊ; எ, ஏ; ஒ, ஓ

Native form

ā, ē, ĵ, ĺ, ⊔, ⊕
K, ʧ, ʈ, t, p
ā, ə, ə, ʦ, ʧ, ĺ, ʋ
D, n, n, n, m
ā, i, ʊ, ʊ, ʧ, ʋ, ɿ
J, r, l, ɭ, ɼ

Grantha form

ँ, ॐ, ॐ, ॐ
g, ʤ, d, b
ś, ʂ, ɕ, ɡ, ʃ
s, ʃ, h

Allophones or phonetic variants

To avoid the unwieldy length which may overrun the present article, the acoustic allophones which are though more important and crucial when this present work of interest is concerned are deliberately avoided and only the articulatory allophones being more simple and handy to describe and explain are taken as the main subject of discussion.

There are altogether 91 allophone: 52 being vocoids and 39 being contoids.
SECTION 2 - Vowel and consonant phonemic charts

This page contains information about vowel and consonant phonemic charts. The text is in a language that includes diacritical marks, indicating the use of specific characters to represent phonemes.

A section of the text is shown in a table format, which is likely to be a phonetic chart or a phonemic chart. The table includes columns for symbols, which are presumably phonetic symbols or transcription marks used in the language discussed.

The text also contains a graph or diagram, which is likely to be illustrative of the vowel and consonant phonemic charts mentioned. The diagram appears to be a phonetic chart, possibly showing the relationships between different phonemes or sounds.

Overall, the page provides a detailed look at vowel and consonant phonemes, possibly in a specific language or linguistic context, given the unique characters and symbols used.

The text is visually structured with clear sections, tables, and illustrations, making it easier to follow and understand the phonetic and phonemic information presented.
இருக்கிறது. 1. முதலில் முனிவரின் தொகுதிகளின் விளக்கப்பட்ட புதிய முனிவர் தமிழ் பிரிவு மற்றும் அறிவிகள் விளக்கப்பட்டுள்ளன. 2. முனிவரின் தொகுதிகளின் விளக்கப்பட்ட புதிய முனிவர் தமிழ் பிரிவு மற்றும் அறிவிகள் விளக்கப்பட்டுள்ளன. 3. முனிவரின் தொகுதிகளின் விளக்கப்பட்ட புதிய முனிவர் தமிழ் பிரிவு மற்றும் அறிவிகள் விளக்கப்பட்டுள்ளன.

334
<table>
<thead>
<tr>
<th>தொலை</th>
<th>முன்</th>
<th>முடிவு</th>
</tr>
</thead>
<tbody>
<tr>
<td>16, 17</td>
<td>14, 15</td>
<td>6, 5</td>
</tr>
</tbody>
</table>

குறித்து கூறப்பட்டதுகையே, வருமான பற்பல விளக்கங்கள் 15 பாகம் முழுக்கும் காட்சியில் முன் முடிவுபடுத்தப்பட்டது.

1:  [ʌ]; [ʌ̞]; [q]; [ˈʌ]; [ˈʌ̞]
13:  [ə]
2:  [ɑː]; [ɑ́ː]; [ɔː]; [ˈɔː]
3:  [i]; [iː]; [ɪ]; [ˈɪ]; [ˈiː]
4:  [e]; [eː]; [ɛː]; [ˈɛː]; [ˈeː]
5:  [uː]; [uː]; [uː]; [ˈuː]; [ˈuː]
14:  [iː]; [ɪ]
15:  [u]; [u]
6:  [uː]; [uː]; [ˈuː]; [ˈuː]; [ˈuː]
7:  [ɛ]; [ɛ̞]; [ɛ]; [ˈɛ]; [ˈɛ]
8:  [e]; [eː]; [ɛː]; [ˈɛː]; [ˈeː]
9:  [o]; [oː]; [ˈoː]; [ˈoː]
10:  [oː]; [oː]; [ˈoː]; [ˈoː]; [ˈoː]
11:  [ʌ]
12:  [ʌ̞-ʌ̞]

மரபு மாற்றங்கள்

i) நுட்பம் ஆட்சியமைப்பு

கால் (காற்றுக்), மாற்று (ஏக்கர்),
முனைப்பு (முனைக்)
தமிழ் (தமிழ்), தமிழ் (தமிழ்),
நாய் (னாய்)
அம்பம் (அம்பம்), முனை (முனை)
நாயகம் (நாயகம்), நவீன் (நவீன்)
சின்னம் (சின்னம்), சின்னம் (சின்னம்)
சின்னம் (சின்னம்), சின்னம் (சின்னம்),
சின்னம் (சின்னம்)
ii) தரை அடையாளை

<table>
<thead>
<tr>
<th>அடையாளை</th>
<th>அடையாளை</th>
<th>அடையாளை</th>
<th>அடையாளை</th>
<th>அடையாளை</th>
<th>அடையாளை</th>
<th>அடையாளை</th>
</tr>
</thead>
<tbody>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
<tr>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
<td>கொடும்</td>
</tr>
</tbody>
</table>

iii) பொருளங்கள்

<table>
<thead>
<tr>
<th>பொருள்</th>
<th>பொருள்</th>
<th>பொருள்</th>
<th>பொருள்</th>
<th>பொருள்</th>
<th>பொருள்</th>
<th>பொருள்</th>
</tr>
</thead>
<tbody>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
<tr>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
<td>பொருள்</td>
</tr>
</tbody>
</table>

336
iv) உணர்வு

கெடை (கடை

எடு (எடு

மலநிலை மருத்துவப் பல்கலைக்கழகம்

தமிழ் (தமிழ்

மருத்துவ (மருத்து

மருத்துவ (மருத்து

மருத்து (மருத்து

துருக்க (துரு

தகவல் (தகவல்

முனை (முனை

தலைக்கலைக் கழகம்

மபை (மபை

மணம் (மணம்

மாம் (மாம்

அவள் (அவள்

அவள் (அவள்

மாம் (மாம்

த்தை (த்தை

ஆண்டு

கெடா செடி செடி > கெடா செடி

மாம் செடி > மாம் செடி

அவள் செடி > அவள் செடி

எடு செடி > எடு செடி

பால் + கெடை பால் கெடை விளக்கம் அறுப்பார்க்க திகழ்கின்றது அறுப்பார்க்க அடும்பார்க்க குறுகிய, நாட்கள் நிகழ்த்தப் பால். மேலும் கெடை விளக்கம் எடு எடு பாலங்கள் செயல் செயல் பாலங்கள் குறுகிய ஆண்டு குறுகிய.
Realization of Tamil Syllables Text To Speech Transferring System Using FPGA

T.Jayasankar
Lecturer/ECE
Anna University Trichy
jayasankar@tau.edu.in

Dr.J.Aruptha Vijaya Selvi
Professor
Kings College of Engg
arputha09@gmail.com

Prof .R. Rajendran
Tamil University
Pudukkottai Thanjavur
raj_ushus@yahoo.com

Abstract

Design and development of the functional architecture of Tamil TTS engine using Field Programmable Gate Array (FPGA) is reported. It is a stand alone hardware system without any operating system (OS) to sense, identify and convert the Tamil mono-syllable text to speech output which is important for the visually impaired persons. The proposed model in FPGAs optimises the parameter extraction to perform efficient speech synthesis.

Keywords — Tamil text to speech, FPGA.

1. Introduction

Text to Speech (TTS) Synthesis is an automated encoding process which converts a sequence of symbols (text) conveying linguistic information, into an acoustic waveform (speech). In recent years, text to speech synthesis technologies for different language are growing rapidly.

Speech synthesis based on syllables seems to be a good possibility to enhance the quality of synthesised speech. As far as the production of speech is concerned, syllable is the minimum possible speech segment which can be spoken in isolation.

A Tamil voice system is expected to pave way for developing creative applications to enable users to hear Tamil content in voice form. The key concept behind this voice engine is Text-To-Speech conversion. This conversion uses the method of concatenating the syllables to generate the required words. A system used for this purpose is called a speech synthesizer and can be implemented in software or hardware. Hardware implementation of the above system is achieved by FPGA [Cyclone -II], and Verilog HDL and synthesizer tools. Minimal hardware makes the system to achieve cost effective, compact, very less power consumption and speed.

Introduction to Tamil Languages

Over 65 million people worldwide speak Tamil, the official language of the southern state of Tamil Nadu, and also of Singapore, Sri Lanka and Mauritius. It will be a boom to Tamilian, if the user interface with the system is in Tamil, that too if it is in the form of speech.

Nature Of Tamil Scripts: The basic units of the writing system in Indian languages are characters which are an orthographic representation of speech sounds. A character in Indian language scripts is close to a syllable and can be typically of the form: C*VC*, where C is a consonant and V is a vowel. There is fairly
good correspondence between what is written and what is spoken. Typically there are about 35
consonant and 18 vowel characters. However, in Tamil there are fewer characters than many of the other
Indian languages. There are 13 vowels and 18 consonants characters. Some of the consonants have more
than one pronunciation and in effect there are 41 phones.

Creation of Tamil Speech Database

To build a unit selection voice, typically a small set of letter is recorded by a native Tamil speaker in a
recording studio. The speaker uttered the sentences into a stand mounted microphone placed in front of
her or his. The speech data was recorded at 44 KHz, mono channel at 16 bits per sample. After the
recording it was down sampled to 16 KHz for further processing.

Experimental Setup

The module used the previously stored phonetics sound to reproduce the word. The actual process of this
project is by pressing a key in the PS/2 keyboard which is connected to the ALTERA DE1 board, a
particular Tamil syllable which is pre-recorded in the SD card should be play backed.

![Fig.1. Data flow in TTS system](image)

Here 2 tools are used,

1. Quartus II tool and,

2. NIOS II IDE.

In the Quartus II tool, by using SOPC builder, the hardware components and their connections are
created. By using NIOS, the software coding is developed. SD CARD acts as the pre-recorded component
(recoded format of Tamil syllables).

SD card is inserted into the SD card driver of the DE1 board and the protocol used here is SPI mode.
When the program is started, the contents of SD card are moved to the FPGA and from there it is sent to
audio codec through I2C protocol bus. On the audio codec the Tamil syllables gets processed and play backed.

Input to the system is the scan code of the keyboard which is connected through PS/2 connector of the DE1. By pressing a key, for e.g., by pressing V the scan code of 2A gets compared with the corresponding recorded Tamil syllables. For V the syllable of “va” will be heard and gets displayed.

If the incoming scan code has no recorded syllable means, it will play back the default values. (i.e.,) the scan code gets compared and the particular memory content is played.

Description and Design of the TTS Transferring System

**KEYBOARD:** The PS2 port is a widely supported interface for a keyboard and mouse to communicate with the host. The PS2 port contains two wires for communication purposes. One wire is for data, which is transmitted in a serial stream. The other wire is for the clock information, which specifies when the data is valid and can be retrieved. The information is transmitted as an 11-bit “packet” that contains a start bit (Logic 0) followed by 8 data bits (LSB First), one odd parity bit (odd parity), and a stop bit (Logic 1). Each bit should be read on the falling edge of the clock.

![Timing diagram of a PS2 port](image)

The above waveform represents a one byte transmission from the Keyboard. The keyboard may not generally change its data line on the rising edge of the clock as shown in the diagram. The data line only has to be valid on the falling edge of the clock. The Keyboard will generate the clock. The frequency of the clock signal typically ranges from 20 to 30 KHz. The Least Significant Bit is always sent first.

The system hardware architecture is shown in figure 3 including CPU, UART, tri-state Bridge, ram and I/O controls, which are all reusable. Such design methods not only makes it Modulization, but also greatly reduce the design cycle of the system.

![Hardware architecture](image)
**NIOS II SOFTCORE PROCESSOR:** Nios II is a high performance 32-bit softcore processor. The processor is configured on an Altera Cyclone II FPGA. Custom instructions are added to improve system performance; furthermore, more on-chip rams can be added to improve data processing capacity.

**SD card:** Many applications use a large external storage device, such as a SD card or CF card, to store data. The DE1 board provides the hardware and software needed for SD card access. The size of the SD card should be less or equal to 2GB. Also, it is required to be formatted as FAT (FAT16 or FAT 32) File System in advance. The system requires a 50 MHz clock provided from the board. The SD 1-bit protocol and FAT File System function are all implemented by NIOS II software. The software is stored in the onboard SDRAM memory.

**SDRAM:** In order to store the reference, the 512 kB SRAM module built in the board is used. There are three memory modules on the Altera DE2: a 4 MB Flash memory chip, an 8 MB SDRAM chip and a 512 kB SRAM chip. While the Flash module provides a vast amount of non-volatile storage, it is very slow with respect to the main system clock. The SDRAM chip is very fast and has a very large storage capacity, but it require a very sophisticated controller to be operated. This makes the SRAM chip an obvious choice. Even though it is not the fastest nor the largest, it has ten times the required storage capacity needed for this project, and it is fast enough (since it can perform a read or write operation in less than 20 ns, i.e. a system clock period) so as to avoid any timing issues. Moreover, it is a fairly simple device and can easily control.

**AUDIO CODEC:** In this project the CODEC is used as both test equipment for other modules and as secondary input/output for the audio system. The DE1 board provides high-quality 24-bit audio via the Wolfson WM8731 audio CODEC. The WM8731 is controlled by a serial I2C bus interface, which is connected to pins on the Cyclone II FPGA. Initialization of the CODEC is through a standard I2C (Inter-Integrated Circuit) bus and sound transfer is through a 3-wire bus, which in this project is defined as a standard I2S (Inter-IC Sound) bus.

![Figure 4. Connection diagram of CODEC part of the DE1 board](image)

The figure 4 above shows the structure of the CODEC connection to the control numbers on the CODEC are generated using the NIOS II processor. This chip supports microphone-in, line-in, and line-out ports, with a sample rate adjustable from 8 kHz to 96 kHz. WM8731 contains A/D, D/A modules with a high
sample rate and quantization precision. We will use 8kHz sample rate and 16 bit quantization precision in this design.

In voice acquisition part, since A / D is the serial data output, a serial to parallel data conversion and control of the SRAM Verilog module is needed. Voice report is communicated with CPU via GPIO; different voice is played according to different verification result. GPIO control is done in Nios IDE. Similarly, since the voice broadcast from FLASH are read out in parallel, thus a parallel to serial data conversion Verilog module is needed.

**TFT LCD display:** The 3.6” LCD module is the active matrix colour TFT LCD module. LTPS (Low Temperature Poly Silicon) TFT technology is used and vertical and horizontal drivers are built on the panel. Horizontal scan can be from left to right or from right to left and vertical scan can be from up to down or from Down to up. We developed the LCD Controller and displayed a Tamil text to the TFT LCD. The main CPU is the Nios II processor.

![Figure. 5 TFT interfacing module](image)

The Nios II processor connects to ext_RAM_bus, 3-D accelerator, 7-segment controller, SDRAM controller, TFT-LCD controller, and so on. The ext_RAM_bus module is a tristate Avalon bus bridge that connects the Nios II processor to flash memory and SRAM, which are the instruction memory and data memory, respectively, used to run the Nios II processor.

New Component menu option in SOPC Builder is used to include the LCD Controller module and specify its signals. The signals between the LCD Controller and Avalon bus became the pins of the Nios II module. The slave side of the module is connected to the Nios II processor and the master side is connected to the SDRAM controller.

**Proposed Design Flow Algorithm**

- Initialize and load the acoustic library, which consists all the audio recordings;
- Load the target or input word; initialize the register counter;
- While (target letter match to memory address) do
-   

342
• Wave files play via audio codec from the corresponding memory address;
• when player done = 1 implies audio player has finished playing the wave file;
• /* end the loop*/
• reset the register counter;
• repeat the procedure for all the target letter;

**FIG.6 DESIGN FLOW OF THE TTS**  

**Simulation Results**

The project is simulated using Verilog in Altera tools. Simulation for the sound of ‘a’ (in Tamil).

**Conclusion**

In this paper, the design and development of functional architecture of Tamil text to speech synthesis is described. Different feature streams are experimented with the system that optimises the parameter extraction to perform efficient speech synthesis.

**Future Work**

This work can be extended for concatenating the Tamil syllables and also produce the natural Tamil voice from FPGA based machine.
References
FaceWaves : Tamil Text-To-Speech with Lip Synchronisation for a 2D Computer Generated Face

A.G. Tamilarasan, Madhan Karky

tamil_2206@yahoo.co.in, madhankarky@gmail.com
Department of Computer Science & Engineering
College of Engineering Guindy
Anna University

Abstract
This paper presents a system for converting Tamil text to speech and synchronizes the lip movements of a 2D computer generated face to the generated speech. This paper describes the components of the system such as, grapheme to phoneme converter, a rule-based phoneme-sound selector, wave merger, lip-phoneme synchronizer and face animator. Presenting the grapheme to phoneme algorithm and the phoneme-sound selector rules, this paper presents the results from the grapheme to phoneme converter and rule based-sound selector. The paper concludes with future extensions of the presented work.

Introduction
Tamil language has strict phonetic rules defined many thousand years ago in the ancient grammar definition Tholkaappiyam. These rules have formed the basis of many text to speech systems that exist in Tamil today. This paper proposes a Tamil text-to-speech and lip-synchronization system that takes Tamil text and a 2D computer generated face as input and converts the Text to speech waves and lip synchronize the face to speak the given text along with audio as an animated video.

The motivation behind this work is to provide such a subsystem for a pure text-to-video system where an entire video can be created purely using natural language. Such a system will be of immense use to physically challenged and also enables common people with very less computer knowledge to create and distribute animation video. One major advantage of having such a system would be to transfer such animation over Internet or mobile as a simple text and converting the text to video using a local client.

This paper is organized into four sections. The second section discusses the background and literature closely related to this work. The third section provides the system architecture and discusses the different modules and their functions. The fourth section discusses the results and issues related to text-to-speech conversion and lip synchronization. The fifth section summarizes and concludes this paper and discusses future research directions.

Background
There have been many research works related to Text-to-Speech and generally the text-to-speech systems fall in one of the two categories namely speech synthesis and phoneme concatenation. Sreekanth Majji and Ramakrishnan A. G. used a labeled database to make a system capable of producing a speech output[1]. John Lewis proposed Naive approach for automatic lip- sync to open the mouth in proportion
to the loudness of the sound [3]. Keith Waters and Thomas M. Levergood demonstrated an automatic lip-synchronization algorithm for automatically synchronizing lip motion to a speech synthesizer in real-time[5]. Masatsune Tamura, Shigekazu Kondo, Takashi Masuko, and Takao Kobayashi used parameter generation algorithm, which is used to generate an audio-visual speech parameter vector sequence[4]. Marc Schroder developed Modular Architecture(MARY) for speech synthesis and has the capability of parsing speech synthesis markup.[2]

The Text-to-Speech system proposed in this paper is based on the Tamil Voice Engine [6] proposed by Madhan Karky et al., The Tamil Voice Engine takes Tamil text as input and uses a phoneme database with over 4000 entries. The words are split into corresponding phonemes using rules and the appropriate phonemes are selected and concatenated. This paper uses a similar system for Text-to-Speech but incorporates more rules to improve the phoneme selection and thereby improves the correctness of generated speech.

System Design

The system proposed for text-to-speech and lip synchronization can be naturally split into two major components. The Text-to-speech component receives text as input and converts the graphemes to phonemes. The waves associated with the phonemes are concatenated along with pauses into a single audio wave as continuous speech. The Lip-Sync component of the system receives a 2D computer generated face with all parts of faces and their coordinates. The coordinates of lips are identified and they are modified for different lip positions for different phonemes. The system maintains a lip coordinate index for phonemes that require modifications. Figure 1 presents the overall design of the system.

Text-to-Speech: The Text-to-Speech component comprises of four modules. Text Processor, Grapheme to Phoneme Converter, Rule Based Phoneme Selector and Merger. The component also comprises of a phoneme database. These modules together convert a given text dialogue to speech.

• Text Processor: The Text processor processes raw input text and tags the text to appropriate tags as words and pauses. Inter-word pause, inter-sentence pause are tagged by the text processor. A document is thus converted to a set of ordered word and pause tags.

• Grapheme to Phoneme Converter: Grapheme to Phoneme converter takes the ordered set from Text Processor as input and processes the word tags and splits them further into phonemes. The grapheme to phoneme conversion is based on an algorithm provided in [6] that uses consonant-vowel combinations of letters in a word. An example of such a grapheme to phoneme conversion is provided below.

\text{தமிழ்கு அழகு பெர்} \text{தமிழ்கு அழகு பெர்}

Will be split into the following phoneme combinations separated by hyphen.

• Phoneme Merger: The phoneme database stores 4646 phonemes and based on the split provided by the grapheme to phoneme converter, Phoneme merger applies certain rules to identify the right set of phonemes for the current split.

<tha+mi+zhuk+ku> <a+mu+dhen+dru> <paer>
The selection of ‘dhen’ in the above example instead of ‘then’ is based on the rule that the sound of the phoneme changes based on where it occurs in a sentence. Similar rules are applied for the hard consonants (ka/ga sa/cha da/ta tha/dha pa/ba Ra/tra) based on their occurrences. ‘kutriyalukaram’ a property that states that words that end with ‘u’ have their meter length cut short by half a meter. This rule is also being handled as a rule for selecting appropriate phoneme. Phoneme merger then chooses the appropriate waves from the phoneme database for the phonemes selected by the Phoneme Selector. The selected phonemes are then merged along with the appropriate pauses. The phoneme merger creates a single wave file that can be attached to a video or used for any text-to-speech application. In this framework, it sends the wave to animator and the phoneme and duration information to lip synchronization modules respectively.

**Lip Synchronization:** The Lip Synchronization component comprises of a Face Generator, Face Animator and Lip Synchronizer. These modules together generate a video of a speaking face whose lip movements are synchronized to the speech generated by the Text-to-Speech component.

- **Face Generator:** The face generator module developed as part of the FaceWaves framework, receives textual description of facial features in Tamil. The descriptions are converted to coordinates and dimensions of various parts of the face. The coordinates of the face are used as input for the lip synchronizer.

- **Lip Synchronizer:** Lip synchronizer receives temporal information about the generated wave such as the phonemes, the start and end time of every phoneme and pauses. Lip Synchronizer uses this information to suggest the lip positions for the face animation at every time instance. An example sentence and how it is converted to a phoneme-pause temporal information is given below.

*Table 1: Duration of phonemes*  
*Table 2: Lip Position*

- **Face Animator:** Face Animator uses the temporal information provided by the lip synchronizer to modify the lip coordinates of the face provided by the Face Generator. Face Animator uses the speech wave provided by the Phoneme Merger and creates frames of faces with varying lip coordinates over the same timeline. As the wave timeline and animation frame timeline are one and the same, the animation video has 100% synchronization accuracy.


<table>
<thead>
<tr>
<th>Phoneme</th>
<th>From(Sec)</th>
<th>To(Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>என்</td>
<td>naan</td>
<td>0.0</td>
</tr>
<tr>
<td>Pause</td>
<td>0.446</td>
<td>0.696</td>
</tr>
<tr>
<td>விளிங</td>
<td>ing</td>
<td>0.897</td>
</tr>
<tr>
<td>கு</td>
<td>ku</td>
<td>0.932</td>
</tr>
<tr>
<td>Pause</td>
<td>1.208</td>
<td>1.457</td>
</tr>
<tr>
<td>வன</td>
<td>van</td>
<td>1.458</td>
</tr>
<tr>
<td>என்</td>
<td>thaan</td>
<td>1.708</td>
</tr>
<tr>
<td>Lip Position</td>
<td>From(Sec)</td>
<td>To(Sec)</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>0.445</td>
</tr>
<tr>
<td>1</td>
<td>0.446</td>
<td>0.696</td>
</tr>
<tr>
<td>3</td>
<td>0.697</td>
<td>0.931</td>
</tr>
<tr>
<td>5</td>
<td>0.932</td>
<td>1.207</td>
</tr>
<tr>
<td>1</td>
<td>1.208</td>
<td>1.457</td>
</tr>
<tr>
<td>13</td>
<td>1.458</td>
<td>1.707</td>
</tr>
<tr>
<td>6</td>
<td>1.708</td>
<td>2.083</td>
</tr>
</tbody>
</table>
Results

Text-to-speech system was tested over a dictionary [7] of 2,00,000 root words. Text was converted to speech with 92% accuracy. Words with non-Tamil characters and missing phonemes can be attributed to the 8% results were the right phonemes were not selected. The same system when tested with 200 named entities (person names and organization names) gives 98% accuracy.

![Fig 2: Lip Positions](image)

The Lip Synchronizer and Face Animator use a phoneme to lip-coordinate mapping. The mapping provides information as lip positions. Figure 2 provides the lip positions for various phonemes. The faces in Figure 2 were generated by the Face animator with input from Lip Synchronizer. As mentioned in the previous section the lip synchronization has 100% accuracy as the same timeline is considered for the wave and for the animation.

Conclusion and Future Work

The current system has certain limitations such as the phoneme recordings are not normalized and the merged phonemes are not smoothened. Wave smoothening and professionally recording the sounds and mastering the waves will be essential to improve the quality of the audio. Secondly, the face animation is currently done for a 2D face. If the similar system can be extended to animate 3D faces the animation would look more realistic. This system we believe would be of immense use to common people with very little computer knowledge to create and distribute 2D animations. Creating mobile agents to communicate text messages or creating automatic news reporters with these animated faces will take this project to the next level.
References
Classical Encryption Techniques for Tamil Text

P.Navaneethan  
Dept. of Electrical & 
Electronics Engineering  
PSG College of Technology  
Coimbatore - 641004, India. 
pnn@eee.psgtech.ac.in

C.L.Brindha Devi  
Dept. of Computer Applications  
K S Rangasamy College of Arts & Science  
Tiruchengode-637209, India.  
clbrindhadevi@gmail.com

N. Karthikeyan  
IIInd MCA, Dept. of Maths and 
Computer Applications  
PSG College of Technology, 
Coimbatore-641004  
karthikeyan513@gmail.com

Abstract

This paper describes the security aspects of the Tamil text that is being stored and transmitted over the Internet. The character set in Tamil language shall be categorized into frequently used Tamil characters and infrequently used Tamil characters. The frequently used Tamil characters are divided into consonants, vowels and combined characters. This paper describes how to encrypt the frequently used Tamil characters using a 16-bit Crypto Index. The Crypto Index serves dual purpose, namely, one to find the algorithm to be used and the other to specify the key to encrypt the consonants & vowels.

Keywords: Encryption, Decryption, Substitution, Crypto Index, Rotation, and Mirroring

Introduction

A novel approach for the encryption and decryption of Tamil text is proposed. A 16-bit Crypto Index is used to select the encryption technique and the corresponding key to be used. Though, an arbitrary substitution of Tamil character with yet another set out of 247! ways is feasible, remembering such a substitution based key is impossible, though equally impossible would be the breaking of one such ciphered text. In Tamil text one will not find things like digram, trigram etc. So it eliminates the brute-force technique for cryptanalysis of ciphered Tamil text. In general, when the size of the key is large, it needs to be stored in a media, which makes the whole scheme insecure. The Crypto Index scheme is sufficiently complex, and the key can be kept confidential to oneself.

CV Based Encryption

In this scheme, only those basic characters, namely, consonants and vowels, which are used to derive all the phonetic characters, are encrypted. For example, the Tamil word கூடியாரு can be represented in CV form as கா - ம் கா கூ. A substitution like கா - ம், ம் - கூ, கூ - ம் for the consonants and ம் - பூ, பூ - பூ, பூ - பூ would have encrypted the CV representation into குடிய பு - பூ, and hence, கூடியாரு will be ciphered as குடிய பு - பூ, which means that the ultimate substitution within Tamil alphabet set as கா - ம், ம் - கூ, கூ - ம், ம் - பூ.
Substitution Schemes

a. Rotation Based Substitution [R]

The given text is converted into CV form. Each character in the CV form is rotated ‘k’ places further down the characters in its set, where ‘k’ does the role of a key. For example, consider the vowel character set with the key as 4. Encryption gives the following result.

Position: 0 1 2 3 4 5 6 7 8 9 10 11
Plain Alphabet: அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ
Ciphered Alphabet: உ ஊ எ ஏ ஐ ஒ ஓ ஔ அ ஆ இ ஈ

The mathematical model for this encryption is:

\[ \lambda(x) = (x + k) \mod n \]

where,

- \( x \) -- position of the character to be encrypted.
- \( n \) -- total number of characters in the vowel or consonant character set.
- \( k \) -- takes the value in the range 1 to 11 for vowels and 1 to 17 for consonants.

Decryption gives the following result.

Position: 0 1 2 3 4 5 6 7 8 9 10 11
Ciphered Alphabet: ஔ ஓ ஒ ஐ ஏ எ ஊ உ ஈ இ ஆ அ
Plain Alphabet: ஐ ஒ ஓ ஔ அ ஆ இ ஈ உ ஊ எ

The mathematical model for decryption is:

\[ \lambda^{-1}(x) = (x - k) \mod n \]

b. Mirroring Based Substitution [M]

The mirroring can be either with respect to a single axis or multiple axes. In Single Axis Mirroring technique, the given character set of size ‘n’ is divided into two equal halves by an axis. The character at position ‘n-1’ is replaced by character at position ‘0’ and ‘n-2’ by ‘1’ and ‘n-3’ by ‘2’ etc. For example, consider the vowels set where \( n = 12 \).

\[
\begin{align*}
\text{Position} & : 0 1 2 3 4 5 6 7 8 9 10 11 \\
\text{Plain Alphabet} & : அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ \\
\text{Ciphered Alphabet} & : ஔ ஓ ஒ ஐ ஏ எ ஊ உ ஈ இ ஆ அ
\end{align*}
\]

The mathematical model for Multiple Axes Encryption is:

\[ \gamma(i) = (2j + 1) (n/\text{m})^{-1} - I \]
where, \( i \) lies in the closed interval \([j(n/m), (j+1)(n/m)-1]\), \( n \) is the number of characters in the character set, \( m \) is the number of axes used for mirroring, for \( j = 0, 1, \ldots, (m-1) \).

c. Transposition Based Substitution

In this scheme, the given character set is stored row-wise in an \((r \times m)\) matrix, where \( n = rm \). The key \( \delta \) is a reordering of \([0,1,\ldots,m-1]\), and is used to encrypt the given character set by assigning to each column the corresponding mapping. Each column of the matrix is then read based on the ascending order of the key values which results in the reordering of the character set. The 18 Tamil consonants are arranged as a \((6 \times 3)\) matrix as in Fig.3.1. In case, \( \delta = [1,2,0] \) then, the resulting reordering permutation \( \alpha \) is obtained by reading each column based on the ascending order of the key \( \delta \), i.e., as per \( \delta^{-1} = [2,0,1] \). The numerical value in each cell of the matrix specifies the position of the consonant in the character set, as shown in Fig.3.1.

| Position | : | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| Plain Alphabet | : | ய | ற | ல | ள | ழ | வ | ஶ | ஷ | ஸ | ஹ | ஺ | ஻ | ஼ | ஽ | ஼ | ஽ | ஼ | ஽ |

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>ய</td>
<td>0</td>
<td>ற</td>
<td>2</td>
</tr>
<tr>
<td>ல</td>
<td>3</td>
<td>ள</td>
<td>5</td>
</tr>
<tr>
<td>ழ</td>
<td>6</td>
<td>வ</td>
<td>8</td>
</tr>
<tr>
<td>ஶ</td>
<td>9</td>
<td>ஷ</td>
<td>11</td>
</tr>
<tr>
<td>ஸ</td>
<td>12</td>
<td>ஹ</td>
<td>14</td>
</tr>
<tr>
<td>ஹ</td>
<td>15</td>
<td>஺</td>
<td>17</td>
</tr>
</tbody>
</table>

\[ \alpha_x = \begin{array}{cccccccccccccccccccc}
\ ய & ற & ல & ள & ழ & வ & ஶ & ஷ & ஸ & ஹ & ஺ & ஻ & ஼ & ஽ & ஼ & ஽ & ஼ & ஽ & ஼ & ஽ \\
\end{array} \]

Fig. 3.1 Row-Wise Distribution of Consonants

The above permutation can be grouped into ‘\( m \)’ groups; i.e., in this case as G0,G1,G2.

Here, \( \text{floor}(i/r) \) gives the group id ‘\( x \)’; i.e., \( x \in \{0,1,2, \ldots, m-1\} \), and hence \( \delta^{-1}(x) \) gives the starting value for the reordering associated with the group \( x \). Moreover, \( (i\%r) \) gives a measure of as to how far ‘\( i \)’ is from the start of the group. For example, 14 is 2 units away from the start, namely 12, in G2; this is nothing but 14\%6. Thus, herefore \( \delta^{-1}[\text{floor}(i/r)] + [i \% r]m \) gives the reordering value for ‘\( i \)’ in group \( x \), and the model for encryption is:

\[ \alpha(i) = \delta^{-1}[\text{floor}(i/r)] + [i \% r]m, \]

where \( r = n/m \) is the number of rows, \( n \) total number of elements, \( m \) number of columns

\[ \alpha(14) = \delta^{-1}[\text{floor}(14/6)] + [14 \% 6] \times 3; \text{i.e., } \alpha(14) = \delta^{-1}(2) + 2 \times 3 = 7; \text{ i.e., } \alpha(\ ற ) = \ வ. \]
In order to get $\alpha^{-1}$, the given character set is stored in an $(r \times m)$ matrix column-wise, as shown in Fig. 3.2. Then, the elements in $\delta^{-1}$ are assigned to the columns of the matrix. Each row of the matrix is then read based on the ascending order of the key value, i.e., $\delta^{-1}$, row-wise. The numerical value in each cell of the matrix specifies the position of the consonant in the character set.

\[
\begin{array}{ccc}
2 & 0 & 1 \\
\delta & 0 & 6 & 12 \\
\mu & 1 & 7 & 13 \\
\tilde{\delta} & 2 & 8 & 14 \\
\tilde{\mu} & 3 & 9 & 15 \\
\tilde{\mu} & 4 & 10 & 16 \\
\ddot{\mu} & 5 & 11 & 17
\end{array}
\]

Fig. 3.2 Column-Wise Distribution of Consonants

The mathematical model for decryption is $\alpha^{-1}(i) = \delta \left( i \% m \right) r + \lfloor i / m \rfloor$, where $r = n / m$ is the no. of rows, $n$ total no. of elements, $m$ no. of columns. For the transposition-substituted value 10, its inverse is 15. This is obtained by applying the decryption model; i.e., $\alpha^{-1}(10) = \delta \left( 10 \% 3 \right) 6 + \lfloor 10 / 3 \rfloor = 2 \times 6 + 3 = 15$ i.e., $\alpha^{-1}(\ddot{\mu}) = \ddot{\mu}$.

Similarly, it can be applied to vowels also.

d. Concatenation of Substitutions

As each substitution (one vowel with another or one consonant with another) could be represented as a permutation, these substitutions can be concatenated so as to arrive at another substitution; i.e., Let $\alpha$ stand for transposition based substitution and $\beta$ stand for rotation based substitution. Then, $\alpha \circ \beta$ (read as $\alpha$ composition $\beta$) stands for a new substitution $\gamma$, where rotation is applied first, followed by transposition. Mathematically, if $\gamma = \alpha \circ \beta$, then, $\gamma(i) = \alpha(\beta(i))$. For example, let $\alpha$ represent transposition based substitution, where $\delta$ is $[2,0,1]$ and $\beta$ represent the set of consonants undergoing left shift by 2 positions and vowels by 9 positions.

\[
\begin{array}{ccc}
2 & 0 & 1 \\
\delta & 0 & 6 & 12 \\
\mu & 1 & 7 & 13 \\
\tilde{\delta} & 2 & 8 & 14 \\
\tilde{\mu} & 3 & 9 & 15 \\
\tilde{\mu} & 4 & 10 & 16 \\
\ddot{\mu} & 5 & 11 & 17
\end{array}
\]
Let, \( \gamma_c = \alpha_c \circ \beta_c \) and \( \gamma_v = \alpha_v \circ \beta_v \). Then,

\[
\gamma_c = \begin{pmatrix}
\text{அ} & \text{ஆ} & \text{இ} & \text{ஈ} & \text{உ} & \text{ஊ} & \text{எ} & \text{ஏ} & \text{ஐ} & \text{ஒ} & \text{ஓ} & \text{ஔ}
\end{pmatrix}
\]

\[
\gamma_v = \begin{pmatrix}
\text{அ} & \text{ஆ} & \text{இ} & \text{ஈ} & \text{உ} & \text{ஊ} & \text{எ} & \text{ஏ} & \text{ஐ} & \text{ஒ} & \text{ஓ} & \text{ஔ}
\end{pmatrix}
\]

**e. Crypto Index Scheme**

To encrypt the consonants and vowels, a novel scheme called Crypto Index Scheme is proposed. In this scheme, a 16-bit key is used, which provides indices to the encryption techniques to be employed. The key is grouped into four groups, namely, G1, G2, G3, G4. The bits b0, b1, b2, b3 (G1) are used to encrypt the vowels based on a Rotation [R], the next 5 bits, namely, b4, b5, b6, b7 and b8 (G2) are used to encrypt the consonants based on a Rotation, and the next 6 bits, namely, b9, b10, b11, b12, b13, b14 (G3) are used to go for further concatenation with either Transposition based substitution or Mirroring [M] based substitution, which is applicable to Consonants (C) and/or Vowels (V) based on its value. G4 is used to
specify the type of rotation (left/right). For example, if the key is  0 000111 0001 0001 then, it can be divided into four groups as follows:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>000111</th>
<th>0001</th>
<th>0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>G4</td>
<td>G3</td>
<td>G2</td>
<td>G1</td>
<td></td>
</tr>
<tr>
<td>b15</td>
<td>b14-b9</td>
<td>b8-b4</td>
<td>b3-b0</td>
<td></td>
</tr>
</tbody>
</table>

G1 -- is used to encrypt only Vowels, G2 -- is used to encrypt only Consonants

G3 -- is used for further concatenation of substitutions applicable to Consonants and / or Vowels

G4 -- Type of rotation (Left or Right)

G3 is converted into a radix-4 number, say F. Based on the radix-4 digits, an appropriate algorithm is chosen for encrypting the Vowels and / or Consonants. The radix-4 digits be symbolically denoted as F1, F2, F3. If F1F2F3 is,

123 ◊ 0 missing ◊ Mirroring on Consonants & Vowels,
023 ◊ 1 missing ◊ Transposition of Vowels, 013 ◊ 2 missing ◊ Transposition of Consonants
012 ◊ 3 missing ◊ Transposition of both Consonants & Vowels

After choosing the algorithm, F should contain only radix-4 values. Since it is to be used as a key (δ) for transposition, replace the occurrence of 3 by 2 in F. If G3 has the combination of two 0’s and 1/2/3 then, the algorithm is applied to both the consonants and the vowels. The consonants and vowels are divided into 3 groups and for the group having the \( F_i \) value 1/2/3, mirroring is applied, and for the other 2 groups rotation by 1/2/3 times is applied based on G4. If G3 has 000 then no substitutions are made.

If G3 has the combination of two 1’s and 0/2/3, then the algorithm is applied only to vowels. The character set is divided into 3 groups and if G3 is 111 then each group is rotated 1 time and for the group having the \( F_i \) value 0/2/3, mirroring is applied, and for the other 2 groups rotation by 0/2/3 times is applied based on G4. Here each group is split into 2 parts and while splitting, the 1st part length should be equal to one more than the value 0/2/3.

If G3 has the combination of two 2’s and 0/1/3, then the algorithm is applied only to consonants. If G3 is 2 2 2 then a three 6-Consonants group is formed. Each group is rotated 1 time and for the group having the \( F_i \) value 0/1/3, mirroring is applied, and for the other 2 groups rotation by 0/1/3 times is applied based on G4. Here each group is split into 2 parts and while splitting the 1st part length should be equal to one more than the value 0/1/3.

If G3 has the combination of two 3’s and 0/1/2, then the algorithm is applied to both consonants and vowels. If G3 is 3 3 3 then rotation applied to both Consonants & Vowels group. Each group is rotated 1 time. Here the group having the \( F_i \) value 0/1/2, mirroring is applied, and for the other 2 groups rotation by 0/1/2 times is applied based on G4.

So the Crypto Index Scheme defines 64 (24+40 = 26) different crypto indices for the group G3.

\[ \text{crypto index} \] is encrypted using Crypto Index Scheme as follows. \[ \text{crypto index} \] is first Converted to CV Form as \[ \text{CV Form} \]. If the key is 0 100001 00010 1001 then
G1 = 1001, G2 = 00010, G3 = 100001, G4 = 0. After applying rotation followed by transposition, we get

\[ \begin{align*}
\gamma & = \text{i.e., c on Consonants} & \delta & = \text{v on Vowels. Hence,}
\end{align*} \]

\[ \begin{align*}
\text{The ciphered text for } & \text{is } \text{panditham}. \\
\end{align*} \]

**Conclusion**

This paper has proposed and illustrated a novel scheme for encrypting the Tamil text using Crypto Index Scheme. The frequently used Tamil character set has 247 characters and hence, 247! possible ways of substitutions, where the key (i.e., the arbitrary substitution) cannot be remembered, in general. Hence, methods have been derived to identify substitution schemes based on a 16-bit Crypto Index, the key that can easily be remembered, but, which makes the schemes complex enough. Moreover, Tamil language has no digrams, trigrams etc. as English has. Hence, it eliminates the brute-force attack by a cryptanalyst.

**References**

Abstract

Tamil is slowly gaining popularity as an active language in social networking on the Internet. This paper aims at statistically analyzing the usage of Tamil words in text streaming social networking sites. This paper proposes an active Tamil text stream reader designed to obtain live usage statistics of Tamil words in Twitter, a text based social networking service. A spatio-temporal dynamic index is maintained by the text stream reader and the word usage and the geo-tags are recorded along with a time stamp. The paper also presents a visualization tool where the data captured in the spatio-temporal dynamic index can be visualized graphically to show what topics are gaining and losing popularity over time and space. The results from the analysis are discussed along with snapshots from the visualization tool. The paper concludes with open questions for future research in active text stream analysis for Tamil language.

Introduction

The number of text streaming social networking sites has increased drastically and the usage of Tamil in these sites has reached considerable proportions. The analysis of which will reflect the current scenario of Tamil usage among web users. These text streams can be analyzed to obtain the usage statistics of words over time and geographic location. This analysis requires a tool which will collect data pertaining to a given Tamil word, analyze it then index it. This paper proposes an overall architecture of such a framework.

The proposed framework contains a text stream retriever which interacts with the site in order to obtain the text stream units for analysis. The obtained data is parsed based on various constraints and the information retrieved is stored in an index. This information is then subjected to analysis.

The results from such an analysis reflect various aspects of Tamil’s web usage. When the results are visualized with respect to time, they depict how the usage frequencies of words vary. This data can also be used to find which topics are gaining and loosing importance. When the results are analyzed with respect to geographic location they provide insights into the global scenario of Tamil. In section 2 we provide an overview of the literature survey conducted. In section 3 we discuss the design of the various modules of the framework which provides us an overview of the spatio temporal properties of Tamil text streams. In section 4 we discuss the implementation of the proposed framework and the results obtained from the analysis. Finally we conclude in section 5 with directions for further studies in the field of text stream analysis for Tamil language.
Background

In the literature there are existing works on text stream analysis. Qiankun Zhao et al. explored the temporal and social information of text streams in order to achieve better results in event detection [1]. Jon Kleinberg analyzed the various techniques for Topic Detection and Tracking and also the various information visualization techniques for interpreting data from a temporal dimension [2]. Nilesh Bansal et al. developed a system called BlogScope [3]. BlogScope is an information discovery and text stream analysis system which follows blogs and analyses their content. Danyel Fisher et al. developed a visualization to track narrative events as they develop in text streams [4]. Le Wang et al. proposed a double time window algorithm for conversation extraction in dynamic text message streams [5]. Vagelis Hristidis et al. explored the techniques for extracting useful information from a collection of text streams and proposed a system for keyword search on textual streams [6].

Design of Tamil Text Stream Analyzer

The Tamil Text Stream Analyzer consists of the following major components:

- Text Stream Retriever
- Analyzer
- Indexer

Figure (1) given below depicts the architecture of the framework.

![Tamil Text Stream Analysis Framework](image-url)
a. Text Stream Retriever

This module is responsible for retrieving the text stream usage instances for a given word. The initial input for this module is obtained from a list of predefined popular root words. In order to prevent the overhead of storing a very large search list comprising of both the root words and their derivatives only the root words are stored and their corresponding derivatives are obtained using a generator. The search query is constructed using the OR operator so that the occurrence of even one of the forms is taken into account. Storing of only the root words in order to increase the memory efficiency is an important design decision. The query is targeted towards the particular site’s Search API. The search results returned by the APIs which are generally XML files are in turn handed over to the analyzer module.

b. Analyzer

The analyzer module is responsible for analyzing the raw data retrieved by the reader. The analyzer parses the given XML data in order to extract the required statistics. Predefined tags containing the required information are identified. The total number of entries for the given word is counted. As the amount of usage history maintained by the Social Networking sites is found to vary for each word, this parameter will not suffice for a conclusion to be drawn. So to provide a uniform interpretation, this module also analyses the per day usage for a fixed number of days in the past for all words. It is also possible to record the origin of the text stream unit provided the sites support geotagging. Currently Twitter, a Social networking site supports geotagging by allowing queries based on the origin of a text stream unit.

c. Indexer

As the usage statistics of words are determined for two parameters namely time and geographic location, spatiotemporal indices are used to store the data. The decision to store, search and update a fixed list of words would result in a static application which cannot be scaled. Hence a dynamic approach is chosen, wherein the text stream instances returned for a keyword are parsed to obtain new words whose root words do not exist in the search list currently. Those words are added to the search list and are analyzed in the future searches.

Index Structure: As text streams are analyzed in a spatio temporal manner, two indices are maintained. The temporal index stores the usage statistics of words with respect to time. The spatial index stores the usage statistics of words with respect to geographic location. A Hashtable data structure is used for the indices. The temporal index is built with the temporal component (Date) as the key.

Implementation

In order to implement the proposed framework, Twitter a prominent Social Networking Site is chosen as the source for text stream. The root word to be searched is taken from the index and the word’s derivatives are found using a generator. The search query for the temporal analysis is constructed from those words and is posted to the Twitter Search API. In order to analyze the word usage spatially we count the number of times the word originates inside TamilNadu. The search API responds with XML data. The analyzer module parses this data and populates the statistical and temporal indices maintained. The text stream units are parsed and the words are subjected to morphological analysis and the resulting root words if not already present, are added to the index.
a. Results

As the source of the text streams is a social networking site (Twitter), they contain conversations which have many stop words. These stop words are not excluded from the search. The search for stop words yields many other words which accompany them in conversations. The analysis also shows that a considerable proportion of the text streams originate outside Tamil Nadu. The user interface is designed with options for the users to visualize the results in various aspects. Options for viewing the results for the top 5 words of the week, comparing the usage statistics of the given words and analyzing the spatial usage distribution of a given word are provided. Figure (2) given below is a screen shot of the text stream analysis system. The bar graph depicts the usage statistics of the top 5 words for a time span of 7 days from 18th of April 2010. The line graph below compares the occurrence count for a given word originating within Tamil Nadu against the total occurrence count.

![Screenshot of the Text Stream Analysis System](image)

**Figure 2: Screenshot of the Text Stream Analysis System**

Conclusion and future work

In this paper we have proposed a framework for spatiotemporal analysis of Tamil text streams. The performance of the current system can be improved by adapting a more efficient indexing mechanism. As a next level of analysis, Topic Detection and Tracking (TDT) and prediction of evolving topics can be applied for Tamil text streams. The usage of Tamil in blogs and search engine queries is on the rise. Blogs and search engine queries have inherent temporal properties, hence this analysis can be extended to these areas also. Related work has been carried out for text streams in English. But Tamil being a highly inflectional language, requires customized searching and indexing mechanisms for efficient analysis.
References


Abstract

While reading old documents, it is difficult to read the content if the pages or print found to be in poor condition. The damage might be because of poor maintenance, poor paper quality, prolonged disturbances like blurring of ink and damages done by bookworms like silverfish and booklouse. The objective of this study is to ascertain the distortions on Tamil scripts by various sources. The factors that cause each kind of distortion, the problems associated with them and the possible solutions for the identification of each kind of distorted text have been discussed in detail. This is extremely useful for researchers engaged in recognizing the distorted documents in any script as same kind of distortion can be found in most of the scripts used globally.

Introduction

The progress of any OCR recognition, which registers the efforts of researches of the last six decades, can lead to the accuracy of 99.90% obtained with the help of any commercial OCR system when the document images are sharp, clear and noiseless. Still, there are several applications were the recognition process miserably fails when a poor image source is used. In addition, even a slight distortion of the image quality can make the accuracy of document recognition fall flat. The occurrences of distortions in scanned images are affected by various factors, which are categorized into four areas [1, 2]. Distorted documents do not include all the ideal properties of a document. Even a slight deter in the quality of source documents results in a downfall in the entire recognition process. Some well-known causes for deterioration on document images include:

(a) Natural calamities
(b) Vertical cuts caused paper folding
(c) Usage of poor quality of ink
(d) Excessive dusty noise
(e) Large ink-blobs merging the disjoint characters or components
(f) Disconnection of arbitrary direction due to paper quality or the presence of foreign material
(g) Floating of ink to the opposite pages or the next pages etc.,
Defects caused during printing

Defects introduced during digitization through camera, copying through photocopier and fax machines

General ink-blobs are nothing but the random occurrence of black or white pixels at or about a co-ordinate point. Even the spread of normal distribution (essentially the standard deviation) is also variable. Disconnection of characters is caused by the white blobs whereas black blobs results in merged characters. Keeping above in mind different kinds of distortions have been observed in Tamil document scripts.

Review Of Literature Over Different Type Of Distortion

In order to reach the goal, an ample study of research outcome in several related areas were surveyed.


Heavily Printed: Even advanced OCR of Roman script stands a failure to recognise the heavily printed characters [25].

Double-sided: Leedham et al [26] attempted the recognition process with the introduction of binarization methods with bleed through defects. Anna Tonazzini et al [27] have drawn more general approaches and statistical methods such as Independent Component Analysis (ICA) and Bline Source Operation (BSS). Dubois and Pathak [28] real samples are used for various distortion models. Like [28] more information regarding this can be accessed using [29]. Gang Zi and Doermann [30] and Gang Zi [31] proposed types of defects taking the base of blurring and mixing techniques.

Faxed document: Bloomberg [32] delivered his message as storing the fax as an electronic image. Oguro et al [33] gave a proposal of three-step solution for restoring fax documents. Randolph and Smith [34] comprised of directional components developed a binary Directional Filter Bank (DFB). The recognition accuracy has gained better percentage by the hands of Hobby and Ho [35]. Natarajan et al [36] have trained the system on distorted documents and then adapted to adjust the parameters of the trained model.
Type written document: Cannon et al [37, 38] have advanced an automatic quality improvement technique for recognition of distorted typewritten images. A new cost function was formulated by Rodriguez et al [39] to segment degraded typewritten digits.

Different kinds of distortion in Tamil scripts

A careful scanning of 200 Tamil documents resulted in the following discussion on the causes of each kind of distortion, the problems associated with them and the immediate need of analyzing the solution to overcome those problems that are obstacles of character recognition. The following enlists the various kinds of distortion.

1. Merging Characters (Touching)
2. Fragmented Characters (Broken)
3. Over Imprinted Characters (Heavily Printed)
4. Electronically Shared documents (Fax etc.,)
5. Double-Sided Documents (Bleed Through)
6. Type Written Documents

1. Touching Characters

This is the most commonly found distortion in hand-written and printed Tamil scripts. It happens due to overlapping of parts of two characters in one or more places in different zones. Segmentation is a process involved in recognition. As OCR depends heavily on the accuracy of the segmentation process, each OCR system has to perform well to maximize the result of segmentation. Documents containing touching characters are magazines with heavy printing, hand written documents and Photostatted documents copied on low quality machines.

A careful analysis and investigation help to classify the touching characters into three types.

(a) Single touching
(b) Multiple touching
(c) Long touching

Figure 1: Touching Characters in Tamil Scripts
In the first case, the neighbouring characters touch each other in only one place, in the second case the merging occurs in more than one place and in the third, the characters merge with each other as hardly separable components. The OCR meets with a drastic decrease in accuracy when touching characters are involved.

A statistical analysis enlists the following observations

(a) Probability is more for the merging of characters at the middle zone rather than upper or lower zones.

(b) In most cases, touching characters in Tamil script either closely resemble some other character or totally differ from valid characters.

(c) Most of the images encounter a single black run at the touching position.

(d) Image characters, which are taken for treatment, are generally of two characters. Merging with one another and overlapping of three or more characters takes rare occasions.

(e) The aspect ratio (width divided by height) clearly distinguishes the touching characters from the usual and isolated characters, which are comparatively smaller than that of the merged ones.

(f) In fact, the identification of vertical thick ink blob differentiates the touching characters that possess the abnormal thickness in it.

(g) From a series of analysis of touching characters, it has been noted that generally a word may contain only a pair of characters touching each other.

(h) Possibility of more than two characters touching is very feeble.

(i) Obviously, the characters of many Indian scripts contain a sidebar at the right end, which is absent in Tamil scripts. Hence, ambiguity is comparatively the least in that particular position.

2. Broken Characters

Resembling the touching characters, the broken characters also cause some obstruction in the character recognition of Tamil scripts.

The following observations have been made on the statistical analysis of broken Tamil characters.

Figure 2: Broken Characters in Tamil Script

(a) Broken characters have common occurrence more in lower zones rather than upper and middle zones.

(b) The appearance of a character is generally not similar in shape like other individual characters. In some cases it may be same in shape of some other character.
(c) Despite the distinct shapes of each character, some resemble other, causing complexity of conception.

(d) Generally, there is a heavy loss of information in Tamil when compared with headline-based characters.

(e) Broken characters generally adhere to the aspect ratio (width divided by height) lesser than other single isolated characters.

(f) The segmentation of vertical and horizontal break of the characters has added one more problem of diagonally broken characters.

(g) The improper spacing of character causes overlapping lines producing difficulty in understanding.

3. Heavily Printed Characters

Sometimes isolated characters turn out to be unidentifiable due to heavy printing, equally significant problem like touching characters. So the problem of heavily printed character of Indian language still posses a need for an innovative solution. This kind of distortion takes the same sources as the touching characters.

Observations after statistical analyses are enlisted below:

(a) Generally, the shape of a heavily printed character may look like some other character.

(b) In most of the characters, when printed heavily, there is a gain of a loop in their structure.

(c) Heavy prints cause damages in almost all zones. Even in clean documents heavy printing sometimes heavily influence the character recognition.

(d) Extending the problem, heavy printing leads to merging of characters, which fall into the touching character category.

(e) Unfortunately the heavily printed characters also occupy the same space as the original characters and hence maintain the same aspect ratio.

(f) Practically, it is difficult to extract the features of heavily printed characters due to their likeness with the originals. They take just a blob of the pixels with height and width of the original characters with no ascenders or descendens to help distinguish them.

(g) The reasons of production of heavily printed characters are as same as that of touching characters.
4. Double-Sided Documents

This is a kind of distortion found generally in very old documents were a text on a side is visible on the other side, which is technically called as show-through or bleeds-through problem. This is one of the most challenging one in distortion problems.

![Figure 4: Backside Text Visible In Tamil Script](image)

The following observations have been made on Backside Text visible documents in Tamil script.

(a) This problem is quite common when papers are thin or poor in quality.

(b) Deep dark printing on pages resulting in bleed through.

(c) Due to the partial appearance of backside characters the original characters on the front are misunderstood.

(d) Binarization processes of these documents have registered a lot of noise pixels.

(e) All segmentation processes prove to be failure in front of this distortion at most of the places.

5. Faxed Documents

Fax machines are considered one of the major sources of text distortion, when they create the problems in recognition of their own. This distortion is very apparently visible in the form of spurious point noise and ragged edges. Fax involving light printing, produces a huge number of broken characters, few touching characters and sometimes only a few pixels of characters. Distortions of fax are listed out as salt and pepper noise, the thickening or the partial omission of figures cause inappropriate threshold sensor outputs, which adds random noise and ill balanced bias that should be overcome.

![Figure 5: Faxed Document in Tamil Script](image)
However, generally it is difficult to restore distorted fax images of fax, as almost all the gray scale
information of the output image is lost when using the binarization process. Even the human beings
sometimes do not recognize distorted images.

The observation read as the list follows:

(a) The width of the stroke is not constant over the document.
(b) Entire document is found with all types of distortions at all the zones.
(c) The quality of the fax document depends upon the fax machine.

6. Type Written Documents

Typewritten documents also encounter the problem of distortion. Typewriters are widely used in
Government offices of India. As far as Tamil scripts are concerned, 80% of official works are processed
by means of type written documents only.

![Typewritten Document in Tamil Script](image)

Figure 6: Typewritten Document in Tamil Script

A sequence of statistical analysis of typewritten characters has brought out the following observation.

(a) The middle zone is terribly influenced by the distortion in the form of merged characters.
(b) Lower zone characters, almost every time merge with the previous upper zone characters
complicating the segmentation of lower zone from the middle zone.
(c) Unequal spacing between lines, words and characters has been observed.
(d) There is a significant change in the shape of upper zone and lower zone characters.
(e) Some times the characters are broken into many parts placed at various base lines. Usually most
of the difficult algorithms have been designed on the basis of headline and baseline.
(f) Extra force applied during typing leads to character distortion of heavy printing.
(g) The technical inability of the typewriter with fixed width or variable grid produces characters of
the same horizontal shape irrespective of the actual shape of the characters.

Conclusion

On the whole, this study has shouldered the task of venturing into the treatment of distorted characters of
Tamil scripts. Aiming at the maximization of the recovery rate of damaged or dilapidated documents of
various Indian languages with special highlights on Tamil scripts was covered under different headings
viz., touching characters, broken characters, heavily printed documents, faxed documents, typewritten
documents and the like. This study not only probes into the character recognition process that deals with
so many problems of distorted Tamil characters, the kinds of distortions on it and the possible solutions
to overcome those stumbling blocks, but also wide spreads ample of opportunities and scope for further researches in the same field of character recognition. Though researches enlist a range of character distorters like spray marks, curved base lines, blurred images, presence of punctuation marks and so on, this study pacts the treatment of only a handful initiation to the researches in the field of character recognition of Indian scripts in general and Tamil scripts in particular.

Reference:


A High Accuracy
Phone Recognition System for Tamil

C. P. Santhosh Kumar
Amrita Vishwa Vidyapeetham, Coimbatore
email: cs_kumar@cb.amrita.edu

N. Deiva Sundaram
Madras University, Chennai

Abstract
Phone recognition systems are used in many applications such as automatic segmentation of speech, keyword spotting, automatic language identification using phonotactic approach, and speaker identification. Phone recognition accuracy and the accuracy of the application developed are highly correlated. In this paper, we present the details of a high accuracy phone recognition system developed for Tamil. We use a hybrid hidden Markov model – neural network to implement the decoder. It was seen that for moderate data sizes, the system outperforms the hidden Markov model – Gaussian mixture model implementations.

Introduction
High accuracy phone recognition systems are very useful for many applications. They can be helpful for phonetic labeling of speech; speech data collected with the word level transcriptions without any segmentation information can be converted to phonetic transcriptions with segmentation information. If this segmentation information has to be generated manually, we will need an experienced phonetician, and doing this manually is not cost effective. Further, due to the co-articulation effects, the segmentation boundaries thus obtained will not be precise. Under such conditions, a consistent phone boundary is the best choice, and a machine assisted segmentation and labeling using a phone recognizer can be more effective and cost effective. Phone recognition systems are also used in automatic keyword spotting systems using phone lattices, language identification using phonotactic approach, and speaker identification. Performance of these systems is highly correlated with the phone recognition accuracy and therefore any effort to enhance the phone recognition accuracy is likely to enhance the performance of the application.

Hidden Markov model –Gaussian mixture models (HMM-GMM) are known for their performance in the development of speech recognition systems, and in such systems Gaussian mixture models are used to model the probability distributions. In speech recognition systems, word n-gram language models and phone n-gram language models can be used to enhance the recognition accuracy. However, for moderate size databases hidden Markov model – neural network structures[1,2,3,4] have been found to offer better performance compared to the HMM-GMM systems, and for large databases both offer similar performances, but HMM-NN systems has less complexity.

Segments of speech (frames) represented in terms of their probabilistic similarity to the phones of any language are often referred as probabilistic features (PF)[1, 2]. Temporal patterns (TRAPS)[1] of log energy from critical bands is a way to derive PFs. In [1, 2], several band-conditioned classifiers are used to derive
PF to be merged by a neural network to estimate the phone posteriors. The use of band conditioned phone posteriors as temporal features was studied in [3]. In this work, they used the hidden layer activation outputs, and is known as hidden activation TRAPS (HATS). In [4], a simplified system offering better results, and requiring less training data due to the splitting of TRAPS features to left and right context has become popular since [4] due to its simplicity and ability to reduce the amount of training data required for similar performances. In this work, we use the implementation in [4] to implement a high accuracy phone recognizer for Tamil.

Keyword spotting systems (KWS)[5,6] are useful for archiving and indexing audio/video documents to be searched using keywords, searching for telephonic conversations that contain words/phrases that are potentially dangerous to the national security, and thereby derive intelligence information. Large vocabulary continuous speech recognition (LVCSR) system based KWS are very popular for their better performance. However, they have many limitations to be applied directly to the Indian environment. To develop an LVCSR system for a language, we need a large labeled speech database to train the acoustic models, and to the best of our knowledge, such a database is not yet available for Tamil. In the Indian context, we tend to use words/phrases across languages, and is fairly common to mix words/phrases from English in native language conversations. LVCSR systems cannot handle multiple languages at the same time, as it needs a language model to enhance the word recognition accuracy, and training a language model catering for multiple languages is practically impossible for many reasons. In this work, we present the details of a KWS systems developed using the lattices [5] generated by the HMM-NN phone recognizer. Another approach that is widely popular uses lattices generated by a phone recognizer [5, 6]. This approach has the advantage that it can be easily ported to multilingual environments, unlike the LVCSR based approach. [5] gives a comparison of different KWS approaches.

**HMM-NN Phone Recognizer**

In the hybrid HMM-NN system, critical band energies are obtained in the conventional way [1, 2, 3, 4], Speech signal is divided into 25 ms long frames with 10 ms shift. The Mel filter-bank is emulated by triangular weighting of FFT-derived short-term spectrum to obtain short-term critical-band logarithmic spectral densities. TRAPS feature vector describes a segment of temporal evolution of critical band spectral densities within a single critical band. The central point is the current frame and 15 frames from the past make a left context (LC) feature and similarly 15 frames from the future make the right context (RC) feature vector. Further, the LC and RC features are windowed to make the transition between the successive frames smooth. Triangular window was used for this purpose. Subsequently, these feature vectors are processed for dimensionality reduction. We used discrete Cosine Transform (DCT) for its simplicity to reduce the 16 dimensional LC and RC feature sizes to 11 [1], [2], [3], [4]. To further enhance the accuracy of the system, we concatenated 15 critical band features for LC and RC to generate input to two separate LC and RC neural networks. Outputs of these classifiers are subsequently merged together using another neural net. Outputs of all neural networks represent phone state posterior probabilities, and phone models have three states each. Details of the implementation can be found in [3], [4] Fig. 1 illustrates the schematic diagram of the implementation of the LC-RC HMM-NN phone recognizer.
Table 1 – Tamil Vowels used in the recognizer with illustrating examples

<table>
<thead>
<tr>
<th>Phone (ASCII)</th>
<th>Alphabet – Example Transcription</th>
<th>Phone (ASCII)</th>
<th>Alphabet – Example Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>ax</td>
<td>அ – அக்கிரமம்&lt;br&gt;/ax m h m aa/</td>
<td>ay</td>
<td>ஏ – ஏல்லு&lt;br&gt;/ae t h l U/</td>
</tr>
<tr>
<td>aa</td>
<td>ஏ – ஏணை&lt;br&gt;/aa d U/</td>
<td>ey</td>
<td>ஏ – ஏணை&lt;br&gt;/ey nn ih/</td>
</tr>
<tr>
<td>ih</td>
<td>ஐ – ஐணை&lt;br&gt;/ih l ax y/</td>
<td>ao</td>
<td>ஐ – ஐணை&lt;br&gt;/ao n dr U/</td>
</tr>
<tr>
<td>iy</td>
<td>ஐ – ஐணை&lt;br&gt;/iy hh ax y/</td>
<td>ow</td>
<td>ஐ – ஐணை&lt;br&gt;/ow m/</td>
</tr>
<tr>
<td>uh</td>
<td>ஐ – ஐணை&lt;br&gt;/uh nn ax y/</td>
<td>aw</td>
<td>ஐ – ஐணை&lt;br&gt;/aw v ax y/</td>
</tr>
<tr>
<td>U</td>
<td>ஐ – ஐணை&lt;br&gt;/ix ax p h p U/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uw</td>
<td>ஐ – ஐணை&lt;br&gt;/uw r/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2 – Tamil consonants used in the recognizer with illustrating examples

<table>
<thead>
<tr>
<th>Phone (ASCII)</th>
<th>Alphabet – Example Transcription</th>
<th>Phone (ASCII)</th>
<th>Alphabet – Example Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>/k ax l v ih/</td>
<td>dh</td>
<td>/p ox nd dh ax n/</td>
</tr>
<tr>
<td>g</td>
<td>/p ax ng g/</td>
<td>nd</td>
<td>/p ax nd dh U/</td>
</tr>
<tr>
<td>ng</td>
<td>/tx ax ng g ax m/</td>
<td>p</td>
<td>/p aa d ax m/</td>
</tr>
<tr>
<td>ch</td>
<td>/k aa t ch ih/</td>
<td>b</td>
<td>/ax m b U/</td>
</tr>
<tr>
<td>jh</td>
<td>/j aa tx ih/</td>
<td>p</td>
<td>/l aa P ax m/</td>
</tr>
<tr>
<td>ny</td>
<td>/ny aa n ax m/</td>
<td>m</td>
<td>/k ax r uh m b U/</td>
</tr>
<tr>
<td>t</td>
<td>/ax t ey/</td>
<td>y</td>
<td>/y ax m uh n ax y/</td>
</tr>
<tr>
<td>d</td>
<td>/k ax d ax n/</td>
<td>r</td>
<td>/m ax r ax m/</td>
</tr>
<tr>
<td>nn</td>
<td>/m ax mn ih/</td>
<td>tr</td>
<td>/ao t tr ax n/</td>
</tr>
<tr>
<td>ddn</td>
<td>/k ax ddn/</td>
<td>dr</td>
<td>/k ax n dr U/</td>
</tr>
<tr>
<td>th</td>
<td>/th ax t h t U/</td>
<td>rr</td>
<td>/ax rr ax m/</td>
</tr>
<tr>
<td>Th</td>
<td>/m ax Th ih/</td>
<td>l</td>
<td>/m aa l ax y/</td>
</tr>
<tr>
<td>h</td>
<td>/p ah t h t ah m/</td>
<td>ll</td>
<td>/uh ll b ll ax m/</td>
</tr>
<tr>
<td>sh</td>
<td>/v ih sh ax m/</td>
<td>lzh</td>
<td>/p ax lzh ax m/</td>
</tr>
<tr>
<td>s</td>
<td>/s ax ng g U/</td>
<td>v</td>
<td>/v ax y ax dh U/</td>
</tr>
<tr>
<td>hh</td>
<td>/p ax hh ax l/</td>
<td>n</td>
<td>/nd aa n/</td>
</tr>
</tbody>
</table>
Lattice based KWS system

Lattices [7] are a way of storing output of a recognizer as an oriented acyclic graph in which each node represents a symbol (phone) and each link represents time boundaries of the phone/word at the end of the link. Fig. 2 shows an example lattice generated by a phone recognizer. It may be seen that at any instant in time, there are multiple possibilities available when the keywords are to be searched. Searching in lattices provides better results than searching in phone strings [5] as it holds several hypotheses in parallel. We follow a lattice based approach in this work [5,6]. Phone lattices were generated from phone posteriori probabilities. In our experiments phone insertion penalty was set to zero for lattice generation to minimize deletion of phones. Fig. 3 shows the schematic representation of the KWS system implemented in this paper, and the likelihood of the keywords are compared with the likelihoods of the hypothesis generated by the background filler phone models working in a parallel loop to decide the presence or absence of a keyword in a continuously spoken sentence.

Fig. 2 – An example of a lattice generated by the phone recognizer

Fig. 3 – A Schematic description of a KWS system

Experiments and Results

We used 1.2 hours of telephone quality speech sampled at 8 kHz for our experiments. In the HMM-NN systems, we used 300 neurons for modeling the probability distribution. For the HMM-GMM systems, we used 16 Gaussians to model the probability distributions, and the models are trained using maximum mutual information criteria using the HTK toolkit [7]. The size of the neural network and the GMMs was chosen to match the limited training data available.
KWS systems are evaluated using Figure-of-Merit (FOM) [6], which is the average of correct detections per $1, 2, \ldots, 10$ false alarms per hour. We used the most frequently occurring 60 words to benchmark our system. The KWS system developed in this using the HMM-NN phone recognizer gave an FOM of 66.25 per cent for the most frequently used 60 words in the test set.

References
7. CUED, HTK toolkit, http://htk.eng.cam.ac.uk/
Face Waves: 2D - Facial Expressions Based on Tamil Emotion Descriptors

Sabitha.Tammaneni, Madhan Karky
tammaneni@yahoo.com, madhankarky@gmail.com
Department of Computer Science & Engineering
College of Engineering Guindy
Anna University

Abstract

This paper aims at recognizing human emotions from textual descriptors and expressing the emotion in a 2 dimensional computer generated face for the Face Waves framework. The main objective of this paper is to map facial expressions for the basic six emotions (Love, Joy, Fear, Anger, Surprise, and Sadness), 27 second level derived emotions, and a third level 79 emotions. Apart from hierarchically classifying these 112 emotions in Tamil, this paper presents an object oriented emotion model, where the emotions at lower levels inherit properties from emotions at higher levels. The paper also describes an 'expression index' built for efficiently storing and retrieving the facial features for corresponding expressions for every emotion. The 'expression index' plays a vital role in mapping the expression on a computer-generated face with time efficiency. Discussing the results of mapping and efficiency of the index, the paper concludes with open questions and future enhancements.

Introduction

Since Charles Darwin’s early work on ‘The Expression of Emotions in Man and Animals’, there have been various proposals to classify human emotions. A few researches have been carried out in mapping expressions to emotions. The expressions to emotion mapping cannot be generalized as every individual has his or her own way of expressing various emotions. In many cases it will not be possible for one to distinguish extreme happiness from extreme sadness just by observing facial expression of a person. In this paper we present a system that can identify emotions from textual descriptions and try to map a corresponding expression on a 2D computer generated face.

The primary aim of this paper is to present an object model for all possible human emotions and their corresponding Tamil textual descriptors. The secondary aim is to map the action units of various parts of a human face to express the emotion. To best of our knowledge, this is the first time that such a work is presented in Tamil and first time an object-oriented model is proposed for representing emotions and their expressions.

This paper is organized into five sections. The second section provides some background information and discusses literature related to this work. The third section presents the classification tree for human emotions in Tamil. The fourth section describes the system design and explains the action units for every expression. The fifth section presents the results and discusses the advantages of an expression index based on the object-oriented emotion tree to increase the time efficiency of expressing emotions on a 2D face for a text-to-video system.
Background

“Facial expression” has drawn interest to a few researchers around the world. Most of the research focus on expressing emotions in a 3D face. Paul Ekman and Wallace F introduced a system called FACS, where each facial expression is described in terms of Action Units (AU’s) [1]. Irfan A. Essa and Alex P. Pentland implemented a computer vision system which is developed by using mathematical formulation which is used for detailed analysis of facial expressions[3]. Prem Kalra, Angelo Mangili, Nadia Magnenat Thalmann and Daniel Thalmann made every expression from one or more MPA’s (Minimum Perceptible Action). One or several simulated muscle actions constitute a MPA [2]. Praseeda Lekshmi V And Dr. M. Sasikumar used multicast Support Vector Machine for Classification of different kinds of facial expressions belonging to the face image they also used Gabor filters for image processing[4]. Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi developed a deformable muscle-based face model that tracks some FCP’s in the real face image sequences and shows the same expressions[5]. Robert Plutchik created a wheel of emotions, which forms the base to this paper[6]. The emotions provided in English are translated in Tamil to present the emotion tree.

Classification of Human Emotions

As explained in section 1, there have been various proposals to classify human emotions. Many classifications do not recognize love as a basic human emotion. We have achieved our classification of emotions based on Robert Plutchik’s classification[6]. The first level of tree is provided in figure 1. Happiness, Fear, Love, Sadness, Anger and Surprise form the first level of emotions. All other emotions are derived from these basic emotions according to Robert.

![Figure 1: Emotion Tree: Level 1](image)

Each of these basic emotions in the first level of tree has sub-emotions as shown in figure 2. Figure 2 gives the second level emotions for two of the basic emotions happiness and sadness.

![Figure 2: Emotion Tree: Level 2](image)
The classification of these emotions runs down to a third level where each of the sub-emotions have sub-sub-emotions with the same base emotion. Figure 3 gives the third level emotions for happiness and sadness. The sub emotion shame(avamaanam) gives raise to four third level emotions. Similarly the sub-emotion cheerfulness(uRsaagam) gives raise to eight third level emotions.

![Figure 3: Emotion Tree: Level 3](image)

The complete tree consisting of 112 emotions is not provided in this paper owing to space considerations.

**System Design**

![Figure 4: Emotion Processor System Design](image)
The design of our emotion processor is given in figure 4. Face Generator will accept the input (face description) from the user, which will analyze the input and then distributes the corresponding characteristics for each part of the face. The character Engine will construct the Human face, which is a 2D computer generated face.

The Text Processor processes a given text document for emotional descriptors. The text lines corresponding to the emotional descriptors are analyzed using a morphological analyzer and the output of the analyzer is fed to the Emotion Identifier. Emotion identifier using the emotion model explained in section 3, identifies the emotion and sends the emotion tag to Face Modifier. The Face Modifier module uses an expression index to modify every part of the face according to the expression. Each part of face is modeled as an object and the face modifier based on the emotion tag controls the attributes of the object. Thus the emotion is mapped to the face that comes as input from the character engine as a set of feature changes termed as an expression of the emotion.

The modified face can now be stored as text in a database for efficient retrieval for the Text-To-Video system. It is to be noted that the Face Generator may generate any type of a face based on the descriptions it gets. The pre-generated face is given as input to the Face modifier along with the emotion descriptors.

**Emotion to Expression**

We use Ekman’s FACS Action Units(AU)[11] to describe the transformations on a 2D computer generated face. The action units specify pull, raise like actions for every part of the face. Those Aus corresponding to 2D face alone have been used to represent the emotion on the face as an expression. The following are examples of two basic emotions and their corresponding action units.

"""(Joy):

<table>
<thead>
<tr>
<th>AU</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Lip corner puller</td>
</tr>
<tr>
<td>5</td>
<td>Upper lid raiser</td>
</tr>
</tbody>
</table>

*Table 1: Action Units for Joy*

"""(Fear):

<table>
<thead>
<tr>
<th>AU</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inner brow raiser</td>
</tr>
<tr>
<td>2</td>
<td>Outer brow raiser</td>
</tr>
<tr>
<td>5</td>
<td>Upper lid raiser</td>
</tr>
<tr>
<td>7</td>
<td>Lid tightener</td>
</tr>
<tr>
<td>20</td>
<td>Lip stretcher</td>
</tr>
</tbody>
</table>

*Table 2: Action Units for fear*
Emotions in the second and third levels of the emotion tree inherit the action units from their parent emotion and have their own action units along with the inherited units.

**Results**

A simple face generated by the Face Generator module and a few samples of the faces modified by our Face Modifier for different emotions are given in figure 5. The features of the face are given as input to the Face Modifier. Emotion descriptors can be natural language text describing the state of the person.

He looked sad

He looked scared at that moment

The above given sentences can be examples of emotion descriptors. Our analyzer breaks the words to identify the root and when the emotion is identified, the corresponding features are retrieved from the expression index. The features are applied over the 2D face that was received as input for modification.

![Figure 5: Original Face and Modified Faces](image)

The face in the center of figure 5 is the original face generated by our face generator surrounded by the corresponding faces modified for six different emotions. Similarly any generated face is modified for one of the 112 emotional descriptors.
Conclusion and Future Work

This paper proposes an object-oriented model for representing emotions hierarchically and proposes a system for applying emotions on a 2D computer generated face using Ekman’s FACS action units. Improving the index for minimizing the number of changes from original face to modified face will be our immediate research that follows. Such an index will improve the time efficiency of creating text-to-video. Applying these emotions for a 3D face can aid to generating 3D animation videos.

References


6

தந்து மிள்ள துவ முழந்த
பிள்ளர் கருத்தில்
மன்னர்கள் பல்கலைக்கழக மிள்ளியில் தொடர்வது

Ilankumaran S, O Sivanadhan
s.ilangkumaran@gmail.com

மேலிருந்து உலகம் குரலில் குரலில் முடிவு அளியும் தரவில் பனையார் இந்தியாவின் ஒரு பகுதியுடன் சேர்ந்து வருவதற்கேண்டு மிள்ளியில் தொடர்வது. போற்றும் மொழிக் கட்டுரையில் மேலிருந்து உலகம் குரலில் குரலில் முடிவு அளியும் தரவில் பனையார் இந்தியாவின் ஒரு பகுதியுடன் சேர்ந்து வருவதற்கேண்டு மிள்ளியில் தொடர்வது.

மேலிருந்து உலகம் குரலில் குரலில் முடிவு அளியும் தரவில் பனையார் இந்தியாவின் ஒரு பகுதியுடன் சேர்ந்து வருவதற்கேண்டு மிள்ளியில் தொடர்வது. UNESCO அறிக்கட்டிய கால்களின் தொடர்வண்டி பல்கலைக்கழகத் தொடர்வண்டி மிள்ளியில் தொடர்வது. இது தொடர்வண்டிகள் மறுசூழ்த்தப்பட்டு செய்யப்படும் விளக்கத்தை உறுதிபடுத்தும்:

• அன்புறுத்தான் பல்கலைக்கழக கல்லறிவன் பெண்களின் விளக்கத்தை
• கல்லறிவாளர் எழுந்து முக்கியத்துவமான இல்லாமலைகளை
• கல்லறிவாளர், மேலிருந்து உலகம் குரலில் குரலில் முடிவு அளியும் தரவில் பனையார் இந்தியாவின் ஒரு பகுதியுடன் சேர்ந்து வருவதற்கேண்டு மிள்ளியில் தொடர்வது.
• இலங்கையின் அழக்கம் மேலிருந்து உலகம் குரலில் குரலில் முடிவு அளியும் தரவில் பனையார் இந்தியாவின் ஒரு பகுதியுடன் சேர்ந்து வருவதற்கேண்டு மிள்ளியில் தொடர்வது.
பலவழியான பயிற்சிவழியாட்டு நடனத்தில் நடன அடர்த்தியை முறைப்படுத்தும் வழிகாட்டி வாய்ந்ததைக் கொண்டு நடப்பதற்கு ஒப்பிட்டுப் பெறுவது தவறாதது. அந்தவழியான பயிற்சிவழியாட்டு நடனத்தில் நடன அடர்த்தியை முறைப்படுத்தும் வழிகாட்டி வாய்ந்ததைக் கொண்டு நடப்பதற்கு ஒப்பிட்டுப் பெறுவது தவறாதது. அந்தவழியான பயிற்சிவழியாட்டு நடனத்தில் நடன அடர்த்தியை முறைப்படுத்தும் வழிகாட்டி வாய்ந்ததைக் கொண்டு நடப்பதற்கு ஒப்பிட்டுப் பெறுவது தவறாதது.
• இன்றி மேல் இந்த தமிழ் பொருள், தமிழ் இந்த புதுச்சொல், இந்த புதுச்சொல்
• புதுச்சொல் தமிழ் பொருள், அந்தை தமிழ், தமிழ் தமிழ், புதுச்சொல் தமிழ் தினமுள்ள
• புதுச்சொல் தமிழ் கிளமுடன் - இரண்டு இரண்டு அுபாயம்
• இந்த பொருள், தமிழ் தமிழ், இந்த தமிழ் தமிழ், அந்தை
• அந்தை தமிழ், தமிழ் தமிழ், புதுச்சொல்
• இன்றி மேல் இந்த தமிழ்

பல பல, பல பல, பல பல, பல பல

மேல் புதுச்சொல் புதுச்சொல் புதுச்சொல் புதுச்சொல் புதுச்சொல் புதுச்சொல்

புதுச்சொல் புதுச்சொல் புதுச்சொல்

மேல் புதுச்சொல் புதுச்சொல் புதுச்சொல்

புதுச்சொல் புதுச்சொல் புதுச்சொல்
தமிழ் விளக்கம்

• என்னும் பொருளாயியாக நம்பப்படும் விளக்கம் முழுமையான விளக்கம் என்று கூறுவதும் முடியுமாது. தொடர்பான வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது. அடுத்து வரலாற்றுச் செயற்கையின் பொருளின் முதல் வரையறுக்கப்பட்டது.

390
• நேரடி சாதமுந்த வல்லாயும் போன்றவர், கல்வி வருவாயும் வாழ்க்கை காரணம் இந்தக் கூற்று அறிவுக் கிழக்கையிலும், மூலச்செயல் வரவு விளையாடாது வந்தது. இது விளையாட்டு, பசின்செயல் காலம் கண்டவராகும். நேரடி சாதமும் வல்லாயும் போன்ற வலுவாசரங்களுக்கு புதுக்காண்டு புனிதத்தேச என்பதால் வருவாயும் நேரடி சாதமும் வல்லாயும் போன்றவரே காரணம் விளையாடவுள்ளது. பசின்செயல் வரவு விளையாட்டு நேரடி சாதமும் வல்லாயும் போன்றவர்களுக்கு புதுக்காண்டு புனிதத்தேச என்பதால் வருவாயும் நேரடி சாதமும் வல்லாயும் போன்றவரே காரணம் விளையாடவுள்ளது. பசின்செயல் வரவு விளையாட்டு நேரடி சாதமும் வல்லாயும் போன்றவர்களுக்கு புதுக்காண்டு புனிதத்தேச என்பதால் வருவாயும் நேரடி சாதமும் வல்லாயும் போன்றவரே காரணம் விளையாடவுள்ளது.
Development of Tamil Digital Libraries: Advances and Challenges

Dr. K. Kalyanasundaram
Lausanne, Switzerland

Introduction

Computers and Internet possibly are the two most important technological innovations of the last century that had a pronounced impact on the humanity. Together they have changed dramatically the way people store information and exchange them with others. On the educational front, the impact in the Indian sub-continent is more pronounced. Ways and means by which information (knowledge?) has been stored and transferred from generation to generation has been evolving over many centuries: from “gurukulam” system [1] (direct largely vocal transmission of knowledge from the teacher to the student) to inscriptions in caves, copper plates to written texts as in palm-leaf manuscripts to printed books. The contents of 26-volume summary - a systematic collection of inscriptions of south India compiled by Prof. Hultzsch, published by the Archeological Survey of India is available online [2]. At the dawn of the 21st C, information storage and exchange is moving to electronic form. Net-based information interchange available widely and at low cost has shrunk the distance barriers that exist amongst friends and relatives living in far off places. Growing number of educational tools are produced in electronic form as e-books, multimedia-based teaching software and even distance/remote education through online portals. In this paper we focus on one important component in distance education and information sharing through the net, viz., digital libraries with exclusive focus for Tamil language.

For any society or community, its cultural heritage is measured by the variety and level of content available in key areas such as literature and performing arts (music, dance and drama). Tamil is one of the oldest and living languages of the Indian sub-continent, with a vibrant history that dates back at least to 2000 years. Tamils have a remarkable cultural heritage, evidenced by huge and rich repertoire in all the key areas mentioned above. Preservation and propagation of this rich cultural heritage in a form compatible with contemporary trend and usage mode is essential. For many reasons, Tamil heritage information is largely preserved in forms that are at great risk. There are thousands of works in science and literature still in palm leaf manuscripts, not yet published in printed book form. With poor storage conditions, existing collection of palm leaf manuscripts and printed books is degrading rapidly. In the case of Sri Lanka, huge and precious collections of literary works of Tamils for several centuries simply vanished when the Jaffna Library got burnt in a major fire incident [3]. Anthropologists are concerned about rapid changes in the life-style of Tamils with a rapid decline in the practice of rich “folklore arts” that are unique. Even in areas such as archeology, Tamils are proud to have a vast collection of temples of varying sizes and artistic features, each with its own unique mural paintings and other treasures [4]. Sadly these are also not being preserved in a form that can guarantee posterity. In view of this Tamil Diaspora have an important moral obligation to start e-preservation efforts in the areas outlined above.
Objectives of digital libraries

It is useful to start with the simple question, “what is a digital library”. Widely accepted definition of a digital library is one of “a comprehensive « networked » information environment, a seamless rich set of tools and resources for the user community, accessible round the clock across the globe with the power of the Net. Digital resources can be in different formats and for different purposes, different target audience(s), accessible online or offline and derived from various « primary, existing » resources in different formats (books, audio/video recordings, manuscripts and photographs). Of late many information sources are produced directly in digital or e-form.

Digital Catalogue of Tamil Materials and Digital Dictionaries

A first step in any digitization effort is assessment (stock-taking) of diverse form of cultural heritage materials that are still available, a catalogue of “who has what and where?” While a catalogue of the library holdings in digital form “searchable” remotely is routine and exists for nearly all the libraries in the west, the same cannot be said about major libraries of Tamilnadu. There is no central inventory of public and private collections of cultural heritage related materials of Tamils within and outside India. Tamil Diaspora has migrated in large numbers to several far-off places (e.g., Sri Lanka, Malaysia, Singapore, Fiji and Mauritius to name the major ones) and developed their own local “identity” over the years.

Academic Institutions of North America and Western Europe have been working together to build “union catalogue”, a central inventory of “who has what and where?” Many countries are building “union catalogue” of regional collections. Two notable ones are i) CRL (Center for Research Libraries, an international consortium of university, college, and independent research libraries [5]; ii) “Worldcat”, sponsored and run by OCLC (Online Computer Library) a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world’s information and reducing the rate of rise of library costs [6]. More than 72,000 libraries in 171 countries and territories around the world use OCLC services. Nearly 1.5 billion items in all world languages held in different libraries worldwide are catalogued here. A notable feature of the Worldcat is that their database can be searched directly in Tamil script in Unicode.

For languages of south India, Digital South Asia Library (DSAL) is a major initiative of CRL managed by the University of Chicago [7]. An important component of the DSAL is their Digital Dictionaries of South Asia [8]. In addition to the classic Madras University Tamil Lexicon, digital searchable versions of important Tamil dictionaries of Fabricius, Kadirvelu Pillai, McAlpin and Winslow are available online. The South Asia Union Catalogue, initiated by CRC and managed by DSAL, intends to become an historical bibliography comprehensively describing books and periodicals published in South Asia from 1556 through the present [9]. In addition, it will become a union catalog in which libraries throughout the world owning copies of those imprints may register their holdings. For an overview of some of the digital south asia library efforts, see the special issue of Focus, vol. 24, number 3 (spring 2005), PDF downloadable from CRL website [10]. Roja Muthiah Research Library based in Chennai is an important and modern library for Tamils with over 100,000 volumes of books, journals, and newspapers in their shelves [11]. A digital catalogue of RMRL library is available via the DSAL gateway.
INFLIBNET is a Govt. of India UGC supported consortium network and is slowly evolving as the major information gateway on the holdings of Indian Universities [12]. Vidyanidhi is another Digital library initiative of India to facilitate the creation, archiving and accessing of doctoral theses [13]. Vidyanidhi is envisioned to evolve as a national repository and a consortium for e-theses through participation and partnership with universities, academic institutions and other stakeholders. “Indcat” project of INFLIBNET is unified Online Library Catalogues of books, theses and journals available in major university libraries in India [14]. Over 11 million books and twenty thousand doctoral theses are included in the “indcat” database.

**Digital Preservation of Tamil Literary works**

One of the earliest preservation modes for printed books, still widely in use, is to take microfilm copies of the work. Three main advantages of preservation as microfilm are compact size, long bench life of microfilms (at least 75 years) and facile browsing the content using a microfilm reader. Modern version of the microfilm readers permits even digital capturing of pages of the microfilm or print out the page. University libraries of major institutions in north America and Europe still prefer microfilm form of preservation. Roja Muthiah Research library of Chennai has been making microfilm editions of Tamil works for more than a decade – possibly the only organized effort for Tamil literary works.

The second popular method of digital preservation of works is in the form of image files scanned at high resolution. The main advantage of this approach is that the digital image preserves all the presentation details (layout, artistic drawings, calligraphy used to present the work). Disadvantages are that the file size are huge and the content of the image files not easily searchable. Digital Library of India initiative, launched as part of a larger “million ebooks” project of the Carnegie-Mellon University focused on the digitization of works in various languages of India. Image files in the form of “tiff” format have been made for several thousand works. Tamil books available in e-form at DLI is only few percent, many duplicate files of the same work scanned at different scanning centers.

Recent initiative of Google to digitize collections of major university libraries of US falls in the same category – ebooks in the form of scanned image files [15]. Interestingly for many European languages, good OCR (optical character recognition) software’s exist and they can be used on the scanned image files of printed books to generate a machine-readable version. Google’s online ebooks interface for browsing the content of their eBook collections in fact has enabled this OCR processing to export the equivalent text of a text page displayed as an image file. Unfortunately for Tamil we do not have high performing OCR software for direct use on the image file version of eBooks.

In the last decade several initiatives have been launched for digital preservation of Tamil works. Mention here must be made of Tamil Heritage Foundation [16], Noolaham.net [17] and that of Pollacchi Nesan [18]. All of them produce etexts as image files in tiff or jpg format. Noolaham focuses on the works (books and newspapers) of Tamil authors of Sri Lankan origin. An urgent need in the digital preservation exercise of Tamil Heritage related materials is a central database (union catalogue) of what has been generated as of date. Without a union catalogue, precious human and machine resources are being wasted in digitization of the same work/object by several initiatives.
Machine Readable Version of etexts

There have been several projects to prepare machine-readable version of ancient works (those in public domain) for free distribution on the Net. Pioneering effort has been Project Gutenberg [19], where over 30000 free eBooks are made available to read on the PC and in a number of portable devices such as iPhone and dedicated ebook readers [20] such as Kindle and Sony eBook Reader. Dr. Thomas Malten of the Institute of Indology and Tamil Studies of the Univ. of Koeln, Germany possibly was the first person to take initiative to digitize Tamil works in machine-readable form in early nineties [21]. Supported by grants from the German Government, he prepared electronic texts in Romanized (transliterated) format of Tamil works of first sangam period and few major works such as Kamba Ramayanam. The etexts used a simple plain ASCII based transliteration scheme so that the etexts can be readily searched. An online search interface was made available to search the entire collections free of cost, an extraordinary feat at a time where the computer usage by the linguists and Tamil Diaspora was not wide, as it happened after the arrival of low cost personal computers. In the early nineties, free Tamil fonts started appearing and also software such as Adami that permitted text input in Romanized format with the option to display the equivalent text natively in Tamil script form. Parthasarathy Dileepan of Tennesse US led a maiden initiative to produce e-version of the entire 4000 verses of nalayira divya pirapandam. Using Adhawin (windows version of Adami) software, it was now possible to view and print the etext verses in Tamil script form. Then came Project Madurai devoted to preparation of etexts directly in Tamil script format. Tamil Virtual University, launched in 2001 has included a digital library section which carries etexts of several Tamil works as web-pages [22].

Project Madurai

Project Madurai (PM) was launched in 1998 as a voluntary Net-based initiative for the digital preservation and propagation of Tamil literary works through Internet, natively in Tamil script form [23]. Volunteers from all four corners of the world use their spare time and personal computer equipment to prepare etexts of Tamil literary works natively in Tamil script form. During the past ten years, over 400 Tamil works, big and small, have been digitized and distributed through a dedicated web-server. The etexts are distributed FREE as formatted text in HTML and PDF formats. Anyone can download for personal use and forward to others. Only requirement for reproduction is that the credit acknowledgement lines included in the header part of the etext be preserved.

The coverage is very broad in scope, spanning a wide time period and include works from early Sangam period to contemporary literature: pattuppATTu, eTTuttokai, patinnenkIzkaNakku, epics, religious works of saivaite, vaishnavite canon, Old and new Testament of Bible in Tamil, ciRappuRaNam, saiva siddhanatha works, works of Bharathi, Bharathidasan and Kalki and works of Sri Lankan and Malaysian Tamil authors to mention broad cateogories. Only condition is that the covered Tamil work should be free of copyrights (work of “public domain”) or the author, their legal heirs give permission for royalty-free reproduction of the work and free distribution of the etext worldwide. The coverage includes all “genre” – poetry, prose and drama, English translations of important Tamil works as well Tamil translation of key literature of other world languages.
Challenges faced by PM during the past decade

Project Madurai is the only initiative devoted to preparing machine-readable (searchable) form of etexts. Nearly all other initiatives of digitization store the etexts in the form of image files in tiff, jpg and other formats. For web-delivery of Tamil etexts in a form usable by any average user of Tamil worldwide, several conditions are to be met. For machine-readable texts, important question is the font encoding to be used. Unlike most of the languages of the west, for Tamil, there has not been universal agreement on the font encoding to be used for electronic version of digital data. If the encoding used is not a standard for use in all computer platforms, then suitable Tamil fonts that work flawlessly in all platforms are to be made available free and the font encoding used should be such that there is no corruption of the data during transfer across platforms. When Project Madurai was launched in 1998, we had only two Tamil fonts (Inaimathi and Mylai) were available that satisfied above requirements. Hence etexts were prepared in these two font encodings. Soon there was a Net-based encoding TSCII (Tamil script code for Information Processing) evolved as a popular encoding for Tamil and Project Madurai started releasing etexts in this 8-bit bilingual encoding. It may worth pointing out here that flawless delivery and processing of Tamil digital materials require use a bilingual 8-bit scheme with standard ASCII scheme as part of the scheme. Around 2000 windows PC started supporting Tamil in the multilingual Unicode encoding scheme. As a consequence, Project Madurai started distributing etexts in Unicode format as well. As of date, PM etext collections are available in two encodings TSCII and Unicode.

Project Madurai has been fortunate to have a steady team (small, max 20. at a time but fully committed) of volunteers who contribute hours in the preparation of the etexts. They are Tamil enthusiasts keen to see our literary heritage preserved in e-form but are based in various far-off places without access to a good collection of printed copies of the Tami literary works. Persons who had time to key-in the work or to proof-read the work typed by other (PM works are proof-read at least once independently by a second volunteer) do not have access to printed books and those who have a good personal collection do not have time to contribute to the project. As Project Managers we have to go periodically to different bookshops in Chennai to procure target books for digitization. Sadly Tamil books covering good part of Grammar, prose and poetry are no longer reprinted due to pronounced shift in the interests of Tamils towards novels in late 20th C. So we are obliged to depend on the precious collections still preserved in major public libraries of Tamilnadu. Janet Library of the Univ. of Koeln, Germany is unique in this context to host possibly the biggest collection of Tamil books outside India, with collections over 60,000. Unfortunately financial resources available to the Indology Institute IITS there (where Prof. U Niklas and Dr. T Malten are associated) are extremely limited and the Institute even had passed through total shut down few years ago.

In order to improve the accessibility of books to our volunteers willing to type in the text, PM adopted the “distributed proof-reading (DP)” approach developed by Project Gutenburg. In DP, scanned image files of printed books are collected and stored in a web-server. Volunteers willing to participate in the eBook preparation access these image files, one page at a time, either for key-in or proof-reading. Using special software, the image of a printed page is displayed on the left in a split-screen window with a text-editor on the right to key in (or proof-read) equivalent text. Using this distributed proof-reading implementation at PM (DP-PM) [24] we have been able to produce nearly 100 ebooks during the past few
years wherein a single eBook was produced as a joint effort by a group of volunteers based in different parts of the globe.

Digital Library of India initiative [25] has been very useful to us in supplying image files of several Tamil literary works and we are very grateful to them for this wonderful effort. Unfortunately funding for the DLI project has ceased, though the collections are still available on the Net. One serious limitation of the DLI collections is that the integrity of the text cannot be guaranteed. The image files have been not systematically checked with the source after editing of the graphic files. For several works often few lines at the top or bottom of the page are missing or words clipped on the right or excessive cleaning of the contrast resulting in pure consonants and akara-varisai abugida characters appear in the same manner (as is the case with palm-leaf manuscripts). With such pitfalls, we are obliged to consult the original printed version to ensure textual accuracy of the etext being produced.

Another important factor to consider is the selection of an “authentic” edition of a given work for the etext preparation. Some works have ancient works reproduced without much of sandi/word splitting so as to make the work more understandable to the lay public. For many ancient works there are textual variants (pĀTa pEdam). One key decisive factor at least for works of poetry in nature is the compliance to the grammar rules (metrics). Checking of a given work for metric accuracy requires a higher level of linguistic expertise, not often available with the volunteers who help us produce the etexts. In this context it is preferable that the entire collection of PM etext collections are vetted by a team of language experts, possibly by a team of university level researchers. Tamil scholars like UVeCa compared several editions of ancient Tamil works to compile “authentic” or “critical” editions.

Project Madurai etext collections at best are raw texts as they were written. In his presentation at the last year’s Tamil Internet 2009 conference, Dr. Jean-Luc Chevillard proposed creation of a second generation of etexts for scholarly research [26]. Suggestion was to produce “critical” editions where individual words of verses are “tagged” with indication of the ciir for each word of the verse. Dr. Chevillard pointed out “Text Encoding Initiative (TEI)” provides a number of modules which can be applied (or adapted) to the Tamil case. An etext can be “metric compliant” or “sandi-split”. Since the meaning of verse can vary with the words, contemporary “sandi-split” versions are in reality an interpretation of a literary work. In principle, with the knowledge of the “sandi-splitting” rules and information on the metrics of a given work, sandi-split versions can be converted to metric compliant versions with the use of a software. In fact In the last TIC 2009 conference, Balasundararaman presented a model software that permit checking of poetic works for their metric compliance (venba metrics) [27]. Project Madurai just started extending the coverage of Tamil works with the inclusion of commentaries of works. There are “classical” commentaries by scholars such as parimelazhakar, naccinarkiniyar that must be preserved in digital form.

The advantage of a machine-readable version of an etext is that all the words of the work can be indexed in a database in several ways, permitting facile search on the occurrence of a given word or a string of words (word sequence) in one or more literary works. With a vast collection of etexts of Project Madurai covering Tamil works spanning a wide time frame over two thousand years, such database-driven search for specific words can be very useful in etymology studies by the language experts. We have provided prototypes of such searchable online interfaces for Tamil works, using databases based on MySQL and

397
php query calls. Dr. Vasu Renganathan of the Univ. of Pennsylvania for example, has undertaken such systematic studies of etymology using etext collections of Tamil literary works.

The digital edition of “tEvAram” published by the Pondicherry Institute of Indology is an excellent example of combining multimedia tools with machine-readable text to bring an ebook of the next generation [28]. For each tEvAram verse, in addition to the text, audio version of the verse and graphic (map showing the location of the town where the temple/deity referenced in the text is given) files are added to enhance the utility of the electronic version. Online websites for the teaching of Tamil of the Tamil Virtual University and of Univ. of Pennsylvania Tamil web have nicely integrated multimedia tools to the electronic text of the Tamil work taken up for detailed study.

**Concluding Remarks**

There are areas such as computer-aided teaching of Tamil online or off-line need a comprehensive digital library. In pioneering efforts, Tamil Virtual University and Tamil Web of the University of Pennsylvania have put together a number of educational tools – lessons as web pages with a number of audio- and video clips integrated. The Ministry of Education of the Singaporean Government also has been supporting development of numerous multimedia tools to aid teaching of Tamil at the primary and secondary school level. Throughout this article we have indicated numerous public-funded and voluntary efforts working in the area of Tamil digitization. An umbrella organization that can interface all these isolated efforts can go a long way in reducing redundancies and accelerate output. In addition to the governmental agencies, academics and IT/ICT professionals have important role to play in this key area of academic and public interest.

**Bibliography:**

5. [http://catalog.crl.edu/](http://catalog.crl.edu/)
6. [http://www.worldcat.org](http://www.worldcat.org)
7. [http://dsal.uchicago.edu/](http://dsal.uchicago.edu/)
8. i) [http://www.crl.edu/focus/article/509; ii) http://sauc.uchicago.edu/](http://www.crl.edu/focus/article/509; ii) http://sauc.uchicago.edu/)
10. [http://www.lib.uchicago.edu/e/su/southasia/about-rmrl.html](http://www.lib.uchicago.edu/e/su/southasia/about-rmrl.html)
11. [http://www.inflibnet.ac.in/](http://www.inflibnet.ac.in/)
13. [http://индcat.inflibnet.ac.in/indcat/](http://индcat.inflibnet.ac.in/indcat/)
15. [http://www.noolaham.net/](http://www.noolaham.net/)
18. i) [http://www.uni-koeln.de/phil-fak/indologie2/](http://www.uni-koeln.de/phil-fak/indologie2/) ii) [http://webapps.uni-koeln.de/tamil/](http://webapps.uni-koeln.de/tamil/)
19. [http://www.tamilvu.org](http://www.tamilvu.org)
20. [http://www.projectmadurai.org](http://www.projectmadurai.org)
21. [http://www.new.dli.ernet.in](http://www.new.dli.ernet.in)
Reducing Digital Divide in Tamilnadu using Data Mining Techniques for better E-Governance

R. Jayabrabu
M.C.A.,M.Phil.,(Ph.D.),
Assistant Professor
Department of Computer Application
School of Science and Humanities
Karunya University
Coimbatore – 641 114
Tamilnadu
E-mail: jayabrabu@rediffmail.co.in

Prof. (Dr). V. Saravanan
M.C.A.,M.Phil.,Ph.D.,
Professor & Director
Department of Computer Applications
Dr. N.G.P Institute of Technology
Coimbatore – 641 048,
Tamilnadu
E-mail: profsaran@hotmail.com

Er. A.K. Balakrishnan
M & B Associates
75 Naichimuthu Gounder Colony
Sanganoor Road
Ganapathy
Coimbatore – 641 006,
Tamilnadu
E-mail: cbebalakrishnan@gmail.com

Abstract
The term digital divide refers to the gap between people with effective access to digital and information technology and those with very limited access. In other words it is closely related to the knowledge divide or knowledge share due to the lack of technology and knowledge. The extraction of useful and non-trivial information from the huge amount of data available in many and diverse fields of science, business and engineering is called as Data Mining. Data Mining techniques and algorithms are the actual tools that analysts have at their disposal to find unknown patterns and correlation in the data. For effective use of E-governance in Tamil Nadu, the digital divide to be reduced. Most of the Government departments are already using E-governance in Tamil Nadu. This is the appropriate time for us to analyze the effectiveness and reach ability of technology to all sectors of peoples. Even the most learned peoples are reluctant in using the technology. This digital divide gap leads to improper usage of Information and Communication technologies. The objective of this paper is to analyze the following two important digital divide issues using data mining and present recommendations for better E-Governance in Tamil Nadu.

a. Improving Quality of Bandwidth/Parameters.
Since, the information and communication technologies are being implemented in the Government at different levels; good bandwidth is needed for constant transformation of knowledge in a proper format. For the better usage of E-Governance, the quality and performance of bandwidth performance has to be increased. In Tamil Nadu, there are so many service providers available for connectivity. But, the expected quality of bandwidth is less than the assured bandwidth. This paper analyses the bandwidth parameters using data mining techniques and suggest a better framework for improving bandwidth across Tamil Nadu.
b. Taking Technology to reduce the gap

Now a day, many new information and communication technologies are introduced. The urban sector is reluctant in using the technology due to the fear of using the technology and also thinking communication/network failure, which occurs frequently. The middle age person still thinks that the technology is very far from them and also is very costlier. The rural sector is unaware of these technologies and they need to be provided infrastructure and training. With the increase usage of mobile phones, convergence of technologies also need to be thought of. This paper analyses the need of urban and rural sector people for the effective reach of technology. Data Mining Techniques are used for data analysis, which leads creation of to better E-governance standards.

The above mentioned parameters are studied by applying data mining techniques such as Association rule mining (determine implication rules for a subset of record attributes), Classification (assign each record of a database to one of a predefined set of classes analysis and Clustering Techniques (find groups of records that are close according to some user defined metrics) and a suitable framework is proposed for better E-Governance.

1. Introduction

When the IT industry increased globally in the 19th century, simultaneously the Internet and the Mobile technologies are also emerged into the world and ruled majority of the people. With this, E-Governance also booms out with the help of some Government Departments around the world. In India, National Informatics Center (NIC) played a vital role for the development of E-Governance in which they incorporate some of the Government related activated like Tax payment, Census Generation, Election Management, Disaster Management,[1] etc.,

In Tamil Nadu, some of the successful E-Governance projects are land registration, call for tender, issue of birth/death certificates, agriculture, e-transaction, RTO, tourism, infrastructure, land/local tax, local body election details, e-ticket etc., [1]. The major scenario in the above-mentioned successful E-Governance is heterogeneous based System. The entire activities of each Government related activities possesses a unique database to store their respective data. This technique is followed in our state and also other states too [2]. As a fact, each Government department maintains their own database as unique and there is no interlinking between various departments/databases. When, the land registration department needs some information about agriculture data, they are not able to access the agriculture database. This leads to minimum usage of the e-governance projects by the citizens. The digital gap increases due to the issue and the important e-governance projects fails after implementation. This paper proposes the use of data mining techniques and a better framework to reduce the digital gap and to interlink the heterogeneous databases. This paper proposed two stages to reduce the digital gap

   a. Improving Quality of Bandwidth/Parameters.
   b. Taking Technology to reduce the gap

2. Using Data Mining Techniques to Reduce the Digital Gap

Data Mining is the technique to explore and analyze the large data sets, in order to discover meaningful patterns and rules [6]. The evaluation of data mining techniques began when the business data are stored
in the database and the technologies were generated to allow the user to navigate the data in the real time.

Recently, the ICT made a proposal for all the state and central Government for the betterment of database maintenance in the near future generation [3]. As we know, now a days, all the Government departments utilizes huge amount of data in their day-to-day work, which leads to maximize the access of current or history of datasets from the database [2]. But, it is not possible to fetch the datasets when they need. This is because of insufficient data, improper format, duplicated data, and some technical problems etc.,; When we discuss on other side, it is also due to less bandwidth, natural disaster, network failure, and loss of data during data transmission and collision of packet with one another etc. As a result of this, the end user cannot able to perform the operation with in the time and also little afraid to continue the E-governance system. Since, a gap is generated between user and existing E-Governance systems. As a result, the Government should concentrate on above set problems for the betterment of E-Governance. By considering these issues, this paper proposes the use of data mining techniques to reduce the digital gap. The major data mining techniques considered in this paper are [6]

a. Association Techniques.

b. Classification techniques.

c. Clustering Techniques.

**Association**: It is method for discovering interesting relations between the variables in the large database. There are different types of algorithm for association rule. They are Apriori algorithm, éclat algorithm, FP-growth algorithm, One-attribute-rule algorithm, Opus search algorithms, and Zero-attribute-rule algorithm [6]. Let us consider the existing E-Governance agriculture database as an example. Suppose, when a user needs a land for the cultivation process with the following features, i.e, good water, larger area, good manpower, and good soil. Based on the above features, the end user can easily search the availability of lands form the existing database with the help of some association algorithm. The one of the best algorithm for technique is Apriori Algorithm.

**Classification**: It is one of the data mining techniques used to predict the group for data instance. Some of the popular classification techniques are decision trees and neural networks [6]. From the existing database, the end user can classify the land with required parameters like state wise, of district wise, area wise and etc by means of tree like structure. By this classification technique, the user can easily classify the required data from the existing database using some protocols. Based on this, the user can identify the locations and nature of the land with a faster manner. Some of the best and easiest algorithms are decision tree and nearest neighbor algorithm that is available in data mining techniques for better classification.

**Clustering**: It defined as collection of data object that are similar to one another within the same cluster and dissimilar to the objects in the other cluster. Clustering algorithms are broadly classified into hierarchical and partitioning clustering algorithm (Jain and Dubes, 1988). Again, the Hierarchical algorithm are Agglomerative and Divisive algorithm and the Partitioning Algorithms are k-means, k-mediod, DBSCAN, CLARA, CLARANS, BIRCH CLIQUE, OPTICS etc [6]., When a person is willing to find the group of land for cultivation respective of location, the user can apply the clustering techniques
with the existing e-governance database to form a new groups based upon the user requirement. Thus the user may satisfy.

This is the appropriate time for us to discuss the effectiveness and reachability of technology to all sectors of people. Even the most learned people are reluctant in using the E-Governance technology. This digital divide gap leads to improper usage of Information and Communication technologies. By using the above specified data mining techniques, the digital gap is reduced which in turn help the state to move towards implementing better and quality of E-Governance projects.

3. Improving quality of bandwidth/parameters for better e-governance:

In general, some of the service providers like BSNL, AIRTEL, etc., are available for network connectivity in Tamil Nadu for good quality of Bandwidth. Bandwidth is defined as amount of data transferred in a given period of time [8]. Since, each service providers are having different qualities of bandwidth. But the expected quality of bandwidth is less than the assured bandwidth. As result the network connectivity in Tamil Nadu reached towards down state. Due to this, the successful E-Governance projects get failed while performing data transactions. By considering the above facts, the quality of service (QoS) need to be improved and also all the service provides are expected to provide guarantees for constant network connections. Bandwidth is one of the major constrain for better E-Governance. Some of the parameters are identified to rectify the poor bandwidth problem. For constant connectivity and the better usage of e-governance, the identified parameters are as follows [8]

a. Availability  
b. Throughput  
c. Data latency  
d. Error rate  
e. Network Traffic  

**Availability:** It is defined as the probability that a device will perform a required function without failure under defined conditions for a defined period of time. In most of the case, availability is an important characteristic of system but it becomes more critical and complex issues on networks. With the help of Data Mining technique the network availability are classified with various parameters and helps the service provided for better network availability. Thus, by applying the classification techniques in network database, availability problem will be rectified.

**Throughput:** It is defined as the rate of communication links or network access. The Throughput is generally measured in bits per second, and sometimes in data packets per second or data packet per time slot. By applying the data mining association algorithm, the service provided will come to normalize the size of the packet for data transformation from one place to another with respect to time and network availability. Based on the mining techniques, the problems are identified and help in future that is not repeated.

**Data latency:** It is defined as how much time it takes for a packet of data to transfer form one destination point to another destination. The latency mainly depends on the nature of the electromagnetic signal.
Thus the latency may be differing from device to device. Hence, data mining association techniques are applied on the history dataset to identify when the problem happens and how the problem happens; Is it happen previously? If yes, what actions are taken to solve the problem?

**Error Rate:** It is defined as the number of received bits that have been altered due to noise and interference while during digital data transmission. The error rate may vary from device to device and software application to application. Thus by applying clustering techniques, the service provider can mine the error rate with respect to the hardware and application software from the previous data. Based on this method, the service provider knows which application software Vs hardware device suppose to minimize the error rate.

**Network Traffic:** It is defined as the data in a network, where the network traffic controller controls the traffic, bandwidth, prioritizing the data packet while during transformation form one point to another. The major part is to measure the network traffic like where the network congestion happens, with this, the classification techniques are applied and the same issue was happened in the previous days or not. Based on the result, the identified problems are rectified.

**Routing Performance:** It is defined as measuring the performance of the router depends upon the load offered of it, i.e. by means of heavy load of test traffic will reveal the performance. Based on the traffic and load the performance may vary. For better performance, the traffic should be shaped and the packet size should be constant throughout the entire process. With the help of data mining classification techniques, the provider can mine the lesser traffic network for better routing performance.

Network problem happens not only due to technical side but also due to natural calamities, breaking of cable, etc. From the above scenario, the Central or State Government has to rework on the above-mentioned areas to improve the bandwidth performance by means of advanced networking technology, Fiber Optic and recent computing technologies will acted as catalyst for improving bandwidths.

In this paper, bandwidth parameters are analyzed with the help of few data mining techniques for network connectivity to improve the bandwidth. The framework is developed to provide better network connectivity for E-Governance. This paper analyses the bandwidth parameters using data mining techniques and suggest a better framework for improving bandwidth across Tamil Nadu.

4. **Proposed frame work for reducing the digital gap:**

In Tamil Nadu, there are more successful E-Governance projects being implemented. But, all the implemented applications are heterogeneous in nature i.e. the databases are not linked for effective usage. Due to the non-linking of databases and availability in different geographical locations, there exist a digital gap. In the proposed framework, a new concept is introduced to reduce this digital gap, instead of storing the data in different location. This paper proposes the creation of data warehouse, which is a subject-oriented, integrated, time-varying, non-volatile collection of data [5][7].

All the existing and emerging E-Governance databases which is heterogeneous in nature and available in geographical locations are combined and get stored in a common place called ‘Data Warehouse’. It may be called as state data warehouse or data repository. The users using a particular E-governance application is able to use the other application also effectively thereby the usage of E-governance applications are increased. Thus, digital gap is also reduced.
From the above figure, the E-Governance technology/applications data are collected from different locations and get stored in different database. This paper proposed a framework in which all the heterogeneous databases are combined and stored in one common place called data warehouse [5]. It contains the summary of all the data, which are made available in a day today process. As per this concept, any one can access any kinds of data at any time by the data warehouse with a faster way. Different data mining techniques are also made available in the proposed framework. By applying these data mining techniques based on the user requirement, the user can mine the data with meaningful order, proper format and in time [7]. Hence, the Tamil Nadu Government E-Governance projects are used more effectively than other State Government projects. With this work, the technology gap is also reduced and the users may utilize the E-Governance by higher level.

![Diagram](image)

**Fig-1: Frame Work for the Betterment of E-Governance.**

**Conclusion:**

In this paper, the importance of digital gaps and the parameters for reducing the digital gap with the E-Governance in Tamilnadu are discussed using different data mining techniques for the better performance. Various network Quality of Services(QOS) parameters such as availability, throughput, data latency, error rate, network traffic and routing performance are considered in data mining perspective to increase the available bandwidth. With the help of proposed framework, the gap also gets reduced between the user and the e-governance systems which enable the Tamilnadu government to implement successful projects.

**REFERENCES:**


8. WWW.compnetworking.about.com, April 2010.
தமிழ் மரப கருத்துக் காலம் குறித்து செய்திகள், பின் நூற்றாண்டு, தமிழகம் களஞ்சியாம் தீர்த்தகங்கற்பயிற்சியானது

திலகம் ஆண்டாண்டான

வாசுமார் வரலாற்று

Technical Consultant, Hewlett Packard, Germany.
Email: ksushini@gmail.com

தமிழ்ப் பதிப்பு, தமிழ் மரப அறிக்கையக் கையெழுத்திட்டு [http://www.tamilheritage.org]

தமிழில் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு. கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மை�ான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மை�ாக மாற்றப்பட்டு. கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு. கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு, கணினிகள் பயன்படுத்தப்படும் மாற்றங்கள் மற்றும் மாற்றங்கள் உண்மையான உண்மையாக மாற்றப்பட்டு.
4. பார்வைகள் அலையும் பட்டியல்களும் கொண்ட வருமானம் கொண்ட கதாத்தாக்குகளில் உள்ள படிமச் செய்திகளை எடுத்துக்காட்டுவது.

1. கொண்டாட்டங்கள் கூறுப்பட்ட தரவுப் பட்டியல்கள் அலையும் பட்டியல்களும் கொண்டவை வருமானவும் செல்வாக்கு
2. படுப்படுத்தல் நுழைவு
3. படுப்படுத்தல் குறிப்பிட்டு வருமான குறியீடுகள் கொண்டவை
4. கொண்டாட்டங்கள் விளக்காக்கு வருமானத்தை அலையும் பட்டியல்களின் admin முகட்டிடைகை
5. கொண்டாட்டம் படுப்படுத்தல் செய்யப் படிமகளின் (table)
6. கொண்டாட்டம் படுப்படுத்தல் செய்யப் படிமகளின் (table)ஆக்கான
7. php அலையும் பட்டியல்களும் கொண்டாட்டம் கொண்டவை வருமானம் குறியீட்டு
MySQL database and the programming languages such as PHP and Python. This is because MySQL has been used for web applications and database management systems. However, it is important to note that the use of MySQL is not always the best choice for every situation.

When choosing a database management system, it is important to consider factors such as the size and complexity of the application, the scalability requirements, and the performance needs. MySQL is a good choice for smaller applications and environments where performance is not a major concern. However, for larger applications and environments with high performance requirements, other database management systems such as Oracle or Microsoft SQL Server may be more appropriate.

In conclusion, MySQL is a popular database management system that is widely used for web applications and database management systems. Its flexibility and ease of use make it a good choice for many users. However, it is important to consider other factors such as scalability and performance when choosing a database management system for a particular application.
இருவரும் அனாறு கல்விக்கேற்ற பயிற்சிக்கு ஒரு போலவே உரையானது. குறிப்பிட்டு கூறுவது ஒரு சாதனை கொண்டாலும் குறிப்பிட்டு காட்டுவதே ஒரு சாதனை. குறிப்பிட்டு கூறுவது ஒரு சாதனை கொண்டாலும் குறிப்பிட்டு காட்டுவதே ஒரு சாதனை.
An Extended Cross Lingual Information Retrieval System for Agricultural Domain using Statistical Document Translation for Tamil Farmers

D. Thenmozhi, Arun Balachandran Ganesan and C. Aravindan
Department of Computer Science & Engineering
SSN College of Engineering, Chennai, India
theni_d, arunbalachandrang, aravindanc@ssn.edu.in

Abstract
Cross Lingual Information Retrieval (CLIR) system allows a user to pose a query in one language and search documents in a different language. In this paper, we extend the use of Tamil-English CLIR system for Tamil farmers by translating the English web pages to Tamil, through which the farmer can pose a query in Tamil and read the information in the same language. We have developed statistical English-Tamil translation engine to translate the English documents that are retrieved from CLIR system to Tamil. The parallel corpus for this statistical engine was built using the text in the domain of Agriculture for Tamil and English languages. The translation model is then trained with this sentence aligned domain-specific corpus. Also a text corpus for Tamil has been built and used in building and training the language model. A Statistical Machine Translation Decoder tool has been used to perform the decoding as the final step.

Introduction
The world of information is huge and expanding; also most of the information is available in English. The non English speaking users still find it a major problem in utilizing this vast resource of information. Accessing this information through queries written in Indian languages is still more difficult. CLIR systems solves this problem by allowing the non English user to specify their information need in their native language and accessing rich information that are available in English.

We have developed a Tamil-English CLIR system for Agriculture society [8] which allows the Tamil farmers to pose Tamil queries and retrieves information from English corpus. The CLIR systems generally display the search result in English. It is appropriate, if the results are displayed in their own language for the users who do not know how to give query in English.

In this paper, we extend the CLIR system by translating the retrieved English documents to Tamil in the domain of Agriculture through which Tamil farmers can enter their query in Tamil, retrieves relevant pages from English corpus and read the information in Tamil.

Machine translation is a sub-field of computational linguistics that investigates the use of computer software to translate text from one language to another. It can use a method based on linguistic rules, when the content to be translated is simple, for example short queries. When the content to be translated is complex in structure, machine translation requires the problem of natural language understanding to
be solved first. Statistical machine translation solves this problem with a probabilistic approach based on bilingual text corpora.

We have built English-Tamil parallel corpus in the domain of Agriculture to train the statistical translation system for translating the retrieved documents in English to Tamil.

Our translation system may contribute in developing Tamil Wikipedia by translating the Wikipedia content that are available in English in the domain of Agriculture to Tamil.

Literature Survey

a. Cross Lingual Information Retrieval System

The Advanced Cross-Lingual Information Access (ACLIA) currently works on two tasks Information Retrieval and Question Answering [7]. For both these tasks, they work for cross-lingual and mono-lingual topics for the languages English (EN), Simplified Chinese (CS), Traditional Chinese (CT), and Japanese (JA) for the tracks EN-CS, EN-CT, EN-JA (cross-lingual) and CS-CS, CT-CT, JA-JA (mono-lingual). [7] created the Evaluation Package for ACLIA and NTCIR (EPAN). The EPAN toolkit contains a web interface, a set of utilities and a backend database for persistent storage of evaluation topics, gold standard nuggets, submitted runs, and evaluation results for training and formal run datasets. [4] developed a Question Answering system that answers complex questions from multilingual sources. They improved the performance of Chinese-to-Chinese and Japanese-to-Japanese subtasks of NTCIR7 IR4QA by means of Mean Average Precision. ACLIA presently not working on the document translation of cross-lingual track.

Many research groups are working on the CLIR system for Indian Languages [6]. [5] developed a Tamil-English CLIR system that uses Tamil morphological analyzer for language analysis. A named entity recognizer is used to identify the named entities in the document for indexing and ranking. They used a bi-lingual dictionary approach for translation. A statistical engine based on n-gram approach is used for transliteration. A simple ontology is used for query expansion. Ranking is based on term frequency. They evaluated the system for the news domain by collecting the English corpus from the magazine “The Telegraph”.

We differ from the above works by extending Tamil-English CLIR system with document translation in the domain of Agriculture which helps Tamil farmers who have English as a language barrier.

b. Machine Translation System

[1] developed a English to Tamil translation system using statistical approach. SRILM toolkit it used for language modeling. Translation model is trained using the parallel corpus in the News domain. They learnt Named entities from this statistical machine translation system. Phrase based decoder is used to translate the English sentences to Tamil.

[3] used the same phrase based approach of statistical machine translation to transliterate word from English to Hindi, Tamil and Kannada languages for the Named-entities. They used GIZA++ for word alignment and Moses for decoding.

We also developed a similar kind of statistical machine translation system that translates English sentences to Tamil by using SRILM toolkit for the language modeling of Tamil and Moses for translation modeling and decoding.

**System Architecture**

The extended CLIR system uses a three phase model that accepts the query from the user in Tamil extracts the documents from English corpus and translates them to Tamil in the domain of Agriculture. This is illustrated in the figure 3.1.

**a. Tamil-to-English Query Translation**

A rule based Tamil-to-English translation engine is developed through which the query is entered in Tamil and translates to English. It uses morphological analyzer to split the query into individual words. These words are translated to English using a Tamil-English bilingual dictionary. When Tamil words gets ambiguous meaning, the correct meaning is obtained based on the context using word sense disambiguation. The obtained English words are re-ordered based on the syntactic structure of English language (Subject-Verb-Object pattern).
b. Searching and Retrieval

The translated English query is given to the existing search engines like Google, AltaVista, etc that retrieves the English documents. We considered top twenty pages that are retrieved by the search engine for statistical document translation. The documents with the extension doc, html and pdf are converted to plain text before given to the English-to-Tamil translation Engine.

c. English-Tamil Document Translation

A statistical English-Tamil document translation engine is developed with the phases namely constructing English-Tamil parallel corpus, training the translation model with the parallel corpus, training the language model for Tamil language and decoding.

Tamil Wikipedia, and the Tamil Nadu Agricultural University websites have been identified as excellent sources that would help in providing the data needed for corpus building. Information in documents has been extracted/retrieved using a semi-automated text extractor and unnecessary attributes to the data is removed either by automation or human intervention depending on the type of attribute. The encoding of the content in Tamil language is converted to Unicode standard. The aligned parallel corpora that is built follows Unicode standard.

Moses toolkit is used to train the translation model using the parallel corpus based on phrase based approach. This model works on the principle of Baye’s rule

\[ P(t|e) = P(e|t)P(t)/P(e) \]

\[ T = \arg\max P(e|t)P(t) \]

where \( T \) is the translation probability.

SRILM toolkit is used to build tri-gram language model for Tamil language. Tamil sentences extracted for building parallel corpus in the domain of Agriculture is used to build this model. A beam search decoder of Moses toolkit for phrase-based statistical machine translation model is used to translate the sentences of English documents to Tamil.

Experimental Evaluation

a. Agriculture Corpus

We extracted the parallel sentences from Tamil Nadu Agriculture University (TNAU) website and Wikipedia content in the domain of Agriculture. Some additional sentences in Tamil from the TNAU website is used to train the language model. The number of sentences used to train the translation model and language model are given the table.

<table>
<thead>
<tr>
<th></th>
<th>Number of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TNAU Site</td>
</tr>
<tr>
<td>Translation model</td>
<td>6135</td>
</tr>
<tr>
<td>Language model</td>
<td>6900</td>
</tr>
</tbody>
</table>

b. Comparison with Google Translation System

We compared the performance of our extended system with the Google translation system by various measures that are listed in the following table.
<table>
<thead>
<tr>
<th>Tamil Queries</th>
<th>Google Translation System</th>
<th>CLIR System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transliterated Approach</td>
<td>Machine Translation Approach</td>
</tr>
<tr>
<td>Document Retrieval for Tamil Queries</td>
<td>From Tamil Corpus</td>
<td>From English Corpus</td>
</tr>
<tr>
<td>Recall value for Tamil Queries</td>
<td>Less</td>
<td>More</td>
</tr>
<tr>
<td>Translation</td>
<td>Available for many Asian/European languages. Tamil is not available</td>
<td>Translates English documents to Tamil</td>
</tr>
</tbody>
</table>

**Conclusion**

This extended CLIR system helps the Tamil farmers to look for rich information available in English corpus by specifying their need in Tamil and view the results also in Tamil language. This system uses a rule based approach for query translation and statistical based approach for document translation. With the available parallel corpus, query translation also can be done statistically in future. Factored based translation model can be used for improving document translation. Query also can be enriched with more semantics by using Agriculture ontology.

**References**

8. Thenmozhi D, Aravindan. C, “Tamil-English Cross Lingual Information Retrieval System for Agriculture Society”, 8th International Conference on Tamil Internet, Germany, 2009
Tamil E-Archives Management

B. Asadullah
Librarian
Mazharul Uloom College, Ambur – 635 802.
Vellore District, Tamil Nadu.

Abstract
From the year 2000 onwards, we had noticed that new innovations were made enormously to serve the human needs in every core area from Agriculture to Engineering, Medical, and Automobile etc., from every corner of the universe in every language of the world, and more especially by Tamil Literature. This has compelled us to tap all these new findings and old traditional findings in a secure way so that these resources can be better utilized in future by our new generation peoples.

In keeping these concerns in mind, I am trying to explain how we can secure all these findings in a secure way using the suitable technology by applying possible low cost of implementation and maintenance. This paper will explain the way, on how to manage the archives of new findings and old traditional findings of Tamil languages in electronic format using pictorial images and formal text format utilizing the best technology at low cost using best available free of cost Open Source Software.

Introduction
Tamil language is spoken and widely used by over 66 million peoples across the world in particular in three countries as official language and in six countries as monitories, where our state of India Tamil Nadu is the largest user of Tamil in every aspect of life. Tamil scripts are preserved in different format from palm leaf to paper, due to accessibility and timely availability. But this preservation does not last for number of years as every functional item has a stop point that is everything has a counted life span. This article will help in preserving archive in computer using latest and freely available licensed open source software.

Electronic Preservation
Due to moderate and vast development of Computer Technology, preservation has become reliable, secure and durable. Here preservation means not only documents, but for vast range of items in electronic format like running text, document, book, journal, pictorial, audio, video and animated format etc., all these above said format are also known as E-Achieves.

Preservation in computer is also a big question, but the answer to this question is having backup of all thing in a compact format, or having a mirror storage which can be made possible using the RAID technology while storing the data in computers. In case of one’s end of life, another mirror image can be utilized in less effort, instead of building whole thing from scratch.

From the above last two clause, we come know that preservation in computer is possible, durable and securable. But maintenance of unlimited e-archives in a computer is not so simple and easily accessible for all end-users.
For this purpose, after wide consultation and analysis, it was concluded that there were a suitable software which fit for the chosen topic, and it is available free of cost in different users operating environment also known as operating system like Microsoft Windows, Linux, under the Open Source Software GNU license.

**Institutional Repository**

An Institutional Repository is an online locus for collecting, preserving, and disseminating in digital form. (Also known as Digital Library). The main objectives for an institutional repository are to create global visibility for an institution's scholarly research and to collect content in a single location, to provide open access to institutional research output by self-archiving it and to store and preserve other institutional digital assets, including unpublished works. Some examples are Fedora, E-Print, IR+, and DSpace, Green Stone.

DSpace is developed by HP Lab and MIT Libraries of US. At the time of writing this article for presentation at the 9th Tamil Internet Conference held at Coimbatore, Tamil Nadu (India) 2010, DSpace 1.6.2 is available for utilization. DSpace is widely accepted and used by different educational institutions across the world.

DSpace is open source software based on BSD License, it is an digital object / assets management system to create, store, search and retrieve digital objects. It allows open access to digital objects and building institutional repositories and the collections are searchable and retrievable on the web. Since it is open source technology platform which can be customized and its capabilities can be extended.

Main features of DSpace are it is a low cost, including all hardware and software components. Technically simple to install and manage, robust, scalable, open and inter-operable, it is modular and user friendly with multi user environment (including both searching and maintenance), it is multimedia digital object enabled and platform independent interoperable. DSpace stores all data as a digital object, which has the following standard model.

<table>
<thead>
<tr>
<th>Objects in the DSpace Data Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object</strong></td>
</tr>
<tr>
<td>Community</td>
</tr>
<tr>
<td>Collection</td>
</tr>
<tr>
<td>Item</td>
</tr>
<tr>
<td>Bundle</td>
</tr>
<tr>
<td>Bitstream</td>
</tr>
<tr>
<td>Bitstream format</td>
</tr>
</tbody>
</table>

DSpace has its own Ingest Process and Workflow Steps in three stage, which verifies all type of documents with their corresponding information, then only it accept the document to be stored in the repository. This method helps in gathering authenticated Tamil document for building a healthy achieve.
DSpace uses the CNRI Handle System for creating identifier, which is required by DSpace to store and locate the independent mechanism for creating and maintaining digital objects.

DSpace uses the Lucene Search algorithm, which helps in searching the tamil text using Fielded search, Boolean, Exact term, proximity and wild cards, fuzzy and range, boosting terms search techniques.

Content Management System

A content management system is a collection of procedures used to manage work flow in a collaborative environment. These procedures can be manual or computer based. The procedures are designed to allow for a large number of people to contribute to and share stored data with controlled access to data, based on user roles. User roles define what information each user can view or edit and aid in easy storage and retrieval of data by reducing repetitive duplicate input with improving the ease of report writing and improved communication between users.

In a content management system, data can be defined as nearly anything documents, movies, pictures, phone numbers, scientific data, etc. These are frequently used for storing, controlling, revising, semantically enriching, and publishing documentation. Examples are web content management, digital asset management, digital records management and electronic content management. Synchronization of intermediate steps, and collation into a final product are common goals of each. The features of content management systems are, it can structure the document repository as per requirements and automatic document versioning control with content categorization and role-based access control. Using varied content types, it can be managed like HTML, Unicode Tamil text, PDF, Documents of different standard of any language, image files, video clips etc., Document can be accessed through web-browser and network drive mapping with full text searching facility in txt, doc, pdf, odt, sxw, html, xls, sxc, ods etc.,

Document Repository System

A document repository system is a systematic way of storing documents along with a set of user defined attributes that need to be filled in at the time of storage. Documents can be searched based on these attributes or on their content. Every organization needs a good planned Document repository system in their specific official language like in our case Tamil language.

Owl is one of the best examples for document repository system, Owl is a powerful system, full of use full features, it allows defining personal sets of attributes for documents. MS-Word of Tamil language, PDF, Unicode TEXT of Tamil, can be word indexed and searched. However documents and files of any type can be stored. Owl has security model for document access.

The documents are uploaded and stored in a folder, uploaded files can be stored in tables in the open source database like RDBMS, MySQL, PostgreSQL for storing file / folder related information.

Conclusion

Managing achieve is not the easy task, when large numbers are being created day to day because of vast development of Tamil language and literature. In this context managing archive has become easy with the help of computers using proper software depending on our usage and requirement.
From the above article, we come to conclusion that DSpace is good digital library software, which helps in managing the vast digital objects of Tamil literature by giving good retrieval facility with modern technology.

Content management system, also helps in managing E-Archive, but its method is different when compare with Institutional repositories software. Document repository system, help in storing e-achieves systematically.

References.

- Bhatt, R K : March towards digitations of Information Resource in India.
- http://www.unesco.org/
- http://www.eprints.org/
- http://www.irplus.org/
- http://www.dspace.org/
- http://www.greenstone.org/
Very long-term digital preservation and archival strategies for Tamil documents

Mani M. Manivannan
Senior Director of Engineering
Symantec Corporation
Chennai, India
mmanivannan@gmail.com

Abstract

The decipherment of Indus script remains controversial. For several years, the inscriptions in Tamil Brahmi script weren't recognized as Tamil. The inscriptions in Vaṭṭeḻuttu, Pallava grantha, and evolving Tamil Brahmi scripts required careful analysis to decipher. Probable errors in inscriptions and reading of the inscriptions have led to multiple interpretations. As writing instruments and technology changed, Tamils have lost valuable literature and public documents. In the short duration that digital Tamil has existed, we are already finding it difficult to retrieve early Tamil writings on the internet. As we embark on the e-Governance and digital archival of precious documents, we have to remember that unless we are careful, we may find it difficult to preserve these documents for the long duration of centuries. In this paper, we will explore some of the strategies to minimize the loss of valuable documents.

1. Background and Overview

Tamils have been creating digital documents in Tamil since the mid 1980s[1] using MS-DOS based editors. Since the early 1990s Tamils have been privately exchanging e-mails in Tamil and with the advent of tamil.net mailing list in 1996, Tamils worldwide started to communicate with each other in Tamil script. Since then there have been multiple Tamil mailing lists, web pages, webzines, blogs that added Tamil digital content. Project Madurai, a community project has been digitizing classical Tamil literature and Tamil books in the public domain and creating a major Tamil corpus with its digital documents since January 1998. Though the Tamils of the diaspora were instrumental in creating a lot of the early Tamil digital content on the web, the mainland Tamil Nadu media have taken to the web and nearly all the Tamil news media have Tamil content on the web. Tamil Nadu government departments have been creating Tamil digital documents in various encodings including TAB, TAM and Baamini. With a nationwide push for e-Governance in India, the Tamil Nadu government is about to embark on one of the largest Tamil digitization efforts.

Before the advent of the standardization efforts to unify the multiple encodings after the TamilNet’97 conference in Singapore, there were multiple fonts each with its proprietary encoding. These discussions led to the TSCII encoding, the first open, non-proprietary Tamil encoding specification in 1998. At the TamilNet’99 conference in Chennai, the Government of Tamil Nadu formally declared two encoding standards TAB and TAM. During this period, neither ISCII, the Indian national standard and Unicode the global standard based on ISCII, had much support among the Tamil developers. However the standards TSCII, TAB and TAM did not stop others from creating fonts based on proprietary encodings.
Baamini has been popular among the Eezham Tamil diaspora before TSCII and it is still used by some in the Government of Tamil Nadu. Other encodings that were used to create Tamil digital documents include Vanavil, Indoweb, Murasoli, Webulagam, Thinathanthi, Dinamani, Thinaboomi, Murasu Anjal, Mylai, Thatstamil, ShreeLipi, Amudham (Dinakaran), Vikatan, Anu Graphics (Pallavar), and Senthamizh (Nakkeeran). These encodings are popular enough that some converters recognize all of these. With strong support for the Unicode encoding in Microsoft Windows and Apple platforms as well as Google’s search and e-mail applications, use of Unicode has started to spread. Tamil Wikipedia is a popular site that uses Unicode as does the Tamil Lexicon at Chicago University. With the National e-Governance plan of Government of India notifying Unicode as the standard for e-Governance applications, creation of Tamil digital documents in Unicode encoding is expected to increase substantially.

INFITT has recognized Unicode as the main standard for Tamil computing. INFITT has also acknowledged that there are commercial applications that don’t yet support Unicode, such as those used in publishing and other industries. For these applications that are not Unicode-ready, INFITT has recognized Tamil All Character Encoding TACE-16 as the only Alternate Standard for Tamil Computing. Since high-end publications are expected to create PDF documents such as Government notifications, textbooks, etc., along with books published by vendors in Tamil using TACE-16 encoding, this will form part of the Tamil digital document collection.

In this paper, we will review the impact of technological obsolescence on the Tamil digital documents created in the past 25 years and compare it to the impact on English language digital documents. We will also study the preservation and archival strategies that are evolving in the rest of the world to address the threats to the stability of digital documents. Drawing inspiration from the seminal paper by Paul Convey [3], we hope that this paper will stimulate in-depth discussions among the technical specialists and the broader audiences interested in preserving not only the historical and cultural heritage of Tamils but also the more mundane public records that the government and the citizens depend on across generations.

2. What is the problem? Digital Dilemma!

All current digital storage media are ephemeral. As Conway’s graph demonstrates the millennia-old Indus valley signs and the ancient Tamil inscriptions, while fragile, are still legible [2]. Even the palm manuscripts have managed to survive long enough to transfer information to the future generations.

![Figure 1. Information Density vs. Life Expectancy](image_url)
Books printed on acidic paper have lasted more than a century. While the newer recording media allow us to pack enormous information in much smaller area, the shelf-lives of the media, the machines that read them and the applications that render them are all very short (see Figure 1).

2.1 Threats to the Digital Storage

Researchers have identified several threats to the longevity, integrity, access and quality of the digital information storage. Some of these are Media decay and failure, Bit rot, Outdated media, Massive storage failures, Network failure, Access component obsolescence, Outdated formats, Applications and systems failure, Natural Disasters, Human and Software errors, External attacks, Insider attacks, Economic failure, Organizational failure, Politics and Censorship.

In the Tamil context, we can cite some prominent examples. Ambalam.com, a webzine that was launched before Tamil Internet 2000, with a pre-eminent Tamil writer “Sujatha” Rangarajan at the helm, is no longer online. The TamilNation.org site, a veritable encyclopedia of Tamil themed collections shut down once, was resurrected later only to shut down recently for reasons unknown (see Figure 2).

![Figure 2. A Major Tamil website disappears](image)

The status of the archives of Tamil digital data from the late 1980s to early 1990s is unknown. The data formats of the old word processors prominent in the MS-DOS and early Windows platforms are not supported by the vendors. In some cases the vendors are no longer around.

The mail archives of one of the first public mailing lists in Tamil (tamil@tamil.net) do not have its oldest archives (starting from October 1996 to September 1998) online. If a researcher were to look for the e-mail where Kōmakaṭ (aka Rajakumaran, the editor of the Malaysian Tamil weekly “Nayanam”) proposed the word iṇaiyam (இணையம்), it will be difficult to get an authentic version of that e-mail from the tamil.net archives.

The encoding and format changes have impacted Tamil digital information in a major way as well. Several early Tamil mailing lists such as tamil.net, agathiyan@yahoo groups.com, tamilulagam@yahoo groups.com, had to switch Tamil encoding from Anjal to TSCII and then recently to Unicode. The archives were however left in their original encoding. ForumHub.com, a pioneer in the web based, public bulletin board successfully migrated from TSCII encoding to Unicode including its archives. Similarly, the webzine Thinnai.com switched from TAB encoding to Unicode successfully including its archives. Project Madurai started with TSCII encoding and it now has files both in TSCII and Unicode encodings. Major encoding conversions seldom go without errors. Initial conversions from
TSCII to Unicode at the Project Madurai site had some minor errors (orphaned characters with dotted circle) which appear to have been cleaned up after user feedback.

The founder of the Agathiyar e-mail group, Dr. Jayabharathi, a Tamil scholar and an author from Malaysia, chose to convert all of the Agathiyar postings in the YahooGroups to Unicode and host them separately at his personal site treasurehouseofagathiyar.net. However, he lost most of the articles that he had hosted at various URLs at the geocities.com site when Yahoo shut down Geocities. Several of the pioneering Tamil mailing lists including Agathiyar were once hosted at the hosting site coolist.com and lost their archives when coolist.com folded. The hosting site eGroups.com was acquired by Yahoo and redesigned as YahooGroups.com. Without Yahoo’s rescue, the pioneering Tamil mailings all would have lost all of their archives had eGroups.com folded like coolist.

Of equal concern is the information deluge [3]. With the advent of Twitter and Facebook, there is an explosion of public digital content that is rich with information. However, our ability to create digital information far outstrips our capacity and infrastructure to store, manage and preserve it over time.

2.2 Why is preservation important?

Tamil scholars know very well the importance of preservation of manuscripts and inscriptions in the service of cultural heritage. In the 19th century, critical works of Tamil literature were saved from certain destruction just in time by U. Vē. Cāmiṇṭāṭaiyer and Ci. Vai. Tāmōṭaraṭā. And yet, they were unable to recover some major works such as Vaḷaiyāpati and Kuṇṭalakēci.

The examples cited for preserving the Tamil digital documents are not unique. This experience is common to English language based sites as well. The shutdown of GeoCities.com did bring down a major collection of articles in English. The U.S. Library of Congress reports that 44% of the sites available on the Internet in 1998 had vanished one year later. Another study says that 27 months after publication, up to 13% of online cited sources are irretrievable [4] leading to link rot, the process by which the collection of links on a website gradually point to web pages that have become permanently unavailable. Some of the U.S. Census data are reported to be inaccessible due to rapid obsolescence of hardware and software formats. The data recorded by two Viking space probes sent to Mars on magnetic tapes got corrupted despite being kept in a climate controlled environment. Individual consumers are not immune either. Their personal collection of digital photos, documents, e-mails on their floppy disks, portable drives, backup tapes, etc., are equally vulnerable. A lot of consumers have audio cassettes, video tapes in BetaMax and VHS formats that are becoming unusable as machines needed to play them are no longer available.

As the examples cited demonstrate even in the short duration of Tamil digital information’s existence, it has experienced the challenges forecast for global digital information storage. This is only going to get more challenging when e-Governance projects are executed in Tamil and millions of Tamil users access e-Governance kiosks to create Tamil documents. The need for planning digital preservation is great.

3. Long-term Digital Preservation

What is long-term digital preservation? And how long is “long-term”? “Long-Term Digital Preservation (LTDP) is a means of keeping digital information such that the same information can be used at some point in the future in spite of obsolescence of everything: hardware, software, processes, format, people, etc.” [5] “Long-term” is taken to be the case where it is impossible for a writer to converse with an
eventual reader and impossible for the reader to clarify uncertainties by asking the writer or the writer’s contemporaries. [6] In other words, if you write a will bequeathing your wealth to your grandchild and save it in a CD, your grandchild should be able to read what you wrote and prove it to others when you are no more. Your grandchild should do so even if the CD drive, the PC, the Windows operating system, CD driver, the software that reads the contents of the CD and the computer monitor all familiar to you are no longer available and even the CD media itself may have failed. It is not easy. It is not cheap. It will require a lot of organization. And in the end, in spite of best efforts it may not be possible to satisfy all the requirements.

3.1 Challenges of Digital Preservation

Unlike analog data like an old black and white photo, digital data does not degrade gracefully. It requires a specific environment, if not an exact one, to be accessible. Digital files on a medium need specific device drivers to access the medium, specific operating systems to understand the file system and launch a specific application program that knows how to interpret the content of the file. These further depend on specific hardware. If any of these fail, the digital data are not accessible. Digital Preservation aims at packaging the digital objects such that their authenticity is provable and the data are accessible as long as there is a need to read that object. Three main methods of digital preservation are technology preservation, technology emulation and information migration.

3.1.1 Technology Preservation

This is the “computer museum” solution that assumes that the computer/equipment, operating systems, original application software, media drives, etc., all can be preserved, and maintained and that the media is readable. This is the method that was used to recover the pictures taken in 1966 by NASA’s robotic probe Lunar Orbiter 1 and stored on 2500 tapes that needed a specialty FR-900 Ampex tape drive, very few of which were made and sold at the cost of $330,000. It took considerable effort by the very few technicians who had the special skills to work with the reader, a lot of salvaged parts and some good luck. Since there is no commercial or practical interest in such equipment, it is impossible to keep such obsolete equipment functional forever. It worked in this case because of the rarity of the digital object that people were trying to retrieve. [7]

3.1.2 Technology Emulation

Emulation uses a special type of software, called an emulator or virtual machine, to translate instructions from original software to execute on new platforms. This is a fairly well understood technology and practical applications do exist. For example, the popular software VMware offers legacy emulation mode to read DVD-ROM or CD-ROM drives. However, it is difficult to predict how many generations of such emulation will be possible. And for very long-term preservation of the order of several decades to a century or more, technology emulation is yet to be proven.

3.1.3 Information Migration

Information migration requires periodic transfer of digital objects from one system software configuration to another or from one generation of computer technology to a subsequent generation with no loss in content or context and little or no loss in structure. This may require that encrypted digital objects be decrypted. Careful processes need to be evolved to guarantee object authenticity and integrity with this method.
Some of the early creators of Tamil digital objects used a migration method to transfer digital data from MS-DOS based applications and file formats across multiple generations of Windows operating system releases and several versions of applications. Even on the web, users migrated data from one provider to another, as well as from one encoding to another using the conversion tools available. As the Government of Tamil Nadu embarks on the e-Governance program, its offices across the state will be migrating their existing digital data files from various encodings such as TAB, TAM, Baamini, Vanavil etc., to Unicode. This migration of information avoids obsolescence not only of the physical storage medium, but of the encoding and format of the data.

In a sense, the Tamil scribes who have been migrating palm manuscripts in Tamil across the centuries have been doing precisely that and have been largely successful in transmitting heritage information to later generations. While this is still error-prone across very long time scale, it is a proven method. This requires that digital objects be constantly migrated to use new storage media, software and computers whenever a current system is in danger of obsolescence.

Dr. Dinesh Katre describes a Loop shaped Life Cycle Model for Digital preservation and how the cycle repeats as obsolescence kicks in. (See Figure 3 and Figure 4) [9]
3.1.4 Distributed Digital Preservation

One of the ways in which the authenticity and integrity of digital objects can be preserved is by Replication. Multiple copies also improve the longevity of digital objects if multiple storage locations are used as well. This technique is also known as Distributed Digital Preservation (DDP). This requires extreme discipline, high organization and dedication while the risk of losing authenticity and integrity is high unless one is careful. Historically, this method has been used successfully by religious orders, monks and traditional scholarly communities dedicated to transmitting specialized knowledge across generations. It is noteworthy that a lot of the Tamil palm manuscripts were found in Saiva and Vaishnava mutts, traditional medicinal families, and traditional Tamil scholar families.

A classic example of DDP is the LOCKSS (Lots of Copies Keeps Stuff Safe)[8] technology used by a consortium of publishers, librarians, and learned societies to support a community-managed failsafe repository for scholarly content (http://lockss.stanford.edu/).

3.1.5 Metadata and Digital Preservation

Metadata (data about data) is critical to the authenticity and integrity of digital preservation. Metadata is concerned with content, context, and structure of the digital data. Content is intrinsic to the digital information and refers to what the object contains. Context indicates who, what, why, where, how and other data associated with the object's creation and is extrinsic to the object. Structure describes the associations within or among individual objects and can be either intrinsic or extrinsic.

The key to successful implementation of a digital preservation effort are metadata.

There are three different types of metadata, all essential to ensure the usability and preservation of the collection over time.

Descriptive Metadata: convey some sense of intellectual content and context.

Structural Metadata: describe the attributes of an object, such as size, electronic format etc.

Administrative Metadata: information related to rights management, creation date of digital resource, hardware configuration etc. [10]

4. Current Standards and Best Practices

There are several standards that address digital preservation that are of interest to us. The Open Archival Information System (OAIS) was initiated by Consultative Committee for Space Data System (CCSDS) in 2002 and became an international standard with ISO 14721 : 2003. This is the reference model for several other standards. LOCKSS and CLOCKSS comply with the OAIS model and are popular with digital librarians and research publications. The Victorian Electronic Record Strategy (VERS) is a practical specification that addresses the immediate problems of handling large digital documents in popular formats. The National Electronic Records Archive (NARA) consolidates several models with an aim to preserve government records. There are several digital preservation initiatives in India as well, with Digital Library of India (DLI) being the one that digitizes books in Indian languages. There doesn’t appear to be any special program anywhere in the world that addresses the unique challenges of Tamil digital documents.
5. Issues of Interest for Tamil documents

Tamil digital documents are constantly being created not only in mainland Tamil Nadu but also in Eezham and the vast Tamil diaspora. While the diaspora largely shares common interests in classical Tamil, when it comes to modern culture, literature, politics, cinema, lifestyle etc., it is quite diverse and differs significantly from the mainland Tamils. Not all of the political opinions of the diaspora are appreciated or even tolerated everywhere. As a result, unpalatable opinions or culturally variant expressions of the diaspora are likely to be ignored if not outright censored by a conservative digital librarian. Selection of the digital content to be preserved is a matter of judgment and since it is expensive to preserve digital content for a long time, it would be tempting to reject the content that one disagrees with regardless of how representative such content may be of contemporary Tamil culture. The digital librarians need only to look at the anthologists that put together the Caṅkam anthologies that have stood the test of time as a guide for selection.

There are some interesting lessons to learn from the fact that the Indus Civilization seals are yet to be satisfactorily decoded and how the Rosetta Stone helped decode Egyptian hieroglyphic writings. The Rosetta Stone had carved text made up of three translations of a single passage: two in Egyptian language scripts (hieroglyphic and Demotic) and one in classical Greek. This is a classic example where Arthur C. Clarke’s Rule of Three worked. As interpreted by Mike Shea (mikeshea.net) this rule becomes “preserve in three formats on three media types in three places”. Extending this to Tamil documents, one could save important Tamil documents in three different encodings (Unicode, TACE-16 and ISO-15919), in three different formats (PDF/A, XML, HTML) in three different places (Tamil Nadu, Malaysia and Canada). Since the Unicode Tamil block could be used to encode texts in Sanskrit and Saurastri and ISO-15919 could be used to encode all Indic languages, only TACE-16 encoding exclusively encodes Tamil. The subscript/superscript modifiers to indicate varga sounds of other Indic languages do not mix well with TACE-16. So, this could serve as the Rosetta stone for a future researcher if there is any doubt as to the identity of the language of a Tamil digital document. Since the Māṇkuḷam Tamil brahmi inscriptions were thought to be a prakrit inscription for a long time, this is not an idle speculation. The metadata also should be used to clearly identify the language and encodings used to avoid any misidentification. Since such triple encoding preservation is expensive this should be considered only for important documents.

Conclusions

The reality is that solution to the LTDP problem is a work in progress. Some acceptable solutions exist for static data such as documents where the current process is to store in standard encapsulated formats such as PDF/A, wrap with metadata as defined by OAIS and migrate when necessary to address media obsolescence and technology following the CDAC-Katre multi-loop life cycle model. Even here, critics are unsure of the selection of PDF/A for an encapsulated model, preferring HTML or XML. More research is being done to understand the preservation requirements for dynamic data.

We hope that this paper sparks an interest in the long-term preservation and archival challenges to preserve Tamil digital information. The Tamil Virtual University could be one of the best Digital Librarians for preserving all Tamil and Tamil related digital documents and could be one of the nodes of a distributed archival center, perhaps using the LOCKSS technology.
Since large number of Tamil digital documents exist with multiple data formats and encoding, it is hoped that accurate metadata is specified to encapsulate these documents in their original format before migrating them to one standard encoding or three standard encodings (Unicode, TACE-16 and ISO 15919) as discussed earlier.

We hope that this paper would spark an interest in and stimulate in-depth discussions among those interested in very long-term preservation of Tamil digital documents.

References
திருச்சபதிப்பில் திரும்பி பிள்ளா ஆக்சியில்

சுதந்திரம

மாணவர்கள் பல்வேறு வகையில், ஒன்றுகொண்டு வருவதில் செய்ததைத் தெரிவு செய்தம். ஆனால் முதல் வருவதில் தள்ளி நிற்கவும் உள்ள காலத்தில் ஆக்சியில் செய்ததை எதிர்க்க வேண்டும்.

இவ்விற்கு முன்னோடியாக முதல் வருவதில் செய்ததைக் கூற்று வைக்க வேண்டும். முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றால் முதல் வருவதில் செய்யப் பட்டுள்ள வேளாண்மை என்றும் தொன்றும். இதன் மூலம் மாணவர்கள் முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றும் தொன்றும்.

மாணவர்கள் தொடர்ந்து வந்து முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றும் தொன்றும். இதன் மூலம் மாணவர்கள் தொடர்ந்து வந்து முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றும் தொன்றும்.

மாணவர்கள் தொடர்ந்து வந்துக்கும் போது தொடர்ந்து வந்து முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றும் தொன்றும்.

மாணவர்கள் தொடர்ந்து வந்து முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றும் தொன்றும். இதன் மூலம் மாணவர்கள் தொடர்ந்து வந்து முதல் வருவதில் செய்ததை எதிர்க்கவும் உள்ள காலத்தில் செய்யப்பட்டுள்ள வேளாண்மை என்றும் தொன்றும். 

428
தொலைக்காட்சி கல்வியில் கவரும் கருத்துடன் அடங்கியியல் படையும் புது எழுத்தராக்கியதற்காக குறிப்பிட்டுள்ளார். இப்பட்டியல் வாய்ப்பாட்டில் புதிதானது விளக்கமான அம்மாள் பிரிவு, பிரிவு அடங்கியியல் குறிப்பிட்டுள்ளார். இப்பட்டியலில் குறிப்பிட்டுள்ள தமிழ் கல்வியும் விளக்கமான குறிப்பிட்டுள்ளார் பிரிவு அம்மாள் குறிப்பிட்டுள்ளார்.


4. TAMIL COMPUTING WORDS , TAMIL VIRTUAL UNIVERSITY DICTIONARIES, ANNA UNIVERSITY ENGLISH-TAMIL COMPUTER DICTIONARY.


6. Digital Dictionaries of South Asia
1. ஐட்ப் (p. 40) [ a| i ], s. gift, எல்லாம்; 2. favour, அதம்;
3. desire, ஐதம்; 4. love, ஐதம்; 5. civility, ஐதம்; 6. poverty, ஐதம்;
7. unripe fruit, ஐதம். அருங்கள், one who is kind; 2. a paupe

7. Google English – Tamil dictionary and Google Tamil – English dictionary

8. Online Tamil Dictionary

Online Tamil Dictionary, - www.tamildict.com/
Tamil dictionary, - www.tamil.net/learn-tamil/tamildic.html
English to Sinhala and Tamil Online Dictionary from Sri Lanka
www.lanka.info/dictionary/EnglishToSinhala.jsp

Dictionaries www.multilingualbooks.com › ... › Online Dictionaries

Tamil English
**dictionary** and **English Tamil dictionary** - FREELANG  www.freelang.net/dictionary/tamil.php

**Tamil English** and **English Tamil Dictionary** Free Online Translation  www.ats-group.net › Languages › Dictionaries  
**Tamil and English dictionary** - dsal.uchicago.edu/dictionaries/fabricius/ - Dravidian Languages  www.yourdictionary.com › Languages  
Sanskrit, **Tamil** and **Pahlavi Dictionaries** - webapps.uni-koeln.de/tamil/  
**alphaDictionary** * Free **Tamil Dictionary** - Free Tamil Grammar  www.alphadictionary.com › Languages  
› Dravidian –

---

**bottle in Tamil** - Google Dictionary  

<table>
<thead>
<tr>
<th>Bottle</th>
<th>Tamil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottle</td>
<td>மாளி</td>
</tr>
</tbody>
</table>

**Related phrases**  
- aspirator bottle  
- magnetic bottle  
- bottle opener  
- bottle feeding  
- blue bottle  
- bottle neck  
- ink bottle  
- bottle-shaped jack  
- green bottle  
- bottle stopper  

**Synonyms**  
noun: flask  

**Usage examples**  
"Since I had to leave the bottle behind, I asked that it be destroyed. At no point did I request preferential treatment; it's not in my nature," Blockum said in his statement.  
Mar 15, 2010 - Jean-Pierre Blackburn - *Tennesse Star*

"I have repeatedly noted that we haven't turned any corners," Petraeus said. "We haven't seen any lights at the end of the tunnel. The champagne bottle has been pushed to the back of the refrigerator. 
And the progress, while real, is fragile..."  
Apr 5, 2008 - David Petraeus - *Boston Globe* (registration)

Winsler, who has been nominated five previous times, said: "I'd be lying if I said I haven't made a version of this speech before. I think I was probably 6 years old and staring into the bathroom mirror, and this would have been a shampoo bottle..."  
Feb 22, 2009 - Kate Winsler - *ABC News*

**Web definitions**  
- A glass or plastic vessel used for storing drinks or other liquids; typically cylindrical without handles and with a narrow neck that can be...

---

http://www.google.com/dictionary?langpair=ta%7Cen&q=bottle&hl=en&aq=f  
4/10/2010
9. என்னும் தில்வாயில் எந்த படி நல்லது என்று அறிக்கையியல் உறுப்புகள்?

‘தற்போது’ என்பது நீண்டுருக்கு பிற்று வேளுடன் அவர்களே விளக்கம் செய்துள்ளன. அவர்கள் விளக்கத்தை இட்டு வரும் படி போன்று அவர்கள் இல்லையால் தொகுதி. விளக்க வடிவ அங்கிலம், அதன் முன்னோடியாக வரும் படி விளக்கம், அதன் முன்னோடியாக வரும் படி விளக்கம் மேல். இல்லையால் மேல் விளக்கம் விளக்கத்தைக் குறிக்கவும் அல்லது வேண்டிய விளக்கத்தைக் குறிக்கவும்.

Utilities

Transliteration

Dictionary

Select Language Tamil

English Word ????? Search There is no result found

10. பல்வேறு வரைபாடுகள் என்று எந்த படி விளக்கப்பட்டது?

tamil online Tamil Wiktionary
DSAL dictionaries J. P. Fabricius’s
Tamil and English dictionary Tamil Moli Akarathi. N.
Tamil-Tamil akarathi. A core vocabulary for Tamil
Tamil lexicen. A comprehensive
Tamil and English dictionary of high and low Tamil Pals e-dictionary
Tamil Virtual University Dictionaries கன்னடம் குறிப்பிட்டிய.
தமிழ் dictionary and Google Tamil – English dictionary Google English – Tamil
1. Google Transliteration IME

New! Download Google Transliteration IME Type a word in English and press SPACE to transliterate. Press CTRL+G (⌘ G on Mac) to switch between English and the selected language.

Dismiss

©2009 Google - Font Guide - Discuss - Help - Google Home

Define

Translate

From Tamil To English swap

1. dream
2. morpheus

Powered by Google Dictionary

‘தமிழ்’(google) குறிப்பிட்டு ஒரு குறிப்பிட்டு பதிமப்படுத்தும் மறைய இல்லை பதிமப்படுத்தும் மறைய இல்லை. இல்லை பதிமப்படுத்து செய்யவுண்டு பதிமப்படுத்து செய்யவுண்டு. இல்லை பதிமப்படுத்து செய்யவுண்டு பதிமப்படுத்து செய்யவுண்டு.

434
2. http://translate.google.com/toolkit/list#translations/active

Google Translator Toolkit is a free online Tamil language typing software. Google Translator Toolkit helps you to type in Microsoft Excel.


You can cut and paste these Unicode Tamil words from Tamilpad into your email messages such as Yahoo or Hotmail or into any editor such as Microsoft Word or Microsoft Excel.


TamilCube's Tamilpad is a free online Tamil language typing software. Tamilpad helps you to type in Unicode Tamil. Just type in English in the left box, by following the Transliteration keyboard mapping given below. The equivalent Unicode Tamil letters appear automatically in the right box. For example, when you type 'VANAKKAM' in English and hit the space bar, Tamilpad will convert it directly into Unicode Tamil script as 'வணாக்கம்'. You can cut and paste these Unicode Tamil words from Tamilpad into your email messages such as Yahoo or Hotmail or into any editor such as Microsoft Word or Microsoft Excel.
http://www.tamilcube.com/dictionary

Enter your English or Tamil word for translation in the search box below and click 'SEARCH'.

This field is required.

- English -> Tamil
- Tamil -> English
- Number -> Tamil word

Modern Online Tamil dictionary (English-Tamil & Tamil-English)

Browse for basic words in online Tamil dictionary: ABCDEFGHIJKLMNOPQRSTUVWXYZ

Tamilpad - TamilCube's Free Online English to Tamil Transliteration tool
TamilCube's Free Online English to Tamil Dictionary and English to Hindi Dictionary
English to Tamil and Tamil to English professional Translation Service

6. இதை மொழிபொழிவையும் கொண்டு நேர்ந்து 

Tamil translation dictionaries - lexicoo...

Online Tamil bilingual and multilingual dictionaries (Tamil <-> English, French, etc). List updated regularly.

www.lexicool.com/dictionaries_tamil.asp [Found on Google, Bing]

Tamil translation from to English tamil ...

All kind of Tamil translation and tamil language related works are done in time according to your needs.

www.tamiltranslator.com/ [Found on Google, Bing]

English to Tamil translation

Your search for English to Tamil translation returned 142 results in the following ..... Translation of Christian Manuscript from English to Tamil ...

www.translatorscafe.com/cafe/PopSearch/English-to-Tamil-tran... [Found on Google, Bing]

Tamil English and English Tamil Dictionary...

Free Online Dictionary Welsh English and Free Online Translation Tamil English Dictionaries.

www.ats-group.net/dictionaries/dictionary-english-tamil.html [Found on Google, Bing]

Tamil Translation Service

Tamil Translation services from Applied Language Solutions high quality, professional, award winning Tamil Translation

www.appliedlanguage.com/languages/tamil_translation.shtml [Found on Bing]

Professional Tamil translation service |...

One-stop language resources portal. Professional Tamil translation service. Free Tamil books, Tamil mobile eBooks, Tamil test papers.

www.tamilcube.com/ [Found on Google]

English to Tamil Translation
English to Tamil language translation service ... Superb English to Tamil translation. In need of a quality Tamil to English or English to Tamil translation?

www.kwintessential.co.uk/translation/to/tamil.html [Found on Bing]

tamil translation - Google Chrome Help

Feb 10, 2010 ... Among the Indian language versions, is Google Chrome also being targetted only for Hindi-knowing people only? Why is there no Tamil option? ...

www.google.com/support/forum/p/Chrome/thread?tid=186d352aa90... [Found on Google]

Tamil translation, English to Tamil tran...

wintranslation.com provides professional Tamil translation service performed by human translators.

www.wintranslation.com/languages/tamil.html [Found on Bing]

Internet Archive: Free Download: QURAN T...

QURAN TAMIL TRANSLATION NOT ARABIC MP3 This audio is part of the collection: Open Source Audio Artist/Composer: QURAN TAMIIL. Keywords: ALLAH islam Muhammed ...

www.archive.org/details/tamilquranmp3 [Found on Google]

English Tamil dictionary - lexicool.com

English Tamil dictionaries ... by language Online dictionaries by subject Lexicool blog Lexicool newsletter Translation courses ...

www.lexicool.com/online-dictionary.asp?FSP=A09B31 [Found on Bing]

Tamil Translation - Free with Linguanaut

Our website Linguanaut helps you get free Tamil translation from our translator volunteers, like how to say hello, welcome, thank you, other greetings and ...

www.linguanaut.com/translation_tam.htm [Found on Google]

Tamil translation English Tamil translat...

Translation India offers Tamil translation and English to Tamil translation services. Get documents, technical, legal, financial and book translation in Tamil language by ...

translationindia.com/indian-languages-tamil.html [Found on Bing]

Tamil - Google & BabelFish translation i...

Thanks Tamil. I also want to add Google Translation (site translation) of member Apropos. I added the line -Submenu, Google Translate, Google Translate ...

my.opera.com/Tamil/blog/google-babelfish-translation [Found on Google]

Amazon.com: Understanding Muhammad: A Ps...

2. தமிழியல் மொழிப்பிட்டு மொழியடைவு அரங்கு

http://www.tamilcube.com/dictionary/mobile/

Download Mobile Dictionaries

TamilCube welcomes you to the world of Modern Mobile Dictionaries! You can download English-Tamil, English-Malay and English-Hindi dictionary software for mobile to any of your mobile phones with Java support, such as any make and model of mobile phone (cell phone), PDA, iPhone and Blackberry. To view the mobile dictionaries in your mobile phone, just follow the simple steps below:
1. Download the zip file for the dictionary you like, unzip into a folder in your PC.
2. Connect your mobile phone, iPhone or Blackberry to the PC and transfer the unzipped software files to your mobile. (This method is the same as how you download games or applications to your mobile device.) Now you are ready to enjoy referring to the dictionary from your mobile phone, anytime and anywhere!
The trial mobile dictionaries contain only the words starting with letter "a".

To view the Tamil and Hindi mobile dictionary, your mobile phone, iPhone or Blackberry must have Unicode support.

When searching the dictionary, use lower case English letters.
வெற்றிகள் கூடிய (Google) இல்லைசையும் கல்வியைச் செய்யவும் பாதுகாக்க வேண்டும். தக்கவருடைய தலை தாழ்த்தும் போர்விளையாக்கும் ஆராய்ச்சியால் அங்கீகாரம் கிடையாது. இது போர்விளையாக்கும் கல்வியைக் கவர்கிறது பேச்சக்கோளமும். அத்துடன், எந்த தேர்வுத்தேர்வு?

பிறப்பு:
கூடிய கூடிய, இல்லைசையும் இல்லைசையும் மேம்படுத்த ஆராய்ச்சி போர்விளையாக்கும் பாதுகாப்பு விலைத்திறன் ஆதி எதிர்முனை. இது போர்விளையாக்கும் நினைவு அநூற்றுத்தோன் போர்விளையாக்கும் நீர்மில்வு என்று தொடர்ந்து வேண்டும் குறிப்பிட்டிக்கொள்ளலாம். இது போர்விளையாக்கும் நீர்மில்வு என்று தொடர்ந்து வேண்டும் குறிப்பிட்டிக்கொள்ளலாம். ஆனால், தக்கவருடைய தலை தாழ்த்தும் போர்விளையாக்கும் என்று தொடர்ந்து வேண்டும் குறிப்பிட்டிக்கொள்ளலாம்.
நிலக்கோட்டின் நிறுவ கால்கள்

புதுக்கோட்டை காலநிலைகள்

புதுக்கோட்டை நிறுவனப் பதிலக் காலநிலையான வரலாறு பற்றி பல்வேறு பல பார்வைகளும் கூறப்பட்டுள்ளது. புதுக்கோட்டை காலநிலைகள் வரலாறு பற்றிய தொல்பொருள்களில் பல பார்வைகள் காணப்பட்டுள்ளன. ஆனால் முதல்நிலைகள் முதல் கடைசி நிலைகள் பற்றிய பல்வேறு பார்வைகள் காணப்பட்டுள்ளன. ஆனால் முதல்நிலைகள் முதல் கடைசி நிலைகள் பற்றிய பல்வேறு பார்வைகள் காணப்பட்டுள்ளன.

முதல் பாலாக் பாத்திரக் கால்கள்

முதல் பாலாக் பாத்திரக் கால்கள் முதல் கடைசி காலநிலைகள் பற்றிய பல்வேறு பார்வைகளும் காணப்பட்டுள்ளன. புதுக்கோட்டை நிறுவனப் பதிலக் காலநிலையான வரலாறு பற்றிய தொல்பொருள்களில் பல பார்வைகள் காணப்பட்டுள்ளன. ஆனால் முதல்நிலைகள் முதல் கடைசி நிலைகள் பற்றிய பல்வேறு பார்வைகள் காணப்பட்டுள்ளன. ஆனால் முதல்நிலைகள் முதல் கடைசி நிலைகள் பற்றிய பல்வேறு பார்வைகள் காணப்பட்டுள்ளன. ஆனால் முதல்நிலைகள் முதல் கடைசி நிலைகள் பற்றிய பல்வேறு பார்வைகள் காணப்பட்டுள்ளன. ஆனால் முதல்நிலை�ள் முதல் கடைசி நிலைகள் பற்றிய பல்வேறு பார்வைகள் காணப்பட்டுள்ளன.
பரத

443
2006 ஆம் ஆண்டில் செயல்பாடு ஆக்கப்பட்டது வரை குடியாரர்கள் நம்பிக்கை இயற்பியல் பயன்பாடு பெற்றுக் கொண்டுள்ளனர். அதன் பின்னர் பிற்பகுதி யாரும் முடிக்கும் மதிப்பு வாய்ந்தது. குடியாரர்கள் தம்மிடம் பராமரிக்கும் குடியியல் பயன்பாடுகளை விளக்கத்துக்கும் பங்களித்தனர். குடியாரர்களும் இந்த பங்களி பெற்றுக்கொண்டனர்.

அதன் பின்னர் தம்மிடம் குடியர்கள் நம்பிக்கை வாய்ந்தது. குடியியல் குடியை விளக்கத்துக்கும் பங்களித்தனர். அதன் பின்னர் தம்மிடம் குடியர்கள் நம்பிக்கை வாய்ந்தது. குடியியல் குடியை விளக்கத்துக்கும் பங்களித்தனர். அதன் பின்னர் தம்மிடம் குடியர்கள் நம்பிக்கை வாய்ந்தது. குடியியல் குடியை விளக்கத்துக்கும் பங்களித்தனர்.

2006 ஆம் ஆண்டில் செயல்பாடு பெற்றுக்கொண்டது வரை குடியாரர்கள் நம்பிக்கை இயற்பியல் பயன்பாடு பெற்றுக்கொண்டுள்ளனர். அதன் பின்னர் பிற்பகுதி யாரும் முடிக்கும் மதிப்பு வாய்ந்தது. குடியாரர்கள் தம்மிடம் பராமரிக்கும் குடியியல் பயன்பாடுகளை விளக்கத்துக்கும் பங்களித்தனர். குடியாரர்களும் இந்த பங்களி பெற்றுக்கொண்டனர்.
கேட்டு என்று விளக்கம் அளிக்கும் நுனியா

3 Chennai Express விளக்கம் முடிக்கத்தை நிறைந்து மாணவர்களின் 1994 க்குள் இருந்து நீர்மையான விளக்குமுறை என்று விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் விளக்கம் வி
பிள்ளாயின் அதிகாரக்குறிக்கை

(Electronic Museum)

பிள்ளாயின் ஆராய்ச்சிகள்

பிள்ளாயின் ஆராய்ச்சிகள் ஆவார்ய கிருஷ்ணவாகுமாராயின் வாசனை அறிக்கட்டுப்படையில் உள்ள பெரும்பாலும் ஒருநாள் விளக்கத்தால் தொடங்கி விளக்கப்படுகின்றன. பிள்ளாயின் ஆராய்ச்சிகள் என்றும் புதிய நிறுவக நிறுவனங்கள் கூறுகிறது. பிள்ளாயின் ஆராய்ச்சிகளின் முன் புதிய நிறுவகங்கள் கூறுகிறது.

பிள்ளாயின் ஆராய்ச்சிகள் - அளவுபோற்ற

பிள்ளாயின் ஆராய்ச்சிகளின் முன் புதிய நிறுவகங்கள் என்பது புதிய நிறுவகங்களில் குறிப்பிடப்படுகின்றது. பிள்ளாயின் ஆராய்ச்சிகள் என்றும் புதிய நிறுவகங்களின் முன் புதிய நிறுவகங்கள் கூறுகிறது.

அ) பிள்ளாயின் ஆராய்ச்சிகள் என்பது புதிய நிறுவகங்களின் முன் புதிய நிறுவகங்களின் குறிப்பிட்டு பிள்ளாயின் ஆராய்ச்சிகள் என்றும் புதிய நிறுவகங்களின் முன் புதிய நிறுவகங்களின் குறிப்பிட்டு பிள்ளாயின் ஆராய்ச்சிகள் என்றும் புதிய நிறுவகங்களின் முன் புதிய நிறுவகங்களின் குறிப்பிட்டு பிள்ளாயின் ஆராய்ச்சிகள்.

447
http://www.chennaimuseum.org/draft/index.htm

isoranak aru athenasieikilek usulakanath ekeka, unapravieie sheni, karikarikaleluk, atmavirapareekul, sarakalee kakan, maruittu karthapam, pitupitukal, cekireeppilek, manaral atmavirapareekul, karukalee athenasieikilek usulakanath. Eka karalagaren athenasieikilek usulakanath, kakan karukalee sarakalee kakan, pitupitukal, manaral atmavirapareekul. 25/04/2010 elo ekkil athenasieikilek usulakanath 2,84,611.
http://virtual-museum-india.blogspot.com


http://ungeslaboratoriumforkunst.dk/index.asp?key=1

http://www.ancientgreece.co.uk/menu.html

449
http://www.nhb.gov.sg/WWW

2010 வரை தமிழில் எச்சாரம் கொண்டு இயங்கிய பல்வேறு வகைகள் கொண்ட பொருள்கள் பராமரிக்கப்பட்டன. இதன் காரணமாக விளையாட்டுக்கு முன்னேற்பது நூற்றாண்டுகள் வரை குறிப்பிட்டது.

http://www.museevirtual-museum.ca/index-eng.jsp

சோட்டுதல் கூட்டம் வழக்கத்தில் செய்யப்பட்டது. இது மறைந்த மக்கள் ஓர் அதிகாரிகளின் மக்களைச் சார்ந்த பொருள்கள் பராமரிக்கப்பட்டன. 2010 வரை தமிழில் எச்சாரம் கொண்டு இயங்கிய பல்வேறு வகைகள் கொண்ட பொருள்கள் பராமரிக்கப்பட்டன. இதன் காரணமாக விளையாட்டுக்கு முன்னேற்பது நூற்றாண்டுகள் வரை குறிப்பிட்டது.

5) எச்சாராயிலான கலாச்சாரான் புனிதிகள் நிறுவலிதல்

அதிகாரிகளின் வாய்ந்த பராமரிக்கப்பட்டதுக் கலாச்சாரான பேருந்துக்கள், பிரித்தனர் கலாச்சார்கள் புருஷர் பெண்கள் மக்கள் வரிப்பர். உலகமும் வெளியும் வில்லையாளர்களும். கலாச்சார்களும் புனிதிகளும் கலாச்சாரான்கள் வரிப்பர். அதிகாரிகள் செயல்பட்டார் வெளியும் வில்லையாளர்களும். கலாச்சார்களும் புனிதிகளும் கலாச்சாரான்கள் வரிப்பர். அதிகாரிகள் செயல்பட்டார் வெளியும் வில்லையாளர்களும். கலாச்சார்களும் புனிதிகளும் கலாச்சாரான்கள் வரிப்பர்.

நிலையம் புனிதிகள் வரிப்பட்டதுக் கலாச்சாரான்கள் வெளியும் வில்லையாளர்களும். அதிகாரிகளின் வாய்ந்த பராமரிக்கப்பட்டதுக் கலாச்சாரான பேருந்துக்கள் வரிப்பர். வெளியும் வில்லையாளர்களும். கலாச்சார்களும் புனிதிகளும் கலாச்சாரான்கள் வரிப்பர். அதிகாரிகள் செயல்பட்டார் வெளியும் வில்லையாளர்களும். கலாச்சார்களும் புனிதிகளும் கலாச்சாரான்கள் வரிப்பர்.
Virtual Museum - Electronic Museum
Online Museum - Virtual Museum
Virtual Tour - Digital Museum
Web Presentation - Site Map
Feedback - Geology
Chemical Preservation - VRML
Virtual Reality Modeling Language - 3D Images
Animated Images - Electronic multimedia guide
Ipod
நமது கதிரிப்புற்று நம் தேர்தல் கலாந்தராத்திற்கும்

ஆண்டவர், முதுகக்கல் மறுவல் இல்லடு இல்லடு (http://www.muthukamalam.com)
19/1, காமா லிங்க பூனைவில் புழு, இல்லடு - 625 531, தொல்லூர், இல்லடு. மைணர்பிளே:n smuthukamalam@gmail.com

நவீகர்வ

2001 ஆம் ஆண்டில் நகர வல்லாம்புக்குத் தெரிந்து கணக்கெடுப்பால் நவீகர்வ நடைபெற்றது. இந்த நவீகர்வத்தை அறிமுகப்படுத்திய நாள் மீது வந்து நவீகர்வ நகர வல்லாம்புக்கு வெளியிட்டு விளக்கம். இந்த நவீகர்வம் பல்வேறு முக்கிய வேலைகளை முடித்துள்ளது. இந்த நவீகர்வம் முழுமையாக புரோட்டுக்கும் பைத்துவின் முக்கியத்துவம் கொண்டது. இந்த நவீகர்வம் முக்கிய வேலைகளை செய்து வந்து பல்வேறு கவர்களை நாட்களிலென வருகின்றது.

மின்கலிப்புற்றுத்து திறக்கும்

2001 ஆம் ஆண்டில் உருசியாவில் அனைத்து நாடங்களும் உருசிய முனையுமாக தொடங்கிய முக்கியமான வேலைத்தடத்தை நடைபெற்றது. இந்த நிறுவனம் தொடங்கியது ஒரு உருசிய நாடு முனையும் தொடங்கியது. முனையாக உருசிய நாட்டின் வேலை வல்லாம்புக்கு வந்து நவீகர்வம் நடைபெற்றது. இந்த மின்கலிப்புற்றுத்து திறக்கும் நோக்கத்திற்கு தமிழ் நாடுகளின் வேலைக்கு வேறு பல்வேறு முக்கிய வேலைகளும் முடித்துள்ளது. இந்த மின்கலிப்புற்றுத்து திறக்கும் வேலைகளும் முக்கியமாக வருகின்றன. இந்த மின்கலிப்புற்றுத்து திறக்கும் வேலைகளுக்கும் இணைந்து நாட்டிலேயே 267 நூற்றுக்கடி மின்கலிப்புற்றுத்து திறக்கும் வேலைகளை நடைபெற்றது.
2003 ஆம் ஆண்டில் ஒன்றியம் 30 ஆம் பிரிவில் “பல நிறைவு” சார்ந்த குழுப்பியில் ஒரு இளம் குழல் லோகியல் நிலை பெற்றது. அல்லது பலவற்றில் ஒன்று குறிப்பிட்டு விளக்கமைக்கும் முறையாகியது வேண்டும். தற்போதுபடி அனுருவையில் குழல் பொறித்து போருந்து செயல்பட்டதோ வருவது 2003 ஆம் ஆண்டில் ஒன்றியம் மறுகையில் இருந்து விளக்கமைக்கும் முறையாக வருவதாக உள்ளது. அனைத்துப் பொறித்து விளக்கத்தே வழங்குவதற்கு வேண்டும் அளவுமுறையில் வருவதாக உள்ளது.

நிலைக் குரல்கள் பன்னாட்டில்

ஆரோ யுரோப்பில் அல்லது பல நிறைவு சுற்றுலாவின் கட்டுப்பாடுகள் அறிவுபெறுகின்றன பல்வேறு பட்டியல் போருந்துகள் கூறும் குரல்கள் குறிப்பிட்டு வேண்டும். உள்ளிட்டு குழல் பொறித்து வருவது 2003 ஆம் ஆண்டில் ஒன்றியம் மறுகையில் இருந்து விளக்கமைக்கும் முறையாக வருவதாக உள்ளது.
புரிந்த நடவடி விருக்கும் செல்வாது எண்ணெய் தொடர்வது பலமக்க தொட்டுப் போது விருக்கும் குணம் தண்டார்களும் புவியாராய் புவியாராய் என்று மேலும் சொல்வது. பொழுதும் என்று குணம் தண்டார்களும் பலமக்க தொடர்வது போது விருக்கும் குணம் தண்டார்களும் பலமக்க தொடர்வது.

காரணி சொல்லுங்கள்

gுனைத் தமிழில் பலமக்க உடல்பட்டை பலமக்க பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை. என்றுப் புதியில் பிள்ளை சிற்றி என்று குணம் தண்டார்களும் பொழுதும் என்று குணம் தண்டார்களும் பொழுதும் என்று குணம் தண்டார்களும் பொழுதும் என்று குணம் தண்டார்களும். அத்துடன் தமிழில் பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை.

புதியும் விளக்கங்கள்

gுனைத் தமிழில் பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை.

புதியும் விளக்கங்கள்

gுனைத் தமிழில் பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை. என்று குணம் தண்டார்களும் பொழுதும் என்று குணம் தண்டார்களும் என்று குணம் தண்டார்களும் என்று குணம் தண்டார்களும். அத்துடன் தமிழில் பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை பலமக்க உடல்பட்டை.

454
5. கல்லறை

ஒன்றும் தமிழ் இல்லை பிறந்தெடுத்து ஒன்று இல்லையால் ஒன்று அணி ஏற்படுத்தும் பெருந்தோற்றம் கட்டமைப்பில் தமிழ் தகர்க்கு விளக்கம் செய்யவும். கல்லறை பயிற்சியில் செய்யப்பட்ட சமயரின் நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தர அடுத்து பெருந்தோற்றம் தொகுதியின் மூன்றாம் நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தர அடுத்து பெருந்தோற்றம் தொகுதியின் மூன்றாம் நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தரவு பதிலிட்டு முடிவு செய்யப்பட்ட நோய் தர அடுத்து பெருந்தோற்றம்
யார்பாட்டின்

(1.P.Number) பிரபமானம் அரசாங்கத்தில் நூற்றாண்டுகள் காட்டியுள்ளன. மொழிபெயர்ப்பில் சிற்றியான பொருள்களை பொருள்கள் மற்றும் பொருட்களை ஒன்றுதொன்றாகக் கொண்டு நூற்றாண்டுகள் காட்டியுள்ளன. முற்படி வேளில் முற்படி வேளில் நூற்றாண்டுகள் காட்டியுள்ளன.

நூற்றாண்டுகள்

(1.P.Number) பிரபமானம் அரசாங்கத்தில் நூற்றாண்டுகள் காட்டியுள்ளன. மொழிபெயர்ப்பில் சிற்றியான பொருள்களை பொருள்கள் மற்றும் பொருட்களை ஒன்றுதொன்றாகக் கொண்டு நூற்றாண்டுகள் காட்டியுள்ளன.

அல்லது

(1.P.Number) பிரபமானம் அரசாங்கத்தில் நூற்றாண்டுகள் காட்டியுள்ளன. மொழிபெயர்ப்பில் சிற்றியான பொருள்களை பொருள்கள் மற்றும் பொருட்களை ஒன்றுதொன்றாகக் கொண்டு நூற்றாண்டுகள் காட்டியுள்ளன.

சிற்றியான பொருள்களை

(1.P.Number) பிரபமானம் அரசாங்கத்தில் நூற்றாண்டுகள் காட்டியுள்ளன. மொழிபெயர்ப்பில் சிற்றியான பொருள்களை பொருள்கள் மற்றும் பொருட்களை ஒன்றுதொன்றாகக் கொண்டு நூற்றாண்டுகள் காட்டியுள்ளன.

சிற்றியான பொருள்களை

(1.P.Number) பிரபமானம் அரசாங்கத்தில் நூற்றாண்டுகள் காட்டியுள்ளன. மொழிபெயர்ப்பில் சிற்றியான பொருள்களை பொருள்கள் மற்றும் பொருட்களை ஒன்றுதொன்றாகக் கொண்டு நூற்றாண்டுகள் காட்டியுள்ளன.
பல உலகம், குழுக்கள் மற்றும் பேச்சுக்களும் நிறைந்தவர். இவர்கள் புரட்சியில் காட்சிகளே விளக்குவதற்கு வேறுபட்ட போர்விளக்கங்களைக் காட்சிப்படுத்துவதற்கு. 

சட்டப்பரங்கள்

நிற்கமலாயின் பாதை, அரசியல் துறவியல்முறையில் மாற்றங்கள் உரியவை என்று பொருளாதாரம் தொடர்ந்து கொண்டன. இவ்வருடன் ரீதியாக துறவியல்முறையில் மாற்றங்கள் பிரதானமாக தொடர்ந்து கொண்டுள்ளன. அதன் நிறுவனம் முன்னேற்றத் தொடர்ந்து கொண்டுள்ளது. இல்லை அல்லது மற்றும் பாதை நிறுவகளின் பாதையில் முன்னேற்றம் என்று கூறியுள்ளனர். இவ்வருடன் துறவியல்முறையில் மாற்றங்கள் பிரதானமாக தொடர்ந்து கொண்டுள்ளது. இல்லை அல்லது பாதை முன்னேற்றம் என்று கூறியுள்ளனர். இவ்வருடன் துறவியல்முறையில் மாற்றங்கள் பிரதானமாக தொடர்ந்து கொண்டுள்ளது. 

புது புதுக்கோட்டை என்று சட்டப்பரங்கள் உரியவை என்று பொருளாதாரம் தொடர்ந்து மாற்றங்கள் காணப்பட்டுள்ளன. இது புரட்சியில் மற்றும் போர்விளக்கங்களை நிறைந்தவர். இவ்வருடன் ரீதியாக மாற்றங்கள் பிரதானமாக கொண்டுள்ளது. அதன் நிறுவனம் முன்னேற்றத் தொடர்ந்து கொண்டுள்ளது. இல்லை அல்லது பாதை முன்னேற்றம் என்று கூறியுள்ளனர். இவ்வருடன் துறவியல்முறையில் மாற்றங்கள் பிரதானமாக தொடர்ந்து கொண்டுள்ளது. இல்லை அல்லது பாதை முன்னேற்றம் என்று கூறியுள்ளனர்.
முதல் கனிம்புல்லி நீர்வாலன்

இய. குறிப்பிட்டு
MA (Tamil), M.Sc (I.T.), M.C.A., M.Phil.,(Tamil)

அடுத்து தேன் அலைவாசர்,

காந்தை தி. அன்னை எனவும்,

பரார்பார் பாலானதுகளும்,

e-mail : sundarbaskar@gmail.com

பதிப்பு

துறையும் பரந்தவாழ்வாலும் நீர்வாலன் அடுத்துப் பெயர்க்கப்படும்படி பாலான வெளியுள்ளது நீர்வாலன் அளவுகளில் இருந்து முடிவு பெறப்பட்டுள்ளது. இன்றும், இன்றைய நீர்வாலன் அளவு பார்க்கப்படும். நீர்வாலன் பார்க்கும் விளக்கவும் கருத்தானார்:

அளவியல்வாலன் பெரும்பான்மையானதும் கல்வி மற்றும் கிராமிகம் மற்றும் அழைக்கும் நீர்வாலன் தொடர்புடையவற்றில் விளக்கம் பெறப்பட்டுள்ளன. அவை வெளியே நீர்வாலனுக்கும் நீர்வாலநிலைக்கும் விளக்கம் வெளியே வந்தது. தொடர்ந்துப் புகழ் மற்றும் விளக்கம் நீர்வாலன் பாலானது தொடர்ந்துப் புகழ் மற்றும் விளக்கம் நீர்வாலன் பாலானது.

இன்றும் ஒன்று லிங்கம் அடுத்து பெயர்க்கப்படும் பாலான நீர்வாலன் விளக்கம் பெறுகிறது. இந்த விளக்கம் நீர்வாலன் பாலானது பாலானது பாலான விளக்கம் வெளியே வந்தது. இந்த விளக்கம் நீர்வாலன் பாலானது பாலான விளக்கம் வெளியே வந்தது.

முறையிட்டு இயக்கு

இன்று தொடர்ந்து பரந்தவாழ்வாலும் நீர்வாலன் அடுத்து பெயர்க்கப்படும்படி பாலான வெளியே வந்தது. அவை வெளியே நீர்வாலனுக்கும் நீர்வாலநிலைக்கும் விளக்கம் வெளியே வந்தது. இந்த விளக்கம் நீர்வாலன் பாலானது பாலான விளக்கம் வெளியே வந்தது.
Index - Concordance - Subject Index

The present study was carried out in the first stage of the project, in the year 2001. The project was sponsored by the University Grants Commission, New Delhi, under the North Eastern Hills University, Shillong, and was completed in the year 2005. The project was an attempt to compile a word index of old Tamil Cankam literature, as a part of the larger project on Tamil Cankam literature. The study was carried out in collaboration with the University of Madras, and the project was supervised by Dr. K. Madhavikutty, the Dean of the Tamil University, Madras. The project was financed by the University Grants Commission, New Delhi, and the Council of Scientific and Industrial Research, New Delhi.

1. 1985, Computer Analysis of Tirukural, S. Baskaran, Cellamuthu, Tamil University.

2. 1993, A word index of old Tamil Cankam literature, Thomas Lehman and Thomas Malten, Institute of Asian Studies, Chennai.
நோக்கு பலக்கடவுளின் 1986-ம் ஆண்டு ராஜேஷ் ஸ்ரீலலேஷ் அயர்தூர் முதலாளர் பலகத்தை அணுக்கியது முதலாளரின் விளக்கத்தை வைக்கப்பட்டது. முதலாளர் அயர்தூரின் நான்கு வருடமானது விளக்கத்தை வைக்கப்பட்டது. விளக்கங்களின் சில அரசுத் தொடர்புகள் கொண்டிருள்ளது. பொருளாதார நோக்கு பலகத்தை அணுக்கியது முதலாளரின் விளக்கத்தை வைக்கப்பட்டது. மேலும் மூன்று முதலாளரின் விளக்கத்தை வைக்கப்பட்டது.

கூறி என்னும் பெயரில் அயர்தூரின் விளக்கத்தை வைக்கப்பட்டது. மேலும் மூன்று முதலாளரின் விளக்கத்தை வைக்கப்பட்டது.

1986-ம் ஆண்டு ராஜேஷ் ஸ்ரீலலேஷ் அயர்தூர் முதலாளர் பலகத்தை அணுக்கியது.
Simple Concordance

KWIC Concordance, Lemma Extractor

Search)}

Tamil Analysis Software

Word Space, Word Form)
Auxiliary Verb

Vb.net

Post-Editing)

Ambiguity)

Pre-Editing)

Occurance)

Vb.net (Rich Text Box)
Control எனும் வருடம் விளைந்த எல்லை நிறுவனத்தின் ஆரம்பகாலக் காலங்களும் அமைந்துள்ளன. இதெல்லை வருடங்கள் போன்றவை தொடர்ந்து ஆரம்பிக்கப்பட்ட அதிச்சந்தப் பொறியல் மற்றும் கல்வியளவு நிறுவனப் போன்றவை முதலில் சேர்ந்து விளங்கும்.}

- நிறுவனங்கள் வருடங்கள் ஆரம்பிக்கப்பட்டு விளங்கும் காலங்களும் அமைந்துள்ளன. இதெல்லை வருடங்கள் போன்றவை தொடர்ந்து ஆரம்பிக்கப்பட்ட கற்பல்லுக்கு மற்றும் கல்வியளவு நிறுவனப் போன்றவை முதலில் சேர்ந்து விளங்கும்.

- இழைத்துள்ள வருடங்கள் நிறுவனங்கள் ஆரம்பிக்கப்பட்டு விளங்கும் காலங்களும் அமைந்துள்ளன. இதெல்லை வருடங்கள் போன்றவை தொடர்ந்து ஆரம்பிக்கப்பட்ட கற்பல்லுக்கு மற்றும் கல்வியளவு நிறுவனப் போன்றவை முதலில் சேர்ந்து விளங்கும் காலங்கள்.
வரிசையில் இணையப்படுத்தும் விளக்கப்படுத்தும் சரியான தொடர்வரிசையில் Database(விளக்கப்படுத்தும்) மற்றும் Index(விளக்கப்படுத்தும்) வரிசையில் குறியீட்டுகளை மற்றும் ஆய்வடைந்த வரிசையில் தொடர்வரிசையில் குறியீட்டுகளை விளக்கப்படுத்தும் வரிசையில் தொடர்வரிசையில் குறியீட்டுகளை விளக்கப்படுத்தும்.
7

ஏழைநிலையில் பங்குதல்

தமிழ் விளக்கம்

உயர்ந்த விளக்கங்கள்
Fostering Tamil Web Communities for Mining Tamil Web Pages

Dr. A. Vijaya Kathiravan, R. Vidhya
Dept. of Master of Computer Applications, KSR College of Technology, Tiruchengode - 637 215
Contact: vijayakathiravan@hotmail.com, vijayakathiravan@gmail.com, Call: +91-9244217777

Abstract

“There are three popular qualities identified with Tamil people in 20th century: Creativity, intelligence and artistic thinking”. With the growing Tamil interest and Internet, the amount of Tamil data doubles every 12-14 months and will increase even more dramatically in the coming year. With an enormous amount of Tamil data stored in web databases and warehouses, it is increasingly important to develop powerful tools for analysis of such Tamil data and mining interesting patterns from it. There is a strong interest in employing methods of data mining to generate models of Tamil related web pages forming web communities. The main intention of this paper is to establish cyber community mining technique for Tamil domain and to identify the Tamil resources available in the free web. This paper proposes a new initiative for forming Tamil web communities with concise introduction about community mining. It also groups research publications and literatures in Tamil using bibliometric analysis. This community mining will yield benefits to all Tamil lovers, who want to be well-versed in a Tamil domain of his own interest. By forming people communities (i.e., people belonging to similar interest) using social network analysis, the domain knowledge in Tamil can be shared. Hence, web community mining may play an important role in forming Tamil Web Communities for gathering Tamil resources and documents of similar interest.

Keywords: Web communities, Tamil communities, social network analysis, community mining, web mining, web structure mining, citation, co-citation and bibliometric coupling.

1. Introduction

The ability to train computers to predict properties based upon knowledge of Tamil computing for Tamil data offers the prospect of automatically screening massive libraries of other language information to produce prediction. Tamil computing is an application of information technology in Tamil to the storage, management and analysis of Tamil information. Tamil information consists of Tamil databases, Tamil articles, Tamil publications, Tamil literatures, Tamil news, agricultural, health care, scientific information and other Tamil related information. With the increase in Tamil interest in web, computer-based Tamil databases, operating systems, hardware, software, tools, search engines and mining techniques have become very essential. Nowadays, web mining outperforms a significant position in information retrieval process using search engines, in which the main ingredients are web content mining, web structure mining and web usage mining. About 40% of the people are likely to surf the web using hyperlinks. Community mining is transpired from web structure mining that aims to
utilize the hyperlinked structural pattern available in web provided that it has some meaningful information hidden inside. A web community refers to group of web pages sharing a common interest like Tamil implicitly or explicitly. Using people communities, people in similar profession can group together as virtual teams to accomplish their tasks.

2. Tamil Community Mining

A Tamil web community is a set of Tamil web pages that provide resources on a specific topic in Tamil. By modeling the Web as a graph and performing several operations on it, it was able to separate the Web in sets of related items in which Tamil community is extracted. The different types of association relationship existing in a community are citation, co-citation and bibliographic coupling. Those have been described in fig 1.

Using the above associations, bibliometric coupling finds patterns in citation graphs; sociometric finds patterns in social networks; collaborative filtering finds patterns in rank graphs; webometric finds patterns in web page links. The main scope of community mining is to measure the similarity of web pages on the web graph and to extract the meaningful communities through the link structure pattern.

2.1 DEFINITIONS OF COMMUNITY

Several different definitions of community were also raised in the literature that has been explained in

Fig 1. Types of associations in Community

Fig 2. Several Different Definitions of Community
A web community is a number of representative authority web pages linked by important hub pages that share a common topic.

A web community is a highly linked bipartite sub-graph and has at least one core containing complete bipartite sub graph.

A set of web pages that linked more pages in the community than those outside of the community could be defined as a web community.

A research community could be based on a single most cited paper and contain all papers that cite it. While each of the above definition characterizes some essential properties of a community, it makes the community mining task rather difficult because of a lack of uniform definition.

### 3. Discovering Communities

In this paper, a community on the web is defined as a cluster of web pages, which share common topics. However, there are many ways to detect the clusters of web communities. One of the key distinguishing features of the algorithms has to do with the degree of locality used for assessing whether or not a page should be considered a community member. On the one extreme are purely local methods, which consider only the properties of the local neighborhood around two vertices to decide if the two are in the same community. Global methods operate at the other extreme, and essentially demand that every edge in a Web graph be considered in order to decide if two vertices are members of the same community. Broder et al. [1] reported on an algorithm of clustering web pages based on the contents. This approach can be applied not only to hypertext but also plaintext. However, indexing web pages accurately is difficult because the contents of web pages are not always meaningful. In contrast to the content-based approach, links in web pages can be reliable information because they reflect human judgment. Botafogo and Shneiderman [2] proposed an idea for abstraction called aggregate based on graph theory. Their algorithm removes ‘indics’ (nodes with high number of out-links) and ‘references’ (nodes with high number of in-links) iteratively to clear the graph. However, removed nodes often become very important elements to understand the web. On the other hand, Kumar et al., [3] defined a community on the web as a dense directed bipartite subgraph, and discovered over 100,000 communities. However, the scale of subgraphs depends on its parameters. This implies the difficulty in detecting communities from the web, since the communities are often somewhat related with each other. As another use of links, Kleinberg [4] and Brin and Page [5] used the link structures for ranking web pages. Their main idea was based on mutually reinforcing that the more a web page is referred, the more authoritative the web page becomes, and the higher the web page ranks. The highly ranked web pages tend to be the representative web pages of communities. There are several data structures such as biconnected components, strongly connected components, bipartite graph (BG), dense bipartite graph (DBG) and Complete bipartite graph or bipartite core (CBG) employed for forming communities.

### 4. Terminologies In Web Community

**Bipartite graph (BG).** A bipartite graph BG (T,I) is a graph, whose node-set can be partitioned into two non-empty sets T and I. Every directed edge of BG joins a node in T to a node in I.

**Dense bipartite graph (DBG).** Let p and q be nonzero integer variables and tc and ic be the number of nodes in T and I, respectively. A DBG (T, I, p, q) is a BG (T,I), where (i) each node of T establishes an edge...
with at least \(p\) (\(1 \leq p \leq i_c\)) nodes of \(T\), and (ii) at least \(q\) (\(1 \leq q \leq t_c\)) nodes of \(T\) establish an edge with each node of \(I\).

**Complete bipartite graph (CBG).** A CBG \((T, I, p, q)\) is a DBG \((T, I, p, q)\), where \(p = i_c\) and \(q = t_c\).

---

**Community hierarchy**

Let the variable \(\text{num\_levels}\) denote the number of levels in a hierarchy for a given data set. A community is denoted with \(C(i, j)\), where \(i (1 \leq i \leq \text{num\_levels})\) is a nonzero integer value that denotes the level and \(j\) is an integer value which denotes unique community identifier at level \(i\). Then, If \(i = 1\), members of \(C(i, j)\) are the web pages. If \(i > 1\), members of \(C(i, j)\) are the communities at level “\(i-1\)”.

**Community \((C(i, j))\)** Let \(p_i\) and \(q_i\) be integer variables that represent threshold values. The community \(C(i, j) = T\), if there exist a DBG \((T, I, p, q)\) over a set of nodes at level “\(i-1\)” with \(p \geq p_i\) and \(q \geq q_i\).

**Cocite** Let \(p_i\) and \(p_j\) be pages. Cocite\((p_i, p_j) = true\), if \(|\text{child}(p_i) \cap \text{child}(p_j)| \geq \text{cocite\_factor}\), where \(\text{cocite\_factor}\) represents a nonzero integer value.

**Relax cocite** Let \(T\) be the set of pages and \(P_j\) be the another page \((P_j \neq T)\). For any page \(P_i \neq T\), if relax\_cocite\((T_i, T_j) = true\) if \(|\text{child}(P_j) \cap \text{child}(T)| \geq \text{relax\_cocite\_factor}\). Here, \(\text{relax\_cocite\_factor}\) is nonzero integer variable and \(\text{child}(T)\) contains the children of the pages of \(T\).
Max flow min cut theorem. According to Floyd and Fulkerson’s Max flow min cut Theorem, an ideal community $C \subseteq V$ can be identified by calculating the s-t minimum cut using appropriately chosen source and sink nodes. In recurse over algorithm, community obtained in one iteration is used as input to next iteration. Finally, mirroring pages in the community are eliminated using “shingling” method. Quite a lot of work has been done in mining the implicit communities of users, web pages or scientific literature from the Web or document citation database using content or link analysis. Forming Communities among users is called as people community. Using bibliographic references in scientific literatures and document citations, scientific literature community can also be formed, which are described in detail.

![Fig. 5: Architecture of Search Engine for mining Tamil Web Communities](image)

5. Proposed Algorithm for Evaluating Tamil Communities

5.1 Main phases

Phase1. Collect web pages related to a Tamil domain.

Phase2. Discover communities on the web for all domains.

Phase3. Discover established relations among the communities.


Phase5. Visualize inter and intra relationship of communities.

5.2 The Detailed Process

Phase1. Preparations: First of all, let a user decide a Tamil domain, which she/he want to explore thoroughly.
Then, source web pages $D$ are collected by using any conventional search engine like Google. Here, web pages of Google's output for the query are downloaded.

**Phase2. Discover Communities:** Use community trawling algorithm or s-t max flow min cut algorithm for community mining. For surveying the picture of communities with future enhancements among communities, only centered web pages in communities are used instead of all the web pages. The centered web page named as core-page is extracted as follows.

1. Count the frequency of links included in $D$.
2. Regard the top $N_1$ links $C$ as the 'core-pages' of communities.

**Phase3. Discover Established Relations:** Measure the relations among core pages by counting the number of co-citations, and regard strong relations as established links. The process is as follows.

1. For every pair of two core-pages in $C$, count the number of links included in both the core-pages.
2. Regard the top $N_2$ pairs as established links $L_1$ (solid lines in Fig. 6).

**Phase4. Discover Future Enhancements:** Measure the relations among core-pages by counting the number of co-citations, and regard weak relations as future links. The process is as follows.

1. For every pair of two cores in $C$ except for $L_1$, count the number of links included in both the cores.
2. Regard the top $N_3$ pairs as future links $L_2$ (dotted lines in Fig. 6).

The movement of communities is shown by established and future links. Therefore, future enhancements are expressed by the combination of these two kinds of relations.

**Phase5. Visualization:** Core-pages and its relations ($C$, $L_1$, and $L_2$) are visualized into 2-dimensional interface to piece out the connections of communities and to understand the potential needs or demands.
Conclusion and Future Work

In this paper, the importance of discovering communities in Tamil for gathering resources has been strongly insisted. Using visualization, an idea for discovering established web pages in a Tamil community is mined by chaining primitive communities to understand potential needs or demands. It could find communities of adaptive granularity. By using techniques such as community clustering, social network analysis and bibliometric analysis, the Tamil world can get more benefits in future. For efficient community mining, algorithms based on machine learning may be used in future. Since this is a general idea that could potentially be used in any application scenarios, where the data can be abstracted to a graph structure, this may be applied to any domain. Its usefulness in other environments such as human relationship network, newsgroups, communication network, etc can be tested and the evolution of communities can be analyzed in future. Since the objects and links in a network are usually dynamic, changes of communities in a time series manner can be observed. From this idea of implementing community mining in Tamil, it has been confirmed that communities are the useful means to share resources and knowledge among people from different community domains.

People community and scientific literature community may act as a powerful knowledge acquisition tool for the Tamil community. The ideas and methods presented in this study can be proved useful by analyzing more domain communities. Future research in community mining focuses on generalizing the notion of community parameterized with coupling factor low for weakly connected communities, high for highly connected communities and optimal for ideal community. Co-boosting is a method for information retrieval from an unlabelled data. By combining this technique with community mining in Tamil, variety of Tamil based community like social community, literature community, political community, student community, Tamil cinema and songs community, research community and other Tamil related communities may be formed in future.

Acknowledgement

“Thou art the real goal of human life, we are yet but slaves of wishes putting bar to our advancement. Thou art the only God and Power to bring us upto that stage”. I sincerely thank INFITT-2010 Conference Selection Committee for having given us this opportunity. My extended gratitude goes to our college superiors, colleagues, friends and family members.

REFERENCES


Dr. A. Vijaya Kathiravan is an Asst. Professor in Department of Computer Applications, K.S.R. College of Technology, Tiruchengode, TN, INDIA. She received her M.Phil. in Computer Science from Bharathiar University, Coimbatore and recently she awarded her doctoral degree in University of Madras, Chennai. Her research interests include information retrieval, web communities, machine learning, data mining, data structures, text mining, NLP, social network mining, leadership assessment and human resource management.
பண்டைய திட்டம் அதன் தொகுப்பு குறிப்பிட்டது

இல்லைக்கால்கள் பாதமின் தருகை

(11- திட்டம் குறிப்பிட்டது)

லர்சு புராண

இல்லையுடன் சேரும் பண்டைக்காலம்பால்

லர்சு பயிர்கள், லர்சு பயிரில்

i-DNS.net International Inc.

பரப்பு அலுவல், சேரும் பரப்புடன், அதிகரிக்கும் பரப்பு

மையக்குறிய: maniam@i-dns.net

முனை

இல்லைக்காலம் திட்டங்கள் குறிப்பிட்டது அதன் தொகுப்பின் இல்லைக்காலம் பற்றாகம் குறிப்பிட்டது (IP Address) அதில் இந்து பயிரில் திட்டங்கள் வகைப்படுத்துகின்றன. பயிரியான அல்லது பயிரியின் வகைப்படுத்துகின்றன. இது இல்லைக்காலம் பற்றாகம் இல்லைக்காலத்தில் பற்றாகம் அல்லது அமைப்பு என்று. அனைத்து, இல்லைக்காலம் வகை ASCII தொகுப்பின் வகைப்படுத்துகின்றன. வகைப்படுத்தல் வகைப்படுத்தல் வகைப்படுத்தல் வகைப்படுத்தல் வகைப்படுத்தல் வகைப்படுத்தல்

அனைத்து 1990களின் பிற்பகலிகளில், இல்லைக்காலம் பற்றாகம் எடுக்கப்பட்டது. ஏற்றம் மறியும் பண்டை பயிரியான பற்றாகம் இல்லைக்காலம் பற்றாகம் இல்லைக்காலம் பற்றாகம். ஏற்றம் மறியும் பண்டை பயிரியான பற்றாகம் இல்லைக்காலம் பற்றாகம் இல்லைக்காலம் பற்றாகம்

1998 இல்லைக்காலம் திட்டங்கள் பற்றாகம் ASCII தொகுப்பு ஏற்றப்பட்டது. மூலம் ASCII தொகுப்பின் வகைப்படுத்தல் இல்லைக்காலம் பற்றாகம் ASCII தொகுப்பு ஏற்றப்பட்டது. மூலம் ASCII தொகுப்பு ஏற்றப்பட்டது. மூலம் ASCII தொகுப்பு ஏற்றப்பட்டது. மூலம் ASCII தொகுப்பு ஏற்றப்பட்டது.

1998 இல்லையுடன் Internet Consortium for Assigned Names and Numbers (ICANN) ஏற்றப்பட்டது. இல்லையுடன் Internet Consortium for Assigned Names and Numbers (ICANN) ஏற்றப்பட்டது. இல்லையுடன் Internet Consortium for Assigned Names and Numbers (ICANN) ஏற்றப்பட்டது. இல்லையுடன் Internet Consortium for Assigned Names and Numbers (ICANN) ஏற்றப்பட்டது. இல்லையுடன் Internet Consortium for Assigned Names and Numbers (ICANN) ஏற்றப்பட்டது.
1. DNS and Domain Name System

The Domain Name System (DNS) is a fundamental infrastructure of the Internet, responsible for translating human-readable domain names into numerical IP addresses. DNS operations are managed by a global network of DNS servers that can be categorized into authority, caching, and delegation servers. The DNS protocol is based on the concept of hierarchical domains, where each domain name represents a unique entity on the Internet.

DNS operates using a client-server model, where DNS clients (e.g., web browsers, email clients) query DNS servers to resolve domain names into IP addresses. DNS servers use recursive or iterative query methods to resolve domain names.

DNS employs a hierarchical naming system, with top-level domains (TLDs) such as .com, .org, .net, and .gov, and country code top-level domains (ccTLDs) like .us and .ca. Second-level domains (SLDs) and subdomains are specified after the TLD and are determined by the owner of the domain.

DNS employs a distributed caching system, where cached DNS records are stored in DNS servers to reduce latency and improve performance. Cache poisoning attacks can result in the placement of incorrect DNS records into the cache, leading to potential security issues.

DNS is managed by the Internet Corporation for Assigned Names and Numbers (ICANN), which oversees the allocation and management of domain names and IP addresses. The DNS protocol is specified in RFC 1034 and RFC 1035.

2. Evolution of the Domain Name System

The Domain Name System (DNS) has evolved significantly since its inception. The DNS was first designed by Paul Mockapetris in 1983, and its initial implementation, version 4 (DNSv4), was deployed in 1985. DNSv4 employed a simple, fixed-size record format, but it became insufficient to handle the rapid growth of the Internet.

DNSv6, version 6, was introduced in the late 1990s to address the limitations of DNSv4. DNSv6 uses a more flexible record format and supports variable-length domains, which allows for more efficient storage and querying of DNS records.

DNSv6 also introduced support for DNSSEC, a set of mechanisms for providing security features for the DNS infrastructure. DNSSEC employs digital signatures to ensure the integrity and authenticity of DNS records, thereby preventing DNS spoofing and other security-related issues.

DNS has since evolved to support additional features such as DNS over HTTPS (DoH) and DNS over TLS (DoTLS), which provide enhanced security for DNS queries. The evolution of DNS continues as the Internet grows and evolves, with ongoing efforts to improve its scalability, security, and performance.

In the future, DNS is likely to evolve to support even more advanced features, such as DNS over QUIC (DoQUIC), which allows for faster and more secure DNS queries.
3. குறிப்பிட்டு விளக்கம் கட்டும் விழா?

IDN என்பது ஆங்கிலத்தில் அறியப்படும் ஒரு பெயராகும், இதன் குறிப்பிட்டு அரங்கங்கள் உரையானது சென்று காட்டுவது, தொழில் பள்ளிகளுக்கு தேசியர்கள் வளர்கிறது. இது விளக்கம் மற்றும் இணையான தொலைக்காட்சியாகும். இது விளக்கம், அதன் பின் முன் வடிவங்கள் போன்றவை. 2000களின் செண்டாம் கூட்டம், இக்கட்டும் விளக்கத்தின் ஒரு பாதையாகும். இது விளக்கம் பின்னால் உள்ள புதிய தொலைக்காட்சிகள் இதன் தொடக்கங்கள் தொடங்கினால்.


4. செல்வல் பின்னால் பழங்குடிய முயற்சிகளின் விளக்கம் கட்டும் விழா?

IDN கால்வரிசை தொடங்கும் விளக்கம் பிறந்தது. இதன் பயன்பாட்டை பிறந்தது. இதன் பயன்பாட்டை பிறந்தது. இதன் பயன்பாட்டை பிறந்தது.

1. முன்னாள் இயங்கு அளிக்கவும்
2. தொழிக்கொள்ளவும்
3. செல்வல் பிறந்த விளக்கம்
4. பாதையாகவிட்டு
1. இடையல். ஆகஸ் infit.org
2. முனையர். ஆர்க். மாற்று tn.gov.in
3. ஆர்க். மாற்று gov.in

ICANN மற்றும் IDN

பதிப்பு 16 IDN ccTLD உரைக்காணாம். இவை உரைக்கு அல்லது உரைக்கு தொடர்ந்து கொள்கிறது. இவைகள் உரைக்கின்றன. பதிப்பு 16 IDN ccTLD உரைக்காணாம்.

1. 16 IDN ccTLD உரைக்காணாம். இவை உரைக்கு அல்லது உரைக்கு தொடர்ந்து கொள்கிறது. இவைகள் உரைக்கின்றன. பதிப்பு 16 IDN ccTLD உரைக்காணாம்.

2. ICANN உரைக்காணாம். இவை உரைக்கு அல்லது உரைக்கு தொடர்ந்து கொள்கிறது. ICANN உரைக்காணாம்.

3. ASCII உரைக்காணாம். ASCII உரைக்காணாம்.

1. தொடர்பு வருகையும் செயல்பாடுகள்.
2. பதிப்பு 16 IDN ccTLD உரைக்காணாம்.
3. வருகையும் உரைக்காணாம்.
A Tamil Web Portal Development with online Dictionary

A.Parimaladevi, III MCA,
Kumaraguru college of Technology, Coimbatore,
priyr.th.cu@gmail.com

A.Muthukumar, Professor,
Kumaraguru college of Technology, Coimbatore
amuthukumar@computer.org

1.0 Introduction
1.1 Purpose

The purpose of this document is to present a detailed description of the mytamilthai.com which is a Tamil web portal development that includes all the details in Tamil. It will explain the purpose and features of the website. This document is intended for both developers and end-users of the system.

1.2 Scope of the project

This site is going to develop mainly for the people who feel uncomfortable while browsing in English. This site includes the information about youth, jokes, sports, medicines and also information for women in Tamil. Instead of searching different sites user can search a single site for his/her entire requirement.

This site also includes the English to Tamil dictionary. User who needs to search the Tamil meaning for a particular English word can effectively use this site. Also if the particular word is not found and if the user knows the meaning, then he/she can add the new word. Administrator who has the rights to see the words in temporary database and add to the existing database of the dictionary.

Some pages of the site can be viewed by all people but certain pages are viewed by the user only after they register their name in this site. Also authorized user can advertise for the jobs they provide with admin assist. Also user can post their comments.

2.0 About index page::

An index page is the first page of web portal system. The site mainly consists of four main divisions. The division at the top is Header with the title and division at the bottom with privacy policy and conditions are displayed in all pages of Tamil web portal system. The next division is for menus which will be displayed only in certain pages.

The menus includes the following
- யுத்தம் - youth
- பெண்கள் - women
- கூறுகள் - jokes
- விளையாட்டுகள் - sports
- சமூகச்சுருள் - medicines

The last division will be content division. In this, there will be some sub-divisions. Left side which includes tabs like
• ஆரம்பத்தில் செய்யுவது - recent news
• வருடமாண்டு புரோக்கம் - about author
• உதவியைத் தருவது - user comments

In recent news, daily news will be updated for eg. About stock market, gold rate, politics, sports, etc. All the details about author will be displayed in next tab and at last if any user posts comments about any particular topic then with admin assist, the comment will be displayed.

The middle sub-division will be the exact content of particular topic. And the last sub-division includes the tabs like
• பதிவு செய்யும் பள்ளியை செய்யும் - login
• கையொண்டு எழுதலும் - jobs available
• குறிப்பிட்டுச் செய்யும் - post comments
• குறிப்பிட்டுச் செய்யும் - keyboard

To login, user needs to register in the site. The information about the registration will be stored in the database. All the jobs advertised will be displayed here. Every one can see all the information about that particular job by just clicking it. User who needs to post the comments, can post here but it will be stored in the database for admin to approve it to display in the site.

3.0 Admin login

Administrator who has all the rights on the system. As admin login, all the tabs displayed in index will be displayed with some extra tabs which includes
• உலகில் -dictionary
• புதுமையைத் தரும் - add new word in dictionary
• புதுமை -new job advertising
• நடும் வாசனை-advertised jobs
• கூட்டமைப்பு-keyboard

The menu will be enabled once admin login. As the menu contains the topics, the information regarding the particular will be displayed in the content page once we select the particular topic in that menu.

Online dictionary is available here. Tamil to English and English to Tamil dictionary is available with the description in both the languages. Admin add words to dictionary and it will be stored in database in Unicode format.

Admin who enters all the information about the job by the information provided by the job provider. Also already advertised jobs will be displayed.

4.0 User login

User after registering, user enters the site and views all the tabs displayed in index page with some extra tabs. User also can add words to dictionary but it will be stored in database. If admin find that word is irrelevant then he has the rights to remove the word from database so that, that word won’t be displayed further unless again adding it properly.

Users who wish to post comments can post here. This is also stored in database for admin use.

5.0 Online dictionary

As described earlier, online dictionary will be displayed both in admin and users login. By typing English word we will get Tamil meaning as well as description in Tamil and the same way if we type in Tamil we will get English meaning with description in English.

6.0 Conclusion

Mytamilthai.com is developed to help all the people who need to know all the information at a single site. Here there is an availability of dictionary with English to Tamil as well as Tamil to English dictionary with glossary. So it is very useful for all people at different countries to the know the meaning for any word they wish. Recent updates are displayed and it will be updated daily.
**Future enhancement:** This project has been developed as a Master’s project and is constrained by time. There is scope for extending the system as per the need.

Job providers advertise for the job availability now. Job seekers at present can view just the job advertisement and need to contact separately by themselves with the information provided by the job provider. Further it is going to extend by job seekers to post their resume for the advertisement displayed in the site through this site directly.

Also now there is an availability of English to Tamil and Tamil to English dictionary only. Further it is going to enhance in different language according to users need.

Also according to users authorized comments the site will be enhanced.

**7.0 References**

1. PHP 6/MySQL® Programming for the Absolute Beginner – Andy Harris –
2. http://www.w3schools.com/PHP
5. http://senthilvayal.wordpress.com
Methods and Options for Videoconferencing in Relation to the Tamil Language in 2010

Eric Miller

On 21 February 2009, Chief Minister Kalaignar Karunanidhi in Chennai made a video call to launch the 3G (Third Generation) network services of BSNL, the Government-run telecommunications company. Today (May 2010), video calls on mobile telephones using BSNL’s 3G network are being routinely made by members of the public in Tamil Nadu. Airtel’s 3G network is also expected to be operational soon.

At this moment in time, as we are on the verge of the Video Call Revolution, it may be useful to look both backward and forward. This might help us to decide what to do with our new videoconference capabilities, as we enter the Age of Videoconferencing in earnest. Previously, I have written three articles about videoconferencing for INFITT -- in 2002, 2003, and 2004. This article summarizes and adds to those.

Videoconferencing, video calls, video chat, or simply, being able to see people as we speak with them through electronic devices, has been on the horizon for many years. Two major ways of videoconferencing are: through one’s computer, and through one’s mobile telephone. In both cases, the hardware is increasingly coming with a video camera built-in. The videoconference camera is generally above the screen. In the case of mobile telephones, there are usually two cameras, one facing the user (for videoconferencing), and one facing away from the user (for optional use for recording still-images and video). Actually, the computer and the mobile telephone are converging, to produce smart phones -- and these are the mobile telephone models that tend to be videoconference-capable.

Skype is most commonly used on personal computers, but it can also be used through smart phones. Skype has brought videoconferencing to wider-than-ever-before public use. Other free video chat programs include those in Gmail and iGoogle, Microsoft’s Windows Live Messenger, Yahoo! Messenger, and Apple’s iChat.

Among online social networks (also known as, social media), Orkut is one of the most advanced in offering the video chat option to users. How will people with similar interests find each other to videoconference with? They could join communities, or become friends, fans, or followers of others, as many people already do on social media.

There are also programs such as Webex, which, for a cost, enable videoconferencing over the Internet, with sharing of files -- for text, electronic drawing, video, etc. -- in various windows. In the 2004 and...

2 http://www.webex.co.in
2005 Chennai-Philadelphia videoconferences I facilitated, we used a simpler (and lower-quality) method for showing Tamil text: we projected it onto the back wall of our room in Chennai (Figure 2). In other videoconferences, I have placed the Tamil text next to the image of the speaker, like in a comic book (Figure 1). The development of ways of producing instantaneous visual translation sub-titles, or captions, is going to be an important part of videoconferencing’s future. This will require voice recognition technology (from spoken to printed words), and automatic translation technology (from one written language to another).

Microsoft’s Natal system will feature motion-sensing technology -- not just for game-playing, but for operating the computer in general. Thus, the Gesture Revolution is upon us, in which a crucial input device is a video camera. Based on what I have seen in the college students I teach in Chennai, many young people today will not rest until they can play games via videoconference in social networks on mobile telephones.

Regarding high-quality types of videoconferencing in Tamil Nadu, India, and beyond:

Early videoconferencing in India (in the 1990s) tended to occur via dedicated non-Internet ISDN lines (three lines together yield 384kbps). However, ISDN lines cost by the minute, and are quite expensive to use -- and are even more so when a bridge is needed, for connecting more than two partners in a videoconference.

The global development is now toward videoconferencing via very high-speed Internet. For example, Reliance -- whose videoconference rooms in their over 200 Reliance World stores across India have led the way in making high-quality videoconferencing visible and available -- previously offered only dedicated ISDN-line videoconferencing; now they also offer Internet videoconferencing.

Access Grid is an “ensemble of resources” that enables many sites -- using interactive multimedia and appearing on multiple screens -- to participate in a videoconference (Figure 4). Polycom, Cisco, LifeSize, and Tandberg are among the videoconference companies that offer telexistence (presence, from a distance), which involves life-size high-definition images of people, simulated eye-contact, and minimum delay-time -- replicating the experience of physically-present meetings as much as possible.

In India, three Governmental entities that are involved with developing videoconferencing -- or the connectivity systems that enable videoconferencing -- are 1) ERNET (Education and Research Network); 2) NIC (National Informatics Centre); and 3) CDAC (Centre for Development of Advanced Computing).

NRENs (National Research and Education Networks) are playing increasingly important roles in Internet development in many developing countries. An NREN is a specialised internet service provider dedicated to supporting the needs of the research and education communities within a country. ERNET (Education and Research Network) is India’s NREN.

---

4 http://www.accessgrid.org
5 http://www.eis.ernet.in
6 http://home.nic.in
7 http://www.cdac.in
8 http://www.eis.ernet.in
Internet2 is a USA-based networking consortium. The technical standards and connectivity that Internet2 involves are very important factors in the global development of the Internet. Founded in 1996 by members of the education and research community, Internet2 provides both leading-edge network capabilities and unique partnership opportunities that together facilitate the development, deployment and use of revolutionary Internet technologies... Internet2 brings academia together with technology leaders from industry, government and the international community...and promotes collaboration and innovation... Internet2 has a section relating to Emerging NRENs. One of the Emerging NREN groups is the South Asia Special-Interest-Group (SA-SIG). SA-SIG’s mission is to help to facilitate high performance networking in South Asia. SA-SIG’s e-mail list -- <southasia-interest@internet2.edu> -- is very much worth joining.

Two annual videoconference showcase events based in the Internet2 community are the Megaconference, which began in 1999; and the Megaconference Jr, which is especially for schoolchildren, and which began in 2004. Megaconferences are marathon events (up to 12 hours) that are webcast live to a global audience. They are composed of numerous brief videoconferences, each between up-to-six parties. I participated in the 2005 Megaconference, and in the 2006 Megaconference Jr (in both instances along with Tamil children, and exploring children’s songs and language learning; from TENET’s IIT-Madras facility). Reliance personnel have participated in a number of Megaconferences.

TEIN3 is another network that enables high-speed videoconferencing and assists in facilitating interesting collaborations. TEIN3 is the third generation of the Trans-Eurasia Information Network. With direct connectivity to Europe’s GÉANT network, TEIN3 offers researchers and educators in Asia-Pacific a gateway for global collaboration with their peers in Europe and other parts of the world. India connected to TEIN3 in March 2010 (through ERNET). Sri Lanka connected to TEIN3 in April 2010 (through LEARN).

An entity to keep in mind in relation to high-speed international videoconferencing for education and development is the Global Development Learning Network (GDLN). The GDLN, which is coordinated by the World Bank, is “a partnership of over 120 institutions in over 80 countries that collaborate in the design of customized learning solutions for people working in development”. A regional association of GDLN, is GDLN Asia Pacific (GDLNAP). Presently there is one GDLN site in India, in New Delhi. It might be good to seek to have a GDLN site in Tamil Nadu also. A GDLN site typically features a large room equipped with top-notch videoconferencing facilities.

---

10 http://www.internet2.edu/about .
15 http://www.tein3.net .
16 Sri Lanka’s NREN is LEARN (Lanka Education and Research Network), http://www.ac.lk .
18 http://www.gdln.org/about .
Teaching Tamil language via videoconference -- on computers and mobile telephones -- could be a very important field. Tamil Nadu could be a world leader in developing language teaching in general by videoconference. These services should be available via Skype or similar programs, 24 hours a day.

At present, Tamil language-learning materials and instruction are available on webpages such as Web Assisted Learning and Teaching of Tamil. Such asynchronous learning processes could also have an optional synchronous videoconference component. The videoconference language-practice lesson-plans could be coordinated with the lessons on the webpages. The on-line tutors -- or language-practice partners -- would need to be recruited, trained, and put in contact with clients. This would involve a lot of work. It could be done as a business, an NGO, and/or an educational project, possibly subsidized by a government. In any case, it should be done, as a way of preserving, developing, and globalising the Tamil language. The early pervasiveness of the English language on the Internet is fading. Other languages are now also entering cyberspace, especially as the audio and video options are becoming more available and convenient.

My dissertation recommends three techniques for teaching language via videoconference: Question-and-Answer Routines, Repetition with Variation, and the Simultaneous Saying and Physical Enacting of Words. My research shows that these are prominent elements of Tamil children’s songs/chants/dances/games -- activities that very likely facilitate the acquisition of spoken language.

Question-and-Answer Routines place language-practice in the interactive context of human relationships, whether one participates in a routine as oneself or as role-playing other characters.

Repetition with Variation gives the learner a sense of control and competence. If only one aspect of a sentence is modified, the learner can still hold onto the grammatical structure of a sentence. Variations can include changes of tense, and substitutions of words (substitution drills); and going from a positive to a negative statement, or going from a command to a question (transformation drills). Repetition with variation is a key aspect of the modern language-teaching approach, the Audio-lingual Method. Systematic and methodical learners tend to especially enjoy this approach.

The Simultaneous Saying and Physical Enacting of Words utilises the entire body -- not just the brain and mouth -- in the language-learning process. The modern language-teaching method, Total Physical Response, is based on this idea.

These three practices are especially good for videoconference language teaching-and-learning, because in the course of videoconference communication it may at times be difficult to make out what a distant person is saying, and these practices makes interpersonal verbal comprehension more likely.

We also found that playing with puppets can add fun to videoconference language-practice. This additional level of mediation and role-play, which enables people to interact with each other indirectly, seems to relax people and take some pressure off them (Figure 3).

Tamil Nadu should take a leadership role in developing ways for many aspects of its culture to be shared via videoconferencing. Training, interactive performance, and discussion about the various aspects of Tamil culture could be made available online and via videoconference.

21 http://ccat.sas.upenn.edu/plc/tamilweb, by Dr. Vasu Renganathan.

aspects of culture could all occur via videoconference. This could be done in part in coordination with the State crafts organisation, Poompuhar; and the annual folk performing arts festival, Chennai Sangamam. Videoconferencing in the classroom is a huge field in the USA and Europe, and students and teachers there are very eager to videoconference with their counterparts in exotic places like India. Videoconference interviews with tradition-bearers are increasingly being held in Folklore and Social Studies classrooms.

Videoconference interviews for employment, and for admission to academic programs, are also becoming commonplace. Chennai and other cities in Tamil Nadu need teletoriums: halls equipped with large screens and videoconferencing facilities. Setting up videoconference systems in halls for single events is too time-consuming and risky. A single teletorium could be used by numerous academic institutions and other groups.

A word should be said about a world-famous experiment in bringing videoconferencing (and general computer and Internet use) to the Tamil countryside, and to other rural areas in India. In the early 2000s, Dr. Ashok Jhunjhunwala -- leader of TENET (Telecommunication and Computer Networking Group, Depts of Electrical Engineering, and Computer Science and Engineering, IIT Madras) -- helped to develop SARI (Sustainable Access through Rural India), which was serviced by n-Logue Communications Private Ltd and other companies. While it seems that this project has for the most part proved non-sustainable, Dr. Ashok Jhunjhunwala continues to champion the idea of videoconferencing for people in the Indian countryside. It may well be that such people may achieve the ability to videoconference through mobile telephones before they achieve it through desktop computers. In any case, there remains a place and need for NGO and Government support for rural and economically-disadvantaged people to utilize videoconferencing for educational, employment, cultural, and other applications.

Apollo Hospitals is one among a number of hospitals in Tamil Nadu -- private and public -- that have strong tele-medicine components, using videoconferencing in the diagnosis process, for staff training, and for other applications. Sad to say, but difficulties for the environment (such as the ash cloud that recently floated above Europe), and inconveniences for travel, tend to cause booms for videoconferencing. Videoconferencing can be thought of as a green activity, in that it can reduce the amount of petrol needed for travel, and it can make it unnecessary for people to travel through delicate natural environments. However, videoconferencing should not be seen as a substitute for physically-present communication -- it is simply a different type of communication, with its own strong and weak points.

---

Years ago, people could often be heard in Tamil Nadu’s browsing centres, singing Tamil cinema songs to distant others. Browsing centres have decreased, as Internet connectivity in the home, school, and office -- and on portable communication devices -- has increased. However, Tamil Nadu has been a world leader in bringing tribal, folk, and classical performing arts into the realm of cinema. Now these arts -- along with the teaching-and-learning of the Tamil language itself -- should be brought into the realm of videoconferencing, to further enable the sweet sound of Tamil to be heard and spoken around the globe.

A videoconference featuring electronic drawing and Tamil words (with transliteration and translation). (Figure 1.)

The words of a children’s song are shown on the Chennai side of the Oct. 2004 Chennai-Philadelphia videoconference. (Figure 2.)

A conversation through puppets. From the Oct. 2004 Chennai-Philadelphia videoconference. (Figure 3.)

An Access Grid videoconference. (Figure 4.)

Eric Miller <eric@storytellinginstitute.org> is Director of the World Storytelling Institute (based in Chennai); and is Assistant Professor of Story and Storytelling at ICAT, the Image College of Animation, Arts, and Technology (also based in Chennai).
Abstract

Most of the educational institutions now have extensive information and communications technology (ICT) in place. The cost of supporting, upgrading and replacing this equipment to provide a robust infrastructure for teaching and learning is increasingly onerous. This brings into question whether alternative network architectures, such as Thin Client computing could provide the required level of functionality with lower long-term costs and/or other benefit.

This paper addresses the use of thin client technology to provide optimum and cost effective solution for IT infrastructures which have the ability to integrate class room management application and school network management application with reliable and stable solution at minimum maintenance cost. The proposed thin client technology will be able to provide effective and secured centralized server based solution of schools with class room management systems integrated.

KEYWORD: Thin Client, school and class room management, server based computing

1 Introduction

This paper addresses the key area of institutional concern for the education sector, that of delivering effective and efficient school and class room management system in a flexible, secure, and accessible way to students in Malaysian Tamil schools. The system will adopt the thin client technology linked with centralized server to implement school and classroom management.

The proposed system will have secure integration with other key educational systems (e.g student records, module registration, and examination scheduling conducting trial exams and distribution of teaching materials), which will be delivered via network services and centralized server technology.

Thin client technology offers major advantages over conventional PC-based class room systems in terms of scalability, economy and sustainability. It can also offer additional flexibility in the range of school-teaching material and multimedia content that can be delivered without needing elaborate installation procedures or additional software to control the security, access and database.

2 Integration of the System

The system integration of Thin client based School and Class room management system with open source technology consists of:
i) **School management System** software designed to automate a school’s diverse operations from classes, exam to school events, calendar and to create powerful online community, by bringing parents, teachers, and students on the common interactive platform.

ii) **Local Centralized Server Based Technology** that consists of storage Server well equipped with database using Linux Edubuntu server platform.

iii) **LAN/WAN network system** which allows thin client terminals to access the centralized server remotely via PXE LAN booting technology. This allows freedom for user to access their desktop from any thin client terminals within the network coverage.

iv) **Class room management System** provides teachers with ability to instruct, monitor and interact with the students either individually, as predefined group or to the overall class.

v) **Thin Client Server** provides solutions for central deployment, configuration and management of thin clients and users connections.

vi) **Thin Client Terminals** that consists of motherboard, display devices, keyboard and mouse, without any preinstalled operating system and hard disk.

![Diagram of system integration](image)

*Figure 1 show the process flow of the system*

Figure 1: System Integration of class room and school management system
2.1 School Management System

The school Management System has something for everyone related directly or indirectly with the school and teaching environment. Some of the key advantages to schools and educational institutions are:

- Easy performance monitoring of individual teaching modules.
- Automated and quick report generation along with process turn around time.
- Centralized data repository for trouble-free data access.
- Authenticated profile dependent access to data.
- User friendly interface requiring minimal learning and IT skills.
- Design for simplified scalability.
- Elimination of people dependent processes.
- Minimal data redundancy.

Some of the advantages to parents are:

- Frequent interaction with teachers.
- Reliable update on child’s attendance, progress report and fee payment.
- Tracking of homework assigned by teachers to their child.
- Prior information about school events and holidays.
- Regular and prompt availability of school updates such as articles, discussions forums, image gallery and messaging system.

2.2 Class Room Management System

Advantages to teaching mechanism:

- Automated student attendance.
- Computerized management of marks and grades.
- Timetable creation in advance.
- Homework assignment to students and approval.
- Efficient and effective interaction with parents.
- Access to forum common to students and parents.
- Access to own and students attendance.
- Power on, power off, Reboot and Login to class room computers remotely.
- Broadcast messages to groups or all network users in seconds.

Some key benefits are:

- Enhanced interaction with teachers, parents and peers.
- On line submission of homework.
- Access to their attendance, timetable, marks, grades and examination schedule.
- Liberty to publish articles and views, and participate in discussion forums.
- Freedom to browse through library books catalogue and identify the book(s) to be issued.
- Prior information about school events and holidays.
2.3 Thin Client System

- Thin client is a general term for a device that relies on a server to operate.
- Thin client has display device, keyboard with mouse and basic processing power in order to interact with the server.
- An ideal thin client device contains no hard drives and CD or DVD-ROM

Plate 1: *Ideal thin client architecture*

3 Advantages of the Technology

By using thin client technology rather than standalone PCs, it is possible to deliver a wide range of computer based educational and examination materials while restricting other resources that are usually accessible to the students if conventional PC system is to be used. With conventional PC based technology, it is difficult to prevent access to the Internet, chat services, mobile devices such as USB drive, documents previously stored by other students etc., which could allow simple cutting and pasting of answers into the assessment or exam sheets by students with thin client technology in place. It is simple for an administrator to disable USB port on thin client terminals for the duration of the assessment or examination time, thus further limiting the ability for student's accessing disallowed information to assist them in the assessment or examination.

Another major attraction of the thin client technology for assessment purpose is that it is very resilient, given the fact that they have no software or moving parts. Therefore there are unlikely to be an issue when the assessment are not been delivered due to faulty desktop devices. This causes unnecessary pressure on the affected student and the additional works involved to the invigilator.

The issue of ensuring that PC's have the appropriate software available also affects PC's which are located in teaching spaces. Traditionally such PC's are left switched off when not in use which means that any automated software updates tend to fail or, worse, try to start when a teacher turns the PC on for a class. This can lead to anti-virus software not being updated, operating system vulnerability not being patched etc. the start up time of a PC system also causes difficulties, when a lecturer arrives in a class room, there will be about 8~10 minutes start up time for the PC and to get the necessary software up and running; if any updated needed to be done this could delay the start of the class. Using thin
client technology there is no need for the software updates and no need to worry about viruses. The user will always get the appropriate version of all the software via central server. The new upload of teaching material will be ready for teaching immediately as the student or teacher starts the class. Figure 2, shows the flow between teachers-students-parents-school in the school and class room management system.

![Diagram of School & Classroom Management System](image)

Figure 2: Data Flow between teachers-students-parent-school in the school & class room management system

4 Conclusion

The school and class room management system with thin client technology have not been fully integrated in the learning process in Malaysia. In last decade, the complexity of existing desktop machines, the capital investment needed for wide area network (WAN) access and lack of educational resources and multimedia content have prevented the potential of thin client based solution become reality.

Recently the convergence of community, business and government organizations in favor of client technology, have started to produce changes in education system. With the use of thin client technology, the teaching system will now can look forward into a new age of centrally manageable teaching technology, with equal access to information will be given to all students regardless of their background and geographical location. Rural students will be given full access to information and knowledge with tremendous reduction in communication time and infrastructure cost, knowledge can be shared with anybody from any part of the world for free. Coupling the centralized server with thin clients and the power of cloud computing, Tamil schools in Malaysia, soon will become the community information hub. The students who benefited with this technology will one day become knowledge based skill workers which will directly uplift the living standard of Malaysian Indian.
8

கல்லிலே பெண் கும்பு எழுதக்

ஸ் கல்லை விளம்பரிக்காம்
Embedding Co-Features in 'Ocr-Friendly' Fonts will go a Longway in Machine Reading of Texts

N D Loga Sundaram
26/15 Kutchery Lane Mylapore, Chennai, India.
Cell: 091 044 9283244798, Alt. 9283772110
selvindls61@gmail.com,

Preamble
The Regal superiority of digital world, crowned by its hyper speed-error-free signal transmission i.e. its unparallel dynamics in information handling. Because IT being made as retrofits to faculties handled by humans it gained popularity very quickly. Today they are ubiquitous products in the hands of informed community.

Every one of us knows well that innovations in IT field are born worldwide, every minute, both in hardware and in software. By each of its million facets, they get deep-rooted mm by mm, in minute-wise human activity around the globe, recognizing that 'Continuous Improvement ' is one of the core theme for remaining in forefront in any business.

Playing field
The overwhelming, inundating expansion of human activity in the digital virtual universe, i.e. in the interface between the human (with their archives), to the current digital machines are limited to, Vision and Sonics. Applications connected with vision and Sonics are already in use, over decades, one after another - Vision to digits and back, Sonics to digits and back, cross platform between Sonics and Vision thro digits.

Archives- CORE instrument in human advancement
As pointed out elsewhere the whole spectrum of human advancement is sourced by the creation of huge voluminous Banks of Knowledge in visual/written format, which bypasses time and distance in human civilization. It is the vessel that accumulates and vends knowledge so that it can be continuously reviewed updated and suitably edited (evolution) to their customized needs on date. Archived written language (vision) is prone less to corruption and distortion than a spoken (Sonics) one and hence they survived perfectly such a long way since Stone Age until this digital era.

Primary in-put medium to human knowledge
Modern civilized humans get their knowledge mostly through sight and sound, both in real world as well from archives, thro 'language' that is called Education'. The other faculties as taste, smell and feeling play lesser part quantitatively which still remains primitive and mostly limited to tangibles in front as non-virtual.
The apparent necessity of quicker machine reading

From times immemorial humans devised machines for many of their day-to-day activities by inherent laziness and to thrive with comfort in competitive world out of their disturbing ambience. This is true in digital world also, which plays an arterial roll on date in ‘human resource’. By nature I.T. is one of the platform erected for high altitude launch of mechanization and automation.

In the present world of communication explosion and consumerism we see people everywhere like to, rather forced to, have anything done quickly and easily, seldom even by fancy mania. The digital gadgets ranging from tiny hand held devices crowded with numerous features to mammoth net search engine servers, are the gadgets white collared people have interface these days, before middling with their bowl of soup. Next to TV and consumer durables, cunning high-tech assisted marketing techniques paved way for deep penetration of desktop PCs with all of its army of useful peripherals and the gadget of the day the cellular Mobile. Hence the universe of digital gadgets is in reign today and its influence extends well beyond informed community

Digital world and its Contents

Aptness and forerunner ship in handling of contents whether in a simple and humble home paperless-office or goliath popular commercial portal; it can be achieved by size of handled quantity as well as by timely presentation. Hence a self-employed free-lancer or a web master, he/she has to handle their contents, as ease and efficient as possible. In most cases, when the source is not from ready-in-the-shelf digital form, they get them digitized through their peripherals. In most of the cases, they are compiled from contents of several other authors, in print media, which warrants first degree of conversion.

Anyone who wares the shoe of a web master or a mere single document initiator they have initial digitization hardships and hiccups by virtue of time and strain it takes to key-in the contents thro’ human eye-fingers route. Therefore they always opt for an easier machine input process if available replacing the laborious ubiquitous keyboard route which causes fatigue, back pain and eye irritation from not so ergonomically designed work place.

Contents in text format

The contents available in texts format instead of an image format is preferred for memory size and more particularly when it has to be subjected to an analytical process as search compare and manipulate with mathematical or logical operators. Compiling is one of the core processes in any creation and hence pick a particular content or theme from the group is the necessity. Therefore text formats are first and foremost choice.

'Always stands in top'

Hence the case of machine reading from print media stands on top of wish list of authors. All the attempts focused to this theme, will be appreciated and hence they command huge market. By virtue of this commercial proposition, several OCR products were brought out already and serve well for about a decade.
Again a review of the need and worth of machine digitization

Replacement of human activity in archiving information thro' digital machine paves way for

1. Unbiased,
2. Quicker,
3. faultless and
4. closely managed

handling of information which leads to omnifarious economics. It shall even be the necessity in future high speed lanes in E-management.

MICR is the forerunner for OCR

Machine reading of printed characters are not new while it is about 30 years old, which is well before the advent of popular use of digital data with computers. It was introduced as magnetic character recognition in bank cheque nos known as MICR cheques. It was devised to avoid human failures and helps in speedy unbiased disposal by using a special kind of ink and a process to suit the machine reading along with safety. Initially it was limited to serial number of cheques (numerals)

Now, design of same numerals of those safety cheques got modified, with simple distinct features in its normal printed glyphs. On date, one can see combination of thinner and thicker strokes or micro sized squares placed at a particular zone with disparity in each numeral's grapheme for providing easily traceable differentiating features between each of the characters in the set.

Bar code Reading

Then the digital era of barcode characters with 59 white and black bars was introduces and it is still a popular true workhorse at every marketing/vender outlet stores. Bar-coded price/product tags embed numerous particulars, which have bearings, both for seller as well as purchaser. As the bar code is devised purely for machine reading process, it is not suitable for normal/average human cognizance.

2D Checkerboard type tags

Now another innovation one can witness is a kind of tag (or nameplate) replacing bar-coded tag. It is a printed square box puzzle checkerboard matrix, which has rows of black and white optically, coded fields. By its 2D nature it contains more data than a barcode and evidently needs different reading writing devices to suit the system in design.

RFID

High volumes of digital activity also bring newer problems and we know about the necessity of safety of contents handled is an indispensable factor in any commercial/legal propositions.

Mechanization in identification or authorization using a radio frequency signals and embedded 3D micro devised chips on to products are in use today for remote sensing at a distance which already started gaining popularity.
Currencies

We know that the safety features are inlaid in Currencies of most of the nations of the world have features for non-human cognizance serves automated or machine recognition.

Embedded Features in Ocr Friendly Fonts

By applying the same basics in MICR, i.e. introducing an inlaid easily traceable secondary features to font characters/glyphs, we can bring about automation/machine reading of contents along with visibility to human eye in print/display media. They shall be embedded with distinct features within the fields of a font character, which can be both in positive or negative videos or combination of both.

Even any of the other durable and safe means, in print friendly high-tech means such as magnetic electrostatic chemical and radioactive are not ruled out. Of course in those case the term 'OCR' itself may have to be substituted.

Since OCR product with these new secondary features shall be designed with sensing mechanism focused to the particular zone of distinction in the field of a graphemes without wasting resources on confusing, inert and noise zones /feature, it is easier to design while the output shall be quicker and errorless.

Another advantage in introducing secondary features into graphemes in print media is that it will relieve many problems in design of OCR product for variations in colour, glyph size, style etc. while giving a font and print designer a free hand to their varieties as secondary features can even be made independent of normal optical visibility.

Forthcoming Contents

Being a newer introduction, it is for forthcoming contents in print media and not for existing contents in printed archives.

Early Deployment

To provide full fledged scope of freedom to font authors for their individuality and to cater their own specific vending zones no suggestions initiated at this stage other than a request to every one in digital planet to look into the scope for innovation and economics. This can cut short the take-off time.

Standardisation

For popularization of any product, standardization and universality is a natural prerequisite leading to general economics and convenience. This should be true for OCR products too. Therefore the developers of OCR-FRIENDLY fonts embedded with unique attributes can come to a common platform for standardizing their process and products so that plurality of vendor specific products has to be avoided since customer specific interests will always prevail upon in long term business.

I P Rights

This basic innovative theme of having a secondary feature in each of font grapheme cannot have intellectual property rights because it is merely a generalized idea re-presented. This lacks newness/state-of-art creative intelligence. But the design rights over a fonts' looks shapes and styles along with process/design of producing a secondary easily traceable visible or other feature remains with the author. Creation of distinct attributes between characters of a set/group when achieved through an
innovative **state-of-art** process in the components of printing and reading like font printing ink or else and in the nature of medium over the print is made, the IP rights can still be enforced.

**Tamil and Universality**

Though the embedding co-features for OCR-FRIENDLYNESS by virtue of universality, it shall be applicable to fonts/glyphs/graphemes of every language of the world, but when introduced in our Tamil as a pioneering work we can attract admiration from digital global villages. There will side effects and problems to be faced by Unicode Consortium in their monster Universal Character Set though their character-encoding scheme are not connected with shapes and looks and other features of fonts and graphemes.

**Conclusion**

I hope by virtue of its rationale printed graphemes embedded with secondary co-features will soon shine with its usefulness in the digital sky and contents with this innovation employed will be flooded with richness of every language, including Tamil, as an avalanche.

The apparent fallout in having maiden digital Tamil contents initially from OCR friendly prints they can easily be converted/used in any other output mode as Sonics or Vision. If we find in the roadmap of having speech to text and text to speech applications until the initial sonic input route is less efficient and more cumbersome, the OCR route can be useful.
Creation of annotated Tamil handwritten word corpus for OHR

MILE Lab, Department of Electrical Engineering, IISc, Bangalore, India.
{nethra, archana, shashikiran, agr} @mile.ee.iisc.ernet.in

Abstract
Annotated datasets form a critical aspect in the development of robust technology for handwriting recognition and can be used for comparing results of different techniques used by various research groups. This paper describes the efforts at MILE lab, IISc, to create a database for the design and development of Tamil Online Handwritten Recognition. 100,000 words have been collected from 500 writers in Tamil, so that as much variations in writing style is captured. The data collected incorporated all the symbols (base characters, Indo-Arabic numerals, punctuations and other symbols). An annotation tool has been developed which helps the study of various styles of writing, stroke directions and presence of delayed strokes. Quality tags like class A, B, C etc has been assigned to the words accordingly. The annotated data is stored in a standard XML format defined by OHWR Consortium.

1. Introduction:
Databases are of great importance in any field of research, and handwriting recognition is no exception. A good database of handwritten data can be used to train and evaluate the performance of the recognition engine. Databases for scripts like Roman and Chinese already exist, whereas no such databases exist for Indic scripts. The database collected at MILE lab, IISc contains a comprehensive collection of words in Tamil, collected from many native Tamil people. Predefined word lists have been used to collect data, where the word list covers all the characters in the language. Here the focus is to develop a comprehensive database to support the development of a robust recognition engine. These databases facilitate comparison of different engines and also allow researchers to focus on recognition methodologies. A large database helps in removing bias of the engine towards particular styles of writing.

Tablet PC and G-Note have been used to collect data. The writer writes with an electronic pen on the electrostatic pressure sensitive writing surface of a Tablet PC or G-Note. The device captures the movement of pen tip on its screen in terms of x, y co-ordinates, sampled at equal intervals of time. It also captures the PEN_DOWN and PEN_UP information. The recognition is challenging because of varying styles of writing the same character. This paper describes how the database of 100,000 words has been collected from different schools and colleges, which involved major field work.

The collected data is annotated at the word, stroke group and akshara level using an annotation tool [2] developed by MILE lab. An akshara in Indian languages is a cluster of graphemes that need to be considered together to obtain the correct Unicode representation. Aksharas can be consonants (C), vowels (V) or a combination of them such as CV, CCV and so forth. The output of annotation is stored in the standard XML format [3] which was proposed by the OHWR consortium.
2. OHWR Consortium funded by TDIL:

A consortium made project was funded by Technology Development for Indian Languages (TDIL), Department of Information Technology, Government of India in January 2007 for research on online handwriting recognition. The project aims at developing Online Handwriting Recognition (OHWR) engines for Tamil, Kannada, Malayalam, Telugu, Bangla and Devanagari scripts. We at MILE Lab, IISc, are developing Tamil and Kannada engines. The academic partners of this project are IIT Madras, ISI Kolkata, and IIIT Hyderabad. The private and public industry partners are Learnfun Systems Chennai, CK Technologies Chennai and CDAC Pune.

3. Characteristics of Tamil handwriting:

Tamil compound characters (aksharas) are formed by graphically combining the symbols corresponding to consonants and vowel modifiers using well defined rules. Segmentation of words in these languages is more feasible than it is for English cursive writing as the characters are written separately without much overlap between them. In Tamil script, the majority of vowel modifiers are written as separate symbols and hence they are recognized separately.

4. Selection of complete constituent symbols:

Tamil: Tamil script comprises 313 characters. Of these, 12 are pure vowels and 23 pure consonants. Thus there are totally 12*23 = 276 consonant-vowel combinations. Apart from these, there are two additional symbols. The set of pure vowels in Tamil and its corresponding transliteration in English is depicted in Fig 1.

![Figure 1. Set of Vowels in Tamil](image)

We have established that only 155 symbols are required to represent all the 313 characters. The details are given below,

1) The vowel modifier for /A/ is depicted by a separate symbol and is written to the right of the consonant. Treating this vowel modifier as a separate class reduces the number of classes. A consonant /T/ combined with the vowel modifiers /a/ and /A/ are shown in two different rows of Fig 2.

![Figure 2. Consonants /Ta/ and /TA/](image)

2) Vowel modifiers of /i/, /I/ and /u/,// create new symbols when combined with the consonants. These new symbols are treated as different classes, thereby adding to the total number of classes. An example of this is shown in Fig 3.

![Figure 3. Consonants /Ti/, /TI/, /Tu/ and /TU/](image)
3) The vowel modifiers of /e/, /E/, /ai/ are separate symbols written to the left of the consonant. These symbols are also treated as separate classes, further reducing the number of classes. Fig 4 shows an example of a consonant in combination with these vowel modifiers.

![Figure 4. Consonants /Te/, /TE/ and /Tai/](image)

4) The vowel modifiers of /o/, /O/ have two separate symbols which are written on either side of the consonant. The consonant combined with vowel /o/ will have the modifier of /e/ to its left and the modifier of /A/ to its right. Similarly a consonant combined with vowel /O/ will have the modifier of /E/ to its left and the modifier of /A/ to the right. Since these symbols are already handled separately, the number of classes reduces further. Example of a consonant combined with these vowel modifiers is shown in Fig 5

![Figure 5. Consonants /To/ and /TO/](image)

5) The vowel modifier /au/ also has two symbols with one written on either side of the consonant. The symbol to the left of the consonant is the same as the modifier of /e/ and the symbol to the right is the same as the consonant /La/. These two symbols are already handled separately, similar to case 4, which also causes a reduction in the number of classes. A consonant combined with vowel modifier of /au/ is shown in Fig 6.

![Figure 6. Consonant /Tau/](image)

Along with the characters, special symbols like full stop and question mark are also incorporated in the symbol list. It is to be noted that in modern Tamil script, Tamil numerals are rarely used. Hence these symbols are not included in our dataset. Hindu-Arabic numerals have been included, and treated as special symbols in our work. The words have been carefully chosen so as to represent all possible symbols used in modern Tamil script.

5. Data Collection for Tamil OHR:

5.1 Criteria for selection of acquisition devices:

The devices used for data collection are the Tablet PC and G-Note. G-Note is more suitable for field work as it is sturdy, affordable and easy to carry. It is also easy for the user to write on a G-Note as the feel is the same as writing on normal paper or pad. The data collected in G-Note is stored as .TOP files. Tablet PC is suitable for individual use. It is heavy and difficult to carry. Also since it is expensive, it cannot be used for field work. The TabletPC data is stored in .txt format.
These devices are shown in Fig 7. Predefined word lists have been used to collect data. A Tamil sample handwritten pages is shown in the Fig.8.

5.2 Selection of Writers:

The criteria for selecting writers for data collection was that the person should be a native writer of the language and one who is currently writing regularly. Students and teachers were primarily chosen for data collection as they write regularly.

6. XML Standard for Annotation:

The output of annotation is stored in a standard XML format which has been defined by the OHWR consortium [3]. This standard XML includes all the details about the data, such as the writer details, the device information, the number of pages and words. The words are truthed at the word level annotation. The aksharas and stroke groups are truthed at the character level annotation. All this information is stored in the XML. The XML also contains information about the quality assigned to each word, akshara, stroke group and stroke. This facilitates separation of Class A/good data from the Class R/reject data.

7. Annotation Details:

Once the data is collected, the first process is to do the word level annotation. The collected set has multiple words on each page; hence determined word boundaries are to be used to obtain the strokes of a word. In word level annotation, each word is labeled, using a tool developed by IIIT Hyderabad. The output is stored in a standard XML format defined by the OHWR consortium.

Next is the character level annotation, where the output of word level annotation is given as input. In character level annotation, words are separated into strokes, stroke groups and aksharas and they are labeled. Quality tags are also assigned to them based on the direction of writing, stroke order and validity of strokes. The output at this level is also stored in the standard XML format. This annotation at character level is performed using a tool developed at MILE Lab, IISc [2].

8. Quality labels for Strokes, Stroke groups and Aksharas:

The strokes, stroke groups and aksharas are assigned various quality labels based on the nature of writing. The labels are defined as follows:

Class A: Denotes words written correctly with the expected number of strokes and in the expected direction. They are automatically segmented correctly by the segmentation module. Based on the statistics of writings from a huge number of native writers, there are multiple sets of stroke sequences valid for many stroke groups.
Class B: Denotes words which require manual segmentation and stroke groups with 10% or less overlap.

Class C: Denotes words where two or more normally separate strokes are written as a single stroke or vice-versa. It also includes strokes with overlap greater than 10% and delayed strokes.

Class D: Denotes words with extraneous strokes or overwriting and strokes written in the opposite direction. However, the resulting stroke groups must have the potential to be properly recognized using offline features.

Class R: This is the reject class, containing wrong words and/or strokes for which the likelihood of recognition is very low.

9. Conclusion:

This paper describes how the database has been created for Tamil Online handwriting recognition. The process of creating a reduced symbol list, which includes all the basic symbols of the character set has been described. The focus is on the process of collecting data, the devices used the criteria for selection of writers and why the reduction in number of symbols is required. This paper also tells how the data can be annotated for further use by researchers.

10. Acknowledgment:

This entire data collection effort was funded by Technology Development for Indian Languages (TDIL), Department of Information Technology (DIT), Govt. of India. We thank Mr. Suresh, Mr. Rituraj, Ms. Chandrakala, Ms. Archana, Ms. Saranya and Ms. Sountheriya for their efforts which made this task successful. We thank Prof. Deivasundram (University of Madras), AVM Matriculation Higher Secondary School, Govt. Boys Higher Secondary School, Sulur, Presidency College Triplicane, Virugambakkam, Chennai and IIT Madras for contributing to the data set. We also thank Dr. Anoop Namboodiri for giving us the word level annotation tool.

References:


Difficulties in developing OCR for Tamil documents and an Integrated OCR Solution

R. DhivyaBharathi  
IT Department, Pre-final year student  
Bannari Amman Institute of Technology  
Erode – 638 401.  
Email id: divyit@gmail.com

A. BalaMurugan  
EEE Department, Pre-final year student  
Anna University  
Tiruchirappalli – 620 024.  
Email id: rajiv.divy@gmail.com

Abstract:
In this paper, the major difficulties arising in the development of OCR for Tamil language is presented with solutions. We have proposed a new approach of using a Neural Network integrated with an Encoded Character String Dictionary. Multiclass Hierarchical Support Vector Machines (MHSVM), Hidden Markov Model (HMM) and Radial Bass Function Neural Network (RBFNN) are used for accurate character recognition.

As MHSVM has shown excellent character recognizing results due to a strong mathematical foundation, it is chosen for the recognition process. HMM is chosen for recognizing online written letters.

RBFNN is integrated with an Encoded Character String Dictionary and trained using the output of MHSVM and HMM for accurate character recognition. The three algorithms used here facilitate the recognition of Online, Offline and Printed Tamil characters which pave the way of developing an Integrated OCR. The main advantage of developing an Integrated OCR is to decrease the memory requirement and to increase the processing time. The proposed model is expected to be 90%-95% accurate.

Keywords: Optical Character Recognition (OCR), Hidden Markov Models (HMM), Multi Hierarchical Support Vector Machines (MHSVM), Radial Basis Function Neural Network (RBF-NN), Encoded Character String Dictionary.

1. Introduction
The common problem in all type of algorithms used in OCR is the set of letters in Tamil which look very much the same, except for some minor differences (Example: la and va). And we face more problems with the slanting letters. Straightening these slanting letters introduces more noise due to the digital nature of the pixels. If we keep them in their original shape, splitting them into letters have to be handled carefully. These problems can be handled by looking at the previous letter or group of letters to decide the next letter using a dictionary of words. MHSVM provides excellent Offline and Printed character recognition. MHSVM requires little prior knowledge which saves a lot of time in training the neural network and minimizes risk. HMM is chosen for recognizing online written letters. Pen-up strokes are modeled using a left-to-right HMM and inter-symbol strokes are modeled explicitly in two states. This shows low performance in case of high lexicon size which can be eliminated using a
statistical method. The statistical tool can be provided by the MHSVM explained earlier. HMM also avoids the segmentation problem, which is extremely difficult and error prone.

2. Proposed OCR Algorithm

The accuracy or efficiency of OCR purely depends on the algorithm we deploy. The efficiency decreases when an algorithm fails to identify a character or if the algorithm detects an unrelated character. We have proposed a method where we can fuse two pattern recognition algorithms and evaluate the efficiency of OCR. Before fusing, the scanned document is preprocessed.

The steps in preprocessing involves

1) Histogram equalization and Gabor Filtering
2) Binarisation
3) ROI extraction
4) Region Probe Algorithm

2.1 Preprocessing Steps

This section describes the preprocessing steps of the scanned document in detail.

2.1.1 Histogram Equalization and Gabor Filtering

The scanned document is first applied to Histogram equalization and Gabor Filtering. Histogram equalization usually increases the local contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. Then the Gabor filter is applied to the scanned document obtained by the previous step by spatially convolving the image with the filter. A Gaussian function [24] multiplied by a harmonic function defines the impulse response of the linear filter, the Gabor filter. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter’s impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function.

\[
\text{Where, } x' = x \cos \theta + y \sin \theta \text{ and } y' = -x \sin \theta + y \cos \theta
\]

In this equation, \( \lambda \) represents the wavelength of the cosine factor, \( \theta \) represents the orientation of the normal to the parallel stripes of a Gabor function, \( \Psi \) is the phase offset, and \( \gamma \) is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function.

2.1.2 Binarisation

The binarisation process involves analyzing the grey-level value of each pixel in the enhanced image, and, if the value is greater than the global threshold, then the pixel value is set to a binary value one; otherwise, it is set to zero.

2.1.3 ROI Extraction

We perform morphological opening on the grayscale or binary image with the structuring element. We also performed morphological closing on the grayscale or binary image resulting in closed image. The structuring element is a single structuring element object, as opposed to an array of objects for both open
and close. Then as the result this approach throws away those leftmost, rightmost, uppermost and bottommost blocks out of the bound so as to get the tightly bounded region just containing the bound and inner area.

2.1.4 Region Probe Algorithm

After all the above process, the image is passed to the segmentation phase, where the image is decomposed into individual characters. For this we have used region probe algorithm.

2.2 Algorithm Fusion

Then we have fused two algorithms meaning that both the algorithms are taken into consideration. While fusing the two algorithms the following points are taken into consideration

1) If one algorithm fails to identify a character, another algorithm may support in identifying the character.
2) If one algorithm gives wrong character another may give a correct one.
3) The possibility for same wrong identification by both the algorithms is less.
4) If one algorithm gives wrong result the decision of choosing the correct result is done by neural network which is discussed later in the paper.

We have chosen SVM, HMM for the fusion and discussed in the rest of the section.

2.3 Hidden Markov Model (HMM)

Hidden Markov Models are suitable for handwriting recognition for a number of reasons [8]. The importance of HMMs in the area of speech recognition has been observed several ago [25]. In the meantime, HMMs have also been successfully applied to image pattern recognition problems such as shape classification [26] and face recognition [27]. HMMs qualify as suitable tool for cursive script recognition for a number a reasons. First, they are stochastic models that can cope with noise and pattern variations occurring is human handwriting. Next, the number of tokens representing an unknown input word may be of variable length.

Moreover, using an HMM-based approach, the segmentation problem, which is extremely difficult and error prone, can be avoided. This mean that the features extracted from an input word need not necessarily correspond with letters. Instead, they may be derived from part of one letter, or from several letters. Thus the operations carried out by an HMM are in some sense holistic, combining local feature interpretation with contextual knowledge. Finally, there are standard algorithms known from the literature for both training and recognition using HMMs. These algorithms are fast and can be implemented with reasonable effort. Kundu and Bahl built an HMM for the English language [28].

However, they require the input word being perfectly segmented into single characters.

The Hidden Markov Model is a straightforward generalization of ordinary probability distributions to the case of randomly generated sequences of discrete or continuous-valued events. A discrete density HMM produces strings $O=O_1...O_T$ of symbols form a finite alphabet $[V_1,...,V_k]$ while the continuous density version creates sequences of real-valued feature vectors $x \in \mathbb{R}^d$. The generation of the observable output $O$ of the model is controlled by a doubly stochastic process. At each time instant $t = 1,...,T$ the model is in one out of $N$ possible states $[S_1,...,S_N]$. The state $q_t$ taken by the model at time $t$ is a random
variable which depends only on the identity of its immediate predecessor state. According to this assumption the state distribution is completely determined by the parameters.

\[ \pi = p(q_t = s_i) \] and \[ a_{i,j} = p(q_t = s_j \mid q_{t-1} = s_i) \]

in other words, the vector \( \pi = (\pi_1, \ldots, \pi_N)^T \) of initial probabilities together with the \((NxN)\) matrix \( A = [a_{i,j}] \) of transition probabilities form a first-order Markov chain.

The actual state sequence taken by the model serves as a probabilistic trigger for the production of the output sequence. The \( q_t \) themselves, however, remain hidden to an observer of the random process. According to a second model assumption, the probability distribution of an output symbol \( O_t \) (or an output vector, respectively) depends solely on the identity of the present state \( q_t \); thus, the distribution parameters.

\[ b_{j,k} = b_j(v_k) = p(O_t = v_k \mid q_t = s_j) \] of a discrete density HMM constitute a \((N \times K)\) probability matrix \( B = [b_{j,k}] \). Consequently, the behavior of an HMM with discrete output is entirely specified by the cardinality \( N \) of the state space, the alphabet size \( K \) and a parameter array \( \lambda = (\pi, A, B) \) of non-negative probabilities, obeying the \((1 + 2N)\) normalization condition

\[ \sum_i \pi_i = \sum_j a_{ij} = \sum_k b_{j,k} = 1 \]

Any of the state-dependent probability density functions (PDF) \( b_j(x) = p(O_t = x \mid q_t = s_i) \) of an HMM with continuous output can be reasonably well approximated by a convex combination

\[ \sum_{k=1}^K b_{j,k} g_{j,k}(x) = \sum_{k=1}^K b_{j,k} N(x \mid \mu_{j,k}, \Sigma_{j,k}) \] of multivariate Gaussian PDFs. The huge amount of statistical parameters found in a continuous mixture HMM as defined above - the model includes estimates of a distribution mean \( \mu_{j,k} \in \mathbb{R}^d \) and a covariance matrix \( \Sigma_{j,k} \) for each of \((N.K)\) index pairs can be drastically reduced if all state-dependent sets \( \{g_{j,k} \mid K=1,\ldots,K\} \) of mixture components are pooled into one global collection of PDFs. The resulting type of model is termed semi-continuous HMM; its output distributions

\[ b_j(x) = \sum_{k=1}^K b_{j,k} g_k(x) \] with \( g_k(x) = N(x \mid \mu_k, \Sigma_k) \) all share the same global set \( g = \{g_1, \ldots, g_k\} \) of Gaussians regardless of the state index \( j \). The semi continuous model is therefore characterized by the statistics \( \lambda = (\pi, A, B, g) \), where density function \( g_k \) is represented parametrically by its mean vector \( k \mu \) and covariance matrix \( \Sigma_k \). Evidently, our notation suggests that \( \pi, A, \) and \( B \) can be interpreted as an ordinary discrete 54 HMM and \( g \) as the codebook of a \( K - \) class probabilistic (soft) vector quantizes, transforming continuous feature vectors \( x \) into a likelihood array \( (g_1(x), \ldots, g_k(x))^T \).

### 2.4 Support Vector Machines (MHSVM)

The utilization of support vector machine (SVM) [9], [10] classifiers has gained immense popularity in the last years. SVMs have achieved excellent recognition results in various pattern recognition applications [10]. Also in offline optical character recognition (OCR) they have been shown to be comparable or even superior to the standard techniques like Bayesian classifiers or multilayer perceptrons [11]. SVMs are discriminative classifiers based on Vapnik's structural risk minimization principle. They can implement flexible decision boundaries in high dimensional feature spaces. The implicit regularization of the classifier’s complexity avoids over fitting and mostly this leads to good generalizations. Some further properties are commonly seen as reasons for the success of SVMs in real-world problems: the optimality
of the training result is guaranteed, fast training algorithms exist and little a-prior knowledge is required, i.e. only a labeled training set. Here, we provide a brief introduction to support vector classification.

For more details and geometrical interpretations please refer to the standard literature, e.g. by Burges [9] or Cristianini and Shawe-Taylor [10].

Consider a two-class classification problem and a set of training vectors \( \{ p_i \} \) with corresponding binary labels \( S_i = 1 \) for the “positive” and \( S_i = -1 \) for the “negative” class. In classification an SVM assigns a label 'S' to a test vector T by evaluating

\[
f(T) = \sum_i x_i S_i K(T, P_i) + b \quad \text{and} \quad S^* = \text{sign}(f(T))
\]

The weights \( a_i \) and the bias \( b \) are SVM parameters and adopted during training by maximizing

\[
L_D = \sum_i x_i - (1/2) \sum_{i,j} x_i x_j S_i S_j K(P_i, P_j)
\]

under the constraints \( 0 \leq x_i \leq C \) and \( \sum_j x_i S_i = 0 \)

with \( C \) a positive constant weighting the influence of training errors. \( K(\cdot, \cdot) \) is the kernel of the SVM. A solution for the \( a_i \) implies a value for \( b \). The SVM framework gives some flexibility in designing an appropriate kernel for the underlying application. Many implementations of kernels have been proposed so far, one popular example is the Gaussian kernel

\[
K((P_i, P_j)) = \exp(-r \| p_i - p_j \|^2)
\]

If \( K(\cdot, \cdot) \) is positive definite, (1)–(2) is a convex quadratic optimization problem, for which the convergence towards the global optimum can be guaranteed. However, obtaining this solution for real-world problems can be quite demanding and requires sophisticated optimization algorithms like chunking, decomposition or sequential minimal optimization [10]. Usually \( a_i = 0 \) for the majority of \( i \) and thus the summation in (3) is limited to a subset of the \( P_i \), which therefore is called the set of support vectors. Extensions of the binary classification to the multi-class situation are suggested in several approaches [9], [12].

2.5 Radial Bass Function Neural Network (RBF-NN)

To improve the accuracy, we have trained RBFNN with the output of both the algorithms. Different samples of Tamil Characters are taken and given as input to both HMM and SVM. If HMM or SVM gives a false character, the neural network is trained with the weightage of both the algorithms and the actual character. This process is done for all the possible false recognition of the two algorithms. During OCR When both the algorithms not giving same character, trained RBFNN is used to retrieve the actual character. This way we can increase the accuracy of OCR. Radial Basis Functions emerged as a variant of artificial neural network in late 80’s.RBF’s are embedded in a two layer neural network, where each hidden unit implements a radial activated function [13].

Due to their nonlinear approximation properties, RBF networks are able to model complex mappings, which perception by means of multiple intermediary layers [14]. Radial basis networks can require more neurons than standard feed forward back propagation networks, but often they can be designed in a fraction of the time it takes to train standard feed forward networks. They work best when many training vectors are available. RBF networks have been successfully applied to a large diversity of applications including interpolation [15], image restoration [16], shape-from-shading [17], 3-D object modeling [18], data fusion [19], etc.
2.6 Encoded Character String Dictionary

Dictionaries [15] obviously look like texts and share many features with other types of texts. However, users typically do not read a dictionary linearly, but access entries on the basis of a key (the headword) in order to retrieve various fields of information associated with that key (pronunciation, grammatical information, etymology, definitions, etc.). Electronic dictionaries now commonly available on CD ROM. In addition, although the display on the screen still looks more or less like a text, the internal representation is rarely that of a linear text. Here, a list of tamil words can be stored which will be used to find the occurrence of next letter of a word. This will enable the Neural Network algorithm to guess the word with more efficiency in case of confusing letters like la and va. These encoded dictionaries can be integrated with the neural network for efficiency in recognizing the characters.

3. Conclusion

In this paper we have proposed a new method of Tamil OCR where we have fused three algorithms to get the maximum possible efficiency by integrating it with a encoded character string dictionary. Our work primarily deals with choosing a better algorithm, and fusing them finally to attain better accuracy. We have chosen HMM, SVM to be fused and finally used Neural network to predict the correct character when there arises a situation where two algorithms yield two different characters. Our future work is to increase the efficiency more by increasing the effectiveness of neural network.

4. References


Recognition of Ancient Tamil Characters in Stone Inscription

S. Rajakumar, Asst. Professor, Department of ECE
Dr. V. Subbiah Bharathi, Dean (A), Department of ECE
J. Navarajan, Professor, Department of ECE
DMI College of Engineering, Chennai-602103, India
E-mail: rajkumarmeae135@yahoo.co.in

Abstract

In our project, we simulate a technique for recognizing ancient Tamil characters in stone inscription of various centuries. Many medical mysteries and valuable historical secrets are hidden in stone inscription which only few people can understand. Though there are number of character recognition, there is no efficient system for Tamil language which has a large character set and huge amount of variation. In particular, there is no system that addresses either the morphological characteristics of stone inscriptions or accounts for ancient characters. We propose a 3-stage approach to increase the recognition efficiency of characters. The image of the sculpture is captured and converted to required dimensions. Segmentation of the character is done using Chan-vese algorithm which uses active contours for separating the characters from background. Chan-vese algorithm has resistance to the morphological characteristics of stone inscriptions as it uses active contours. The noise in the image is removed using the area property. Zoning of characters is done by which each character is divided into six regions each viz., three horizontal and three vertical regions. Feature extraction is carried out by extracting the geometric features of the character contour. These features are based on the basic line types that form the character skeleton. 54 features (nine for each region) representing the number and normalized length of line segments of is extracted. The feature vectors so generated are applied to neural network (which is trained in advance with the features extracted from the characters of training database) for pattern recognition. Post processing of the pattern obtained is done to identify the exact character.

Objective: The main objective of this thesis is to convert ancient Tamil characters of various centuries to modern character by training the computer. This thesis explains how the characters are extracted and recognized using neural networks.

Problem Definition: Though there are a number of Character Recognition systems available, Character recognition in Stone Inscription is still a challenging task due to the following factors:

- Texture of inscriptions
- Noise present in image
- Complex characters
- Matching characters of different centuries
- Large character set
- Size of letters etc.
Existing Systems:

Edge detection is performed by computing gradients in an image. The contour of points with maximum gradient forms the edges in the image. Various operators used for these systems are Sobel operator, Roberts operators, Prewitt operators, Canny operator etc. These systems cannot be applied to stone inscriptions as the surface of the stones has ruggedness.

As shown in the above figure edge detection systems cannot be used in stone inscriptions due to its texture. Also the morphological characteristics of the stone inscriptions must be addressed during segmentation. In Thresholding, the mean values of the pixel values are calculated. Then, the histogram of the image is obtained. The mean value of the pixels is applied over the histogram to obtain the threshold value. Once the threshold value is obtained, all pixels having intensity greater than threshold are converted to the value of 255 and rest of the pixels is converted to 0. By performing the above operation, Binary image of the image is obtained having the value of 255 for characters and 0 for background of the image.

As shown in the above figure, when thresholding is applied the neighbours of the character is also taken as character due to the degradation function of the image. When the degradation function becomes significantly higher, the character cannot be segmented succesfully.

Proposed System:

In our project, we propose a morphological segmentation that utilises the features of adaptive thresholding combined with dilation of centroids that eliminate the disadvantages of classic thresholding methods. The various steps involved in it are

- Specifying the ROI
- Morphological segmentation
  - Conversion to Gray image
  - Adaptive thresholding
  - Applying Median filter
  - Use of Centroids
Specifying the ROI: Filtering a region of interest (ROI) is the process of applying a filter to a region in an image, where a binary mask defines the region. For example, you can apply an intensity adjustment filter to certain regions of an image. The multiROI function is used to obtain the region of interest from the image. The number of region of interest must be specified prior to running the program. Once it is specified, a figure window opens as shown in the image. The required region is selected by the polygon outline. The mask representing the region of interest is formed as shown in figure. This mask is used to extract the required character from the image.

Morphological segmentation: A grayscale image (also called gray-scale, gray scale, or gray-level) is a data matrix whose values represent intensities within some range. MATLAB stores a grayscale image as an individual matrix, with each element of the matrix corresponding to one image pixel. In RGB format, each pixel has three values one each for Red, Green and Blue colours. The mean value of the image for each colour are different and hence thresholding cannot be applied directly for RGB image. Hence it is converted into gray image before applying thresholding. The Gray value of the pixel can be calculated as

\[ I = .3(R) + .59(G) + .11(B) \]

Where,

- \( I \) = Gray image intensity value
- \( R \) = Red component of RGB image
- \( B \) = Blue component of RGB image
- \( G \) = Green component of RGB image

Once the gray image is obtained, adaptive thresholding is done to convert it into a binary image. The histogram of the image is calculated for the image and the threshold value is calculated. All pixels having value greater than the threshold are converted into 0 and other pixels are converted into 1 as shown in figure.

Median filtering is similar to using an averaging filter, in that each output pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel. However, with median filtering, the value of an output pixel is determined by the median of the neighborhood pixels, rather than the mean. The median is much less sensitive than the mean to extreme values (called outliers). Median
filtering is therefore better able to remove these outliers without reducing the sharpness of the image. The medfilt2 function implements median filtering. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. The noise which occupies large pixel values cannot be removed using median filter. The area property of the bounded region is used to remove these noises. The bwlable function is used for implementing the above step. The command, L = bwlable(BW,n), is used which returns a matrix L, of the same size as BW, containing labels for the connected objects in BW. n can have a value of either 4 or 8, where 4 specifies 4-connected objects and 8 specifies 8-connected objects; if the argument is omitted, it defaults to 8. The elements of L are integer values greater than or equal to 0. The pixels labeled 0 are the background. The pixels labeled 1 make up one object, the pixels labeled 2 make up a second object, and so on. The regionprops command is then applied over the matrix L to obtain the area of each bounded region.

The threshold of area is selected as 500 and bounded regions having area lesser than this are eliminated. The resulting image contains only the binarised characters. The region of interest mask obtained from the multiROI function is overlapped with the noise removed image to extract the specified character from the image. Once the required character is segmented, universe of discourse function is applied over it to obtain the boundary of the character. Due to this function, the variations in the size of letters is removed as all characters will be transformed into the same size.

The character after being resized is cross-correlated with every image in the database. In image processing, cross-correlation is a measure of similarity of two images as a function of a time-lag applied to one of them. This is also known as a sliding dot product or inner-product. The correlation is done in 2 dimensions and is obtained as

$$ R(i,d) = \frac{E[(X_i - \mu)(Xd - \mu_d)]}{\sigma_i \times \sigma_d} $$

Where

- ‘i’ is the input character
- ‘d’ is the image from database
- ‘E’ represents expectation value
- ‘X’ represents mean
- ‘\mu’ represents intensity of individual pixel
- ‘\sigma’ represents the standard deviation

The above equation is applied over every pixel in each image in database. The correlated values obtained are given as input to the neural network. The neural network selects the image that has maximum correlation in the database as the recognised character.
References:


Design and Evaluation of Omnifont Tamil OCR

**Tushar Patnaik**, CDAC Noida  
tusharpatnaik@cdacnoida.in

**Shalu Gupta**, CDAC Noida  
shalugupta@cdacnoida.in

**CV Jawahar**, IIIT Hyderabad  
jawahar@iiit.ac.in

**Santanu Choudhury**, IIT Delhi  
santanuc@ee.iitd.ernet.in

**A G Ramakrishnan**, IISc, Bangalore 560012.  
ramkiag@ee.iisc.ernet.in

IISc Bangalore has developed a recognition engine for Tamil printed text, which has been tested on 1000 document images of pages scanned from books printed between 1950 and 2000. IIIT Hyderabad has developed a XML based annotated database for storing the 5000 images of scanned pages and the corresponding typed text in Unicode. CDAC, Noida has developed an efficient evaluation tool, which compares the OCR output text to the reference typed text (ground truth) and flashes the substitution, deletion and insertion errors in different colours on the screen, so that the design team can quickly identify the issues with the OCR and make corrective steps for improving the performance. IIT Delhi has proposed and developed a novel scheme for segmenting only the text regions from document images containing pictures. The OCRs uses Karhunen-Loue transform (KLT) as features and a support vector machine (SVM) classifier with RBF kernel in a discriminative directed acyclic graph (DDAG) configuration. They assume an uncompressed input image of the document page, scanned at a minimum of 300 dpi with 256 gray levels (not binary or two-level). Tamil OCR currently gives over 94% recognition accuracy at the Unicode level, evaluated on over 1000 printed pages, some of them also containing old Tamil letters.

**The features of the OCRs are:**

15. Omnifont: Any normal font used by books is handled. We don't say it is font-independent, because ornamental or styled fonts cannot be handled.

16. **Merged Characters**: To a certain extent, the OCR is capable of identifying and segmenting the merger between two adjacent characters in a old, printed book.

17. **Noise Tolerance**: Certain types of breaks in the character are handled successfully.

18. **Old Tamil or Kannada Script**: The pre-1970 (prior to script revision due to E V Ramasamy Naicker) Tamil CV combinations such as NA, RA, nA, Nai, Lai, Lai and nai are all recognized, along with the revised representations of the same. Similarly, old Kannada (halegannada) characters of La and zha and their vowel combinations are all handled seamlessly.
19. **Unicode Output**: The output is given in UTF-8.

20. **Testing**: Both OCRs have been tested by CDAC, Pune using an annotated corpus of over 1,000 document images of pages scanned from books printed between 1950 and 2002.

21. **Consistency**: The OCRs produce consistent and graded performance with font, size and quality variations.

22. **Future Enhancement**: The current average performance of 94% for Tamil and 84% for Kannada at the character level is without the use of any language model for postprocessing. Thus, there is a good potential to improve the performance of both OCRs further.

Medical Intelligence and Language Engineering Laboratory has teamed up with Bookshare.org, an International non-profit organization, to provide Tamil and Kannada digital books (copyright free or permitted by authors) online to print-disabled people (visually challenged, old people with vision disabilities and people with other disabilities that make it impossible for holding a book and turn pages of it). A Text-to-speech engine in the respective language will also be provided to the registered user, who can then directly listen to the printed content on their desktop or laptop. We look forward to partners, who can give us copyright free books (hard or soft copies) or direct us to sources of the same. They are also welcome to directly partner with bookshare.org or Worth Trust at Chennai or Enable India at Bangalore.

Figure 1 shows a screen shot showing the performance of the system, as well as the convenience and use of the GUI. Figure 2 shows the confusion matrix shown by the evaluation tool, which helps in identifying common confusion and improve the OCR accordingly. Figure 3 shows the evaluation tool, comparing the XML annotated Tamil text for the page and the OCR’s output in Unicode. Use of such convenient tools accelerated the development of the OCR accuracy, and it is currently giving a performance of over 95% for good quality printed pages.

*Fig.1. A screen shot from the OCR, showing the input image and the output text.*
Fig. 2. Display of Confusion matrix for a page recognition.
Fig. 3. Evaluation Tool showing substitution, insertion and deletion errors.
A Novel Hybrid SVM-Neural Approach to Recognize Handwritten Tamil Characters

N. Shanthi
E-mail: shanthimoorthi@yahoo.com
K. Duraiswamy
E-mail: drkduraiswamy@yahoo.co.in
K.S. Rangasamy College of Technology
Tiruchengode, Tamil Nadu, India

Abstract

In this paper a Tamil handwritten character recognition algorithm based on hybrid SVM-Neural approach is presented. Initially SVM and Neural network is individually tested to know the performance of each classifier. The recognition accuracy of SVM is 90% and the recognition accuracy of neural network is 81%. So a hybrid approach is used to improve the recognition accuracy.

Data samples are collected from different writers on A4 sized documents. They are scanned using a flat bed scanner at a resolution of 300dpi and stored as grey scale images. Various preprocessing operations are done on the input document image to enhance the quality. The characters are segmented and normalized to uniform size of 64 X 64. These uniform sized characters are projected onto a grid of fixed size 8 X 8. The pixel density is calculated for each grid and used as the feature vector. These features are given to well known support vector machines in the first stage. Few characters with low recognition accuracy are selected and are given to neural classifier. Since few classes are given to the neural network the recognition accuracy is improved. The new hybrid approach yields a recognition accuracy of 91.25%.

1. Introduction

Machine simulation of human reading is one of the areas, which has been the subject of intensive research for the last three decades, yet it is still far from the final frontier[1][2]. Works on offline recognition of handwritten characters has been done well on Chinese, Arabic and few other scripts of the other nations[13][14][15]. However, there is less progress towards recognition of handwritten characters of Indic scripts. Recognition of handwritten Indian scripts is difficult because of the presence of vowel modifiers and compound characters. Most of the existing works are concerned about Devnagari and Bangla script characters, the two most popular languages in India. Few works are reported on the recognition of other languages like Telugu, Oriya, Kannada, Punjabi, Gujarathi [4] etc., There are few attempts to recognize printed or handwritten Tamil characters [3][6][8][9] [17][18]. This paper proposes a recognition system for handwritten Tamil characters. The paper is organized as follows. The following section presents a brief description of Tamil language. The proposed hybrid SVM-Neural approach is presented in section 3. Finally, in section 4 conclusion and future works are given.

2. Tamil Language

Tamil, which is a south Indian language, is one of the oldest languages in the world. Sanskrit has influenced it to a certain degree. Its history can be traced back to the age of Tolkappiyam, the earliest extant Tamil grammar generally to 500 B.C. Tamil language is a member of the Dravidian / South Indian
family of languages. Among the Dravidian language it is least influenced by 'Sanskrit' though there is a
certain degree of influence. Most Tamil letters have circular shapes partially due to the fact that they were
originally carved with needles on palm leaves, a technology that favored round shapes. The Tamil script
is used to write the Tamil language in TamilNadu, SriLanka, Singapore and parts of Malaysia.

Tamil is the official language of the Indian state of TamilNadu, classical language in India, and has
official status in India, Sri Lanka and Singapore. With more than 77 million speakers, Tamil is one of the
more widely spoken languages in the world. The script for Tamil Language was conceived since time
immemorial and the present form is after undergoing various changes at various periods of time. During
its long cherished existence, it underwent various changes, corrections, modifications and has arrived at
the present form based on its applicability and adoptability in the ever changing writing method and
devices from time to time. The present form of the Tamil script is believed to have been evolved based on
its “adoptability” for handwriting mechanisms.

2.1 Tamil Alphabets

Tamil alphabet consists of 12 vowels, 18 consonants, a special character (akh) and 216 combinational
alphabets obtained by combining the vowels and consonants. Tamil language is relatively simpler
compared to other Indian scripts. The rules for character composition are far fewer than in other Indian
scripts. The only category of composition allowed is of Consonant-Vowel type, where a structure
Corresponding to a consonant and another corresponding to vowel are combined to form a C-V type
character with a unique shape. Even though the total number of alphabets is 247, the graphical characters
required to represent all of them are only 106. The C-V combination results in unique shape when
\(\text{அ} \), \(\text{இ} \), \(\text{ஈ} \), \(\text{உ} \), \(\text{ஊ} \) are combined with consonants. When the remaining vowels are combined with consonants,
\(\pi\), then the combination results in horizontally isolatable structure. For example by adding \(\text{அ} \) with \(\text{ஜ} \) we
get \(\text{க} \). Similarly when all the consonants are added with \(\text{ஜ} \), we get 18 unique characters. Then by placing
the character \(\pi\) after these character we get one alphabet. Following \(\text{அ} \) we get \(\text{அ} \pi\); following \(\text{ச} \) we get \(\text{ச} \pi\);
following \(\text{த} \) we get \(\text{த} \pi\). In this manner there are eighteen alphabets with the character \(\pi\) coming after
different consonants. So the composite characters can be recognized as the sequence of \(\text{அ}, \text{ச}, \text{த}\)
characters. Due to this flexibility the total number of characters to be recognized is reduced to 106. If
these 106 characters are recognized then all the 247 characters can be recognized. In addition to this the
other unique characters to be recognized are \(\text{ே}, \text{ை}\) and \(\pi\).

3. Hybrid SVM-Neural approach

For relatively small class problems a single classifier system provides acceptable accuracies. But when the
number of classes is more, the recognition accuracy is less. So hybrid approach can be used to recognize
complex character recognition problem.

Classification with multiple classifiers has been an active research area since early 1990s. Many
combination methods have been proposed, and the applications to character recognition problems have
advantage over individual classifiers. Generally, a character recognizer involves three tasks:
preprocessing, feature extraction and classification.

3.1 Preprocessing

Various preprocessing operations are performed prior to recognition to enhance the quality of the input
image.
Preprocessing operations performed prior to recognition are:

1. **Thresholding** is to convert a gray scale image into binary by determining a value for gray-scale (or threshold) below which the pixel can be considered to belong to the writing and above which to the background; Here Otsu’s method of histogram-based global thresholding algorithm is used[11].

2. **Skew detection and correction** is to detect the skew present in the image and to correct it. A document skew detection algorithm using wavelet transform and horizontal projection profile [19] is used in this work.

3. **Line segmentation**, the separation of individual lines of text; Horizontal histogram profile is used for segmenting the lines[10].

4. **Slant correction**, Removal of angle between the vertical direction and the direction of the strokes; a new method for slant estimation is used based on the combination of the projection profile technique and wavelet transform.

5. **Character segmentation**, the isolation of individual characters; Vertical histogram profile is used for segmenting the characters[10].

6. **Skeletonization** is the process of peeling off of a pattern as many pixels as possible without affecting the general shape of the pattern [12]; Here Hilditch’s algorithm is used for skeletonization.

7. **Character size Normalization**, Size normalization is a transformation of an input image of arbitrary size into an output image of a fixed pre-specified size, while attempting to preserve structural detail[7]; Bilinear interpolation technique is used to convert the random sized image into normalized image of size 64X64.

Figure 1 shows the sample input image and Figure 2 shows the preprocessed image of the last line.

![Sample input image](image1)

*Figure 1. Sample input image*

![Preprocessed image](image2)

*Figure 2. Preprocessed image*
3.2. Feature extraction

Feature extraction is the problem of extracting the information from the preprocessed data, which is most relevant for classification purposes, in the sense of minimizing the within-class pattern variability, while enhancing the between-class pattern variability [20]. Selection of feature extraction method is the most important factor in achieving high recognition performance [5]. Here a simple feature of pixel density is used as feature. An 8 X 8 grid is superimposed on the character image, and for each of the 64 zones the average number of pixels is computed which results in 64 features. The pixel density varies from 0 to 64. The Figure 3 shows the zoning of character image.

![Figure 3. Zoning of character image](image)

3.3. Classification

The final goal of character recognition is to obtain the class codes of character patterns. On segmenting character or words from document images, the task of recognition becomes assigning each character or word to a class out of a pre-defined class set. Different classifiers are used for recognition. Support vector machine and artificial neural network are used in this work for classification.

3.3.1 Support vector machines

The SVM is a new type of hyperplane classifier, developed based on the statistical learning theory of Vapnik[16], with the aim of maximizing a geometric margin of hyperplane, which is related to the error bound of generalization. The application of SVMs to character recognition has yielded state-of-the-art performance. Initially SVM alone is used for training and classification of the Tamil characters.

SVMs are mostly considered for binary classification. For multiclass classification, multiple binary SVMs, each separating two classes or two subsets of classes, are combined to give the multiclass decision. Here “one-against-one” approach in which k(k−1)/2 classifiers are constructed and each one trains data from two different classes. For training data from the $i^{th}$ and the $j^{th}$ classes, the following binary classification problem is solved:

$$
\min_{w^i, b^i, \xi^i} \frac{1}{2} (w^j)^T w^j + C\left(\sum_i (\xi^i)_i\right)
$$

subject to

$$(w^j)^T \phi(x_i) + b^j \geq 1 - \xi^i, \text{ if } x_i \text{ in the } i^{th} \text{ class},$$

$$(w^j)^T \phi(x_i) + b^j \geq -1 + \xi^j, \text{ if } x_i \text{ in the } j^{th} \text{ class},$$

$$\xi_i \geq 0.$$

Here max-wins voting strategy is used for classification. The recognition accuracy using SVM is 90%.
3.3.2 Artificial neural networks

An artificial neural network is defined as a computing structure consisting of a massively parallel interconnection of adaptive neural processors. The main advantage of artificial neural networks lies in the ability to be trained automatically from examples and good performance with noisy data. Artificial neural networks have been widely used in the field of character recognition produced promising results.

Back propagation feed forward neural network (BPN) is used for training and testing the network. The recognition accuracy is only 81%. Neural network took large time to train and the recognition accuracy is also low. So we have decided to use the benefits of both the classifiers.

3.4. Introduction to SVM-Neural Classifier

Even though the recognition accuracy of SVM is higher when compared to the other classifiers, still there is a scope for further improvement. The analysis of the Tamil character set reveals that a few characters are closely identical to each other in structure. So when all the characters are given to SVM then there is a difficulty in recognizing the identical characters. The characters identified and their percentage of recognition using SVM is shown in table 1. For example when the classification result of character ஏ is analyzed, the result shows that many a times the character ஏ is recognized as ஓ or ல. This is because of the reason that these characters look identical in structure. Hence in order to solve this difficulty a hybrid two stage classifier approach is proposed.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Characters with low recognition accuracy</th>
<th>Similar characters</th>
<th>Recognition accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ஏ, எ, ர</td>
<td>ஏ, எ, ர</td>
<td>78.45</td>
</tr>
<tr>
<td>2</td>
<td>஧</td>
<td>஧</td>
<td>72.8</td>
</tr>
<tr>
<td>3</td>
<td>ஞ</td>
<td>ஞ</td>
<td>77.7</td>
</tr>
<tr>
<td>4</td>
<td>ண</td>
<td>ண</td>
<td>81.29</td>
</tr>
<tr>
<td>5</td>
<td>ன, த</td>
<td>ன, த</td>
<td>74.85</td>
</tr>
</tbody>
</table>

The hybrid approach consists of 2 stages. SVM is chosen as the classifier in the first stage since it produced the highest recognition accuracy. When the performance of the SVM classifier is analyzed it illustrated that the 64 × 64 sized image with overlapping zones produced the best recognition accuracy. So those 225 features are calculated and given as input to the SVM classifier in the first stage. The output of SVM revealed that the recognition accuracy of the characters ஏ, ஧, ஞ, ண and ன are low because they are identical to other characters in structure. The groups of such identical characters are (எ, ஏ, ர), (ஓ, ஒ), (ஞ, ல), (ழ, ல) and (ன, ல, ன).

In the second stage a separate classifier is designed for each group of identical characters. Both SVM and ANN are used in the second stage and the performances of both the classifiers are analyzed for all the groups. Table 2 shows the characters with low recognition accuracy, their identical characters and improved recognition accuracy using SVM-SVM approach. Table 3 shows the improved recognition accuracy of SVM-Neural approach. From the Tables it is evident that the performance of ANN is good in the second stage when compared to SVM. Hence in the proposed hybrid approach SVM and ANN are combined sequentially to improve the classification performance. This hybrid approach improves the overall recognition accuracy by 1.5%.
This hybrid approach improves the overall recognition accuracy by 1.5% from the original 90% and improves the recognition accuracy of each identical character by 6% to 12%. SVM and ANN are combined effectively in the proposed hybrid approach to improve the overall recognition accuracy and the recognition accuracy of identical characters.

The architecture of the proposed hybrid SVM – Neural approach is shown in Fig.4.

![Fig.4 Architecture of the Proposed Hybrid SVM-Neural Approach](image-url)
4. Conclusion

This study presents a novel hybrid system to recognize offline handwritten Tamil characters. The experimentation result shows that the simple SVM approach produces a cumulative recognition accuracy of 90% and the recognition accuracy for different characters varies from 72% to 99%. When neural network is used for recognition, the recognition accuracy varies from 63% to 98% with an overall recognition accuracy of 81%. But when SVM-neural approach is used the recognition accuracy is increased to 91.5%. This recognition accuracy is achieved with a simple feature of pixel densities in various zones of the images. The main recognition errors were due to abnormal writing and ambiguity among similar shaped characters. Future work can include more robust invariant features to achieve better discrimination power and neural classifier can be constructed for other characters.

References

9) S.Hewavitharana and H.C.Fernando, A Two Stage Classification Approach to Tamil Handwriting Recognition, Tamil Internet 2002, California, USA, 118-124, 2002.
10) N.Shanthi and Dr.K.DuraiSwamy. Preprocessing algorithms for the recognition of Tamil Handwritten Characters. 77-82, Third International CALIBER 2005, Kochi.
Effective Tamil Character Recognition in Tablet PCs using Pattern Recognition

Ferdin Joe J
Department of Computer Science & Engineering,
Einstein College of Engineering,
Tirunelveli, e-mail: ferdinjoe@gmail.com

Dr. T. Ravi
Department of Computer Science & Engineering,
KCG College of Technology,
Chennai. e-mail: travi675@yahoo.com

Prof. R. Velayutham
Department of Computer Science & Engineering,
Einstein College of Engineering,
Tirunelveli. e-mail: rsvel_kumar@yahoo.co.uk

Abstract

Tamil Character Recognition is done for many applications using various novel methodologies. But the recognition of Tamil characters in Tablet PCs is a challenging one. Various methodologies like SVM, Preprocessing, interpolation etc were used in the past but were not satisfactory to friendly-user point of view. In this paper we introduce a new methodology pattern matching for Tamil Character Recognition. This method was tested to be a better approach for Gujarati scripts. As our Classical Language Tamil characters also has more curls and twists like Gujarati scripts, we utilized the same methodology for Tamil Character Recognition. This methodology uses a series of operations on the digitally written character. Scanning, Inversion, Grayscale conversion and trimming are done in the preprocessing section. Back propagation using neural networks in the middle tier and at last pattern matching algorithms are employed. In this last phase instead of pattern matching done, we come forward with a modified pattern matching algorithm. We have termed this algorithm as Thellodai Line Fitting Algorithm. As a result of this series of operations, the recognized output is obtained. For tablet PCs the output was simulated in Matlab’s Neural Network and Image Acquisition toolbox. After analyzing the simulated results, the recognition rate of this methodology is found to be about 70 - 80% on an average where as the other methodologies were tantalizing between 60s and 70s for character recognition of the Classical Language Tamil.

Index Terms—Pattern Classification, Thellodai Algorithm

Introduction

This paper deals with the new methodology implied for the effective pattern matching of Tamil character recognition in tablet PCs. Tablet PCs English characters are easily recognized by just scribing on the touch module. But for Tamil Characters, the mechanism is not available as of now in the case of Tablet PCs. English characters are easy and faster to recognize because, it has only 26 characters. There lies a difficulty for Tamil Character Recognition. It has 247 letters excluding the old scripting letters. So there is a need to classify the Tamil characters in a simple way and reduce the character set to a minimal extent.

Gujarati scripts as in [1] are given with a set of matching. They are checked for templates and pattern matching of the particular character. This is possible for Gujarati scripts because the number of characters used are less. But when this technique is implied Tamil, certain difficulties are experienced in the
performance of execution. Tamil Characters are totally 247 in number, excluding the other language influential characters.

Similarly some methodologies are given for handwritten Tamil character recognition also. In section II, the difficulties in execution of the various methodologies are discussed. In section III, the newly proposed Thellodai line fitting algorithm is discussed in detail. In section IV, the proposed model is elucidated. In section V, the performance measures and the modules for real time implementation are discussed and finally section VI concludes the entire work with the possibilities to execute this project in real time.

Related Work

The base work is selected as [1], because Gujrati also has curvy writings similar to Tamil. So we are taking the technology used in [1] as the back bone for our proposed model. The model stated in [1] fails for Tamil because, Tamil characters are large in number and they take larger time when compared to the Gujrati scripts.

Various methodologies are being proposed for Tamil Handwritten Character Recognition. The models stated in [3], [5], [6], [7], has various good features for larger devices or enterprise level solutions. But these methodologies are found to be less effective when we take in the case for Tablet PCs.

Tablet PCs normally deal with the device driver synchronization, character compression and well classified set of characters. Tamil character set consists of 247 characters totally. In the previous models proposed in [3], [6] and [7], the classification is not efficient to the expected levels of Tablet PC synchronization. English characters find easier in the tablet PC because, it has only 26 singlet characters for classification. After thorough analysis on the Tablet PC architecture, the problem statement is framed as follows: “There lies a need for classifying all the 247 letters in Tamil to a maximal extent. The number of patterns used should be reduced to a minimal extent. A new algorithm based on line fitting for pattern recognition needs to be framed. This system should be developed in such a way that the device driver synchronization for Tablet PCs is to achieved.”

Thellodai Algorithm

As mentioned in the problem statement at the end of the previous section, we have developed a new algorithm based on line fitting. We term it as Thellodai Line Fitting algorithm. Thellodai pronounced /θɛɭəlˈɒðəɪ/. Thellodai means “Clear stream” in Tamil. As per this algorithm concerned, this has the well classified set of Tamil characters. All the 247 characters are classified efficiently and 35 patterns are framed. Many letters are recognized as a combination of two or more other patterns.

The Thellodai algorithm has a predefined trained set of 35 characters. These characters are termed as Frequently Used Patterns Uf, Rare Patterns Rp and Original Pattern Op.

Algorithm 1: Thellodai Line Fitting Algorithm

Function match()
{
Var Uf, Rp, Op;
Wait for input;
Set the combination;
Decide on the permutation;
If permutation is Uf
}
The sub patterns are matched;
}
If permutation is Rp
{
    Rp sets are matched;
}
If permutation is Op
{
    Move to next input;
}
If match is formed
{
    Indent to next space and wait for input
}
}

The above algorithm is framed specially for the Tamil character set. The patterns possible are reduced to a minimal set of just 35 patterns. These patterns are studied for the usage. According to this study, many of the characters are not used nowadays for casual usage.

So we don’t need to concentrate much on these characters. Thellodai algorithm is included as an additional feature in our proposed system and it will be good for rarely used characters also. The frequently used characters are indexed in an ajax based system and it will show the possibilities of character identification.

Proposed Model

The proposed model is based on the architecture mentioned in [1]. The system is simulated in matlab 7.0 using the image acquisition toolbox and neural network toolbox. The image acquisition toolbox is used for simulating the patterns and neural network tool box for simulating the frequently used pattern set.

*Fig 1: Expected output for the Tamil Character Recognition in Tablet PCs (Modelled).*

Fig 1 shows the expected output of the simulated set of data. For example, if we scribe “ah”, it will give the possibility for “aah” also. The main aim is to study the efficiency of handwritten character recognition for aharam, aaharam, eharam, eeharam, uharam, ooharam, aeharam, aeharam, oharam, ooharam and ouharam set of characters. In Fig1, the character “ah” is scribed by the user. The system identifies as “ah” and gives the possible extension for “aah” also. The scribed character is displayed in black and the permuted extension is displayed in grey.
The recognition sector deals with the following processes: Scanning, Inversion, Grayscale conversion and trimming are done in the preprocessing section. Back propagation using neural networks in the middle tier and at last pattern matching algorithms are employed. In this last phase instead of pattern matching done, we come forward with a modified pattern matching algorithm. We have termed this algorithm as Thellodai Line Fitting Algorithm. As a result of this series of operations, the recognized output is obtained.

Preprocessing Phase

The preprocessing phase has the steps of Inversion, grayscale and thinning. During inversion, the character input is converted to the inverse color. Then during the grayscale phase, it gets converted to black and white. During thinning the edges are fitted.

Supervised phase

During the supervised phase, the Neural network principles come into act. The back propagation and trained intelligence modules will analyze the output of the preprocessing phase. Then by using the trained intelligence, the extensions are predicted. A popular and simple NN approach to the problem is based on feed forward neural networks with back propagation learning. In the training step, we each training sample is represented by two components: possible input and the desired network's output given that input. After the training step is done, we can give an arbitrary input to the network and the network will form an output, from which we can resolve a pattern type presented to the network. As per the requirement of the Neural Network algorithm the information content that is area within the borders needs to be digitized. This is achieved by dividing this area into a 12 x 12 grid. This 12 x 12 grid is scanned, and every time 1 is detected that particular block is flagged. This procedure is referred to as digitizing.
The undigitized output is obtained in this phase. This output is now fed as input to the post processing phase.

![Digitized output as a result of the post processing phase](image)

**Postprocessing phase**

Post processing phase is the final phase for this proposed methodology. This phase implies the Thellodai algorithm for the efficient pattern matching of the digitized tamil characters.

![Output obtained as a result of Thellodai algorithm in the post processing phase.](image)
The purple colored cell in Fig 4 indicates that the extension is possible for this character for “key” and “keey” characters. When the user scribes on the purple extensions, it will automatically guide the user to incorporate the extension for the particular character to be finalized.

**Performance Measures**

The performance of the proposed methodology are found for the various subsets aharam, aaharam, eharam, eeharam, uharam, ouharam, aeharam, aaeharam, aiharam, oharam, ooharam and ouuharam.

The combination of 35 singlets is combined with each other and the characters are found for efficiency of matching. It is found that, the subsets of aharam, aaharam, aeharam, aaeharam, aiharam, oharam, ooharam and ouharam are in no need for any difficulty in recognizing the character. But eharam, eeharam, uharam, ouharam finds little difficulty. This is just for the reason for some characters like “Do” are in need of totally different shape transformation when compared with the other characters.

<table>
<thead>
<tr>
<th>Character Subset</th>
<th>Recognition rate in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aharam</td>
<td>80</td>
</tr>
<tr>
<td>Aaharam</td>
<td>77</td>
</tr>
<tr>
<td>Uyir</td>
<td>79</td>
</tr>
<tr>
<td>Mei</td>
<td>80</td>
</tr>
<tr>
<td>Eharam</td>
<td>72</td>
</tr>
<tr>
<td>Eeharam</td>
<td>70</td>
</tr>
<tr>
<td>Uharam</td>
<td>71</td>
</tr>
<tr>
<td>Ouuharam</td>
<td>69</td>
</tr>
<tr>
<td>Aeharam</td>
<td>76</td>
</tr>
<tr>
<td>Aaeharam</td>
<td>78</td>
</tr>
<tr>
<td>Aiharam</td>
<td>80</td>
</tr>
<tr>
<td>Oharam</td>
<td>75</td>
</tr>
<tr>
<td>Owharam</td>
<td>74</td>
</tr>
<tr>
<td>Ouuharam</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 1: Recognition rates of various subsets of Tamil character series.

![Graph showing the performance and recognition rates of various subsets of Tamil character set.](image)
As shown in Fig 5, it is evident that the efficiency of the character recognition is found to be nearing 80% for most of the character subsets. Only 4 out of the 14 subsets, the efficiency is in the early 70s.

The recognition rates for eharam, eeharam, uharam, ouuharam are found to be nearing the early 70s because there are chances for asymmetric matching. When comparing with the recently existing systems mentioned in [1], [3], [13], it is clear that our proposed model will be better in its every aspect. The fidelity of the letters will be more. Handwritten performance of the user will increase.

There are many advantages for this system. When comparing to the transliteration system, the user is in no need to have English phonetic proficiency to type in Tamil. The proper pronunciations and spellings are possible. There is an issue of confusion with the l-la and zha characters. This problem can be eliminated by using the proposed approach.

Tablet PCs as of now are not designed for Tamil Characters. This is because of the lack of proper hardware modifications. The modifications need to be done in such a way that Tamil characters can also be recognized. After this issue is fixed, real time implementation is possible for this proposed methodology.

Conclusion

From the above elucidated proposed methodology, it is clear that this technology is successful at the level of simulation. The efficiency for most of the cases comes above 75% and the Thellodai Line Fitting algorithm holds better than many of the existing pattern matching algorithms. The real time implementation needs the Unicode data of Tamil characters to be included in the Tablet PCs. If it is done so, then the real time implementation can be done. Tamil Tablet PCs are no more longer to wait.

References


**Ferdin Joe J** is currently a PG scholar with the Department of Computer Science & Engineering, Einstein College of Engineering, Anna University @ Tirunelveli. He graduated in Computer Science and Engineering from Anna University @ Chennai during 2009. He is a member of IAENG and Chennai IT Professional User Group. He is an independent consultant and providing IT solutions for Academicians and Researchers. His areas of interest include Data Mining, Bioinformatics, SOA, Web 2.0 and Network Security. He has published nearly 10 papers in International Conferences, 1 paper in International Journal and 5 papers in National Conferences. He has a passion towards dedicated research in core computing principles. To know more about him and his research work, please log on to http://www.ferdin.co.nr.

**Dr. T. Ravi** is currently Professor and Head with the Department of Computer Science & Engineering, KCG College of Technology, Anna University @ Chennai. He was previously with the Department of Computer Science & Engineering, Dr. MGR University, Chennai. He graduated in Computer Science & Engineering from Madurai Kamaraj University during 1991. He obtained Masters Degree during 1996 and PhD during 2009 in Computer Science & Engineering from Jadavpur University, Kolkata. He has nearly 18 years of teaching experience. He has published many papers in International Conferences and Referred International Journals. His areas of Interest include Data Mining, Networks and Bioinformatics. He is a recognized supervisor in Anna University, Coimbatore and Hindustan University, Chennai. Currently he is guiding research scholars from the above mentioned universities. To know more about him, please log onto http://www.travi.co.nr.

**Prof. R. Velayutham** is currently Professor and Head with the Department of Computer Science & Engineering, Einstein College of Engineering, Anna University @ Tirunelveli. He graduated in Computer Science & Engineering from Madurai Kamaraj University. He obtained his masters degree from Madurai Kamaraj University. He is currently pursuing PhD from Anna University @ Tirunelveli. He has published 20 papers in International and National Conferences and 1 paper in National Journal. His areas of interest are Cryptography and Network Security. He has nearly 10 years of experience in teaching. He was previously with Noorul Islam University, Nagercoil.
9

தமிழ் சிற்றகால காலக்குறிப்பிகள்
கலனிக் குறிப்பிறகாலாக
Design Of Intelligent Robotic Arm Writing Tamil For Physically Challenged
‘A boon for physical ailment’

Seima Saki S, ECE
seima.skct@gmail.com

Sabita Devi D, Mechatronics
sabi.skct@gmail.com

Sri Krishna College Of Engineering And Tech., Coimbatore,

Manoj S, Computer Technology,

PSG College of Technology, Coimbatore s.manoj89@gmail.com

Mr.R.Vimalathithan, Senior Lecturer,

Sri Krishna College Of Engineering And Tech, Coimbatore athivimal@gmail.com

Abstract
Physically impaired people find hard to subsist in this challenging world. Here forth is advancement in technology to provide comfort to handicapped and visually disabled to convey their thoughts through pen and paper. Inventions made so far didn’t prove a great amount of benefit to them. The working model proposed here is much novel and proves to be a well-worn technology. The central notion is to design a robotic arm that writes upon dictation. We use speech analysis techniques to convert it into text. Writing is brought by finding the mathematical equation of each alphabet, character or a number and converting into values that drives the arm. We use 3 degrees of freedom design for X,Y and Z axis motion. The values of the equations are stored in the memory of the micro controller and are sent to motors corresponding to the incoming text. Hence not only one but many styles of writing can be brought up by a single system by feeding their corresponding equation values. This not only assists disabled persons, but also supports people who need a proxy for their hand writing. This technique is much practical and shows best results.

Introduction
Researchers spend their valuable time in inventing novel methods to impede the disabilities of physically impaired. Previously developed devices were not that effective to make the needy feel relieved from all his necessities. Some tasks like signing the documents and writing in his own hand writing seemed to be impossible. Can a robot take up these jobs of a human…?

The technology invented here is to couple human voice with the mechanical arm. Speech is analyzed by software and converted to digital format. The hand writing is mathematically formulated as discussed below. Further, the system allows selecting upon different styles among the ones present in their memory. The rest of the paper is organized as follows: Section II presents a brief overview of
Fundamental Building Blocks

Our aim is to make the arm trace the letters in accordance with the speech. Arm movement can be brought up with pulleys and stepper motor. For this, each alphabet, number or a special character is to be represented by a unique values, so called as ‘fonts’, and fed into a memory. And motor carries out the activity in relation to those values. And we need a microcontroller to take up memory and control activities. We require a speech analyzer to convert human language into text and text is processed further into digital data. Microcontrollers work with digital inputs. So the speech has to be converted into digital string. To position and orient the robotic arm, manipulators of 6-DOF or 3-DOF can be employed. The DOF of the manipulator or the arm are distributed into sub assemblies of each axis.

Given below is the outline of technical blocks of the robotic humanoid hand.

![Figure 1: Basic Building Blocks-Overview](image)

Hardware Descriptions

Micro Controller Unit:

We use the basic 8051 family micro controller Atmel 89C51. It has 4KB on-chip program space, 128bytes RAM, 2 timers, 4 I/O ports, 1 serial port and 6 interrupt sources. 8051 is a 8-bit processor meaning that the CPU can work on only 8 bits at a time. Data larger than 8 bits has to be broken into 8-bit pieces to be processed by the CPU.
Micro controller unit used here performs the task of matching the incoming text with already stored binary equivalents of the letters and alphabets. MCU makes decision of which data has to be written to the port where motors are interfaced. We require 3 motors for 3 Degrees of Freedom.

Stepper motor interface:

Stepper motor is widely used device that translates electrical pulses into mechanical movement. In applications such as hard disk, dot matrix printers, and robotics the stepper motor is used for position control. The most common stepper motors have four stator windings that are paired with a center-tapped. This type of stepper motor is commonly referred to as a four phase or unipolar stepper motor. The center tap allows a change of current direction in each of two coils when a winding is grounded, thereby resulting in a polarity change of the stator. The stepper motor shaft moves in a fixed repeatable increment, which allows one to move it to a precise position. This repeatable fixed movement is possible as a result of basic magnetic theory where poles of same polarity repel and opposite poles attract. The step angle determines number of rotations per minute and the delay given for each step determines the speed of rotation. Due to incremental steps, accuracy is high.

Pulley Mechanism for Navigation:

Movement of the arm can be brought about by using a manipulator that is fixed to a pulley belt. Pulley is operated by a stepper motor. The setup of stepper motor controlling the pulley with its shaft is shown in the figure. For x,y and z axis movement we use three pulleys with motors in it. Motion of one arm affects the other there by bringing up a relative translation over the entire 3 dimensional space.

Figure 2: Stepper Controlled Pulley System
Construction

Module 1- Mechanical Perception:

3 DOF design:

To enhance robot hand dexterity, the robot should be designed to have a redundant number of degrees of freedom. In order to avoid ill poised nature and to minimize its inability to act as a human hand, we go for a 3 Degrees of freedom design. The overview of the design is shown in the figure. X,Y and Z are the Cartesian co-ordinates, \( q_1, q_2 \) and \( q_3 \) denotes vector of joint angles. \( L_1, L_2 \) and \( L_3 \) is the distance between two vector of joint angles. Movements in Z axis is used to lift the whole set up and place it in its origin for the start of a new page or a new line and movement in X and Y axis is for writing on the plane.

![Figure 3: 3 DOF Design](image)

Mechanical setup:

The entire built set up is shown in the figure. Pulley is supported by metal rods. As mentioned above we use three pulleys. Each of the pulley is driven by stepper motor, which gets signals from the micro controller for its movement. The strength of the pen on the paper is adjusted by varying the Z axis dimension. Also the speed of the stepper motor determines the accuracy of writing.

![Figure 4: Mechanical Setup](image)
Module 2- Programming Conception:

Speech to text:

The speech is converted to text using the third party software as mentioned above. The text data is converted to binary form and sent serially to MCU. Hence each word maps to certain set of binary values, i.e. set of letters in binary form. The pronunciations of words vary from person to person. Hence it is hard to differentiate between words having similar pronunciation. Bear can be interpreted as bare or vice versa. These types of verbal mistakes are much common with this type of systems and can be minimized by training the system continuously with most common words. If the software used is able to compare the words with the context of the sentence and then check for verbal errors, then it would be much efficient. As a key point, the text is converted to digital context and serially sent to the controller.

Text identification:

The values of each alphabet, character or number is already stored in heap of memory of the controller. This heap of values maps to set of locations which contains the values that are to be sent to the pen in order to bring about the letter. These values are nothing but the mathematical formulation of letters, i.e. the equations formed by those letters when endorsed in a 2 dimensional XY plane.

Mathematical formulation of text:

The entire text is imagined as combination of lines and curves in XY plane. Let the size of character be in the ratio of 1:2 (width : height) for English and 1:1 for Tamil. Hence each character is expressed in equations of x and y. For the word to be written, values of the entire axis have to be given simultaneously. For slanting lines, slope has to be specified and for curves its corresponding curvature values has to be specified. With this type of method, any character can be represented and be stored as specific fonts in the MCU memory. This allows multiple hand writings to be written by a single system.

Figure 5: Text Mapping To The Memory

Figure 6: Graphical Outlook of a Sample Letter
Conclusion

The algorithm for formulation of the text was derived. Speech analysis was found tricky with many errors. But it was minimized when the system was trained with the common words. Mechanical arrangement was designed and the errors were significant while retrace. Yet it proved to be efficient. If this technology is released as a real-time product it would really be a boon for physical ailment.

References

LaaLaLaa - A Tamil Lyric Analysis and Generation Framework

Sowmiya Dharmalingam & Madhan Karky
sow_86@yahoo.co.in | madhankarky@gmail.com
Department of Computer Science & Engineering
College of Engineering Guindy, Anna University

Abstract

Over 1000 Tamil lyrics are being written every year. Tamil films, advertisements, private pop albums are the main contributors of Tamil lyrics. This paper presents LaaLaLaa, a framework for Tamil lyric analysis and generation. For the analysis part, a set of 2000 Tamil lyrics collected over a period of 50 years penned over different themes is mined for various statistics and morphological patterns. The analysed statistical data is then used for the generation of the lyrics. This paper explains the different components in the framework such as lyric generator, rhyme finder, lyric miner, template selector, morphological generator and WordNets for lyric generation. We introduce three scoring mechanisms for the properties of flow, rhyme and meaning of a lyric. These scoring methods are used to compare the quality of the generated lyric. Discussing the results, the paper concludes with open problems and future directions.

1. Introduction

Tamil is a vibrant language with a rich grammar, vocabulary, an inherent poetic flavour and music is synonymous with its culture. Tamil Lyrics have evolved in a dramatic fashion from the many historical poems of the Sangam literature including Ettuthogai and Pathuppaattu, narrative poem Silappatikaram, compositions of the Tamil Saiva saints such as Appar, Thirugnana Sambanthar, Manikkavaasagar and compositions of Tamil hymns known as Thiruppugazh to the medieval period philosophies dominated by songs to the present day’s rap culture. There are about 1000 lyrics being written every year as private albums, jingles and as original soundtracks of mainstream movies.

In this paper, we propose LaaLaLaa, framework that can analyse lyrics for various statistics on words, their morphological patterns and generate lyrics in Tamil concordant to the input music and selected theme. Computational creativity is particularly very challenging, as it requires understanding and modelling knowledge, which almost cannot be formalized. Generating meaningful lyrics to a given tune and in a given domain can be treated as an optimisation problem that aims at maximizing the various features of lyrics such as meaning, rhyme and flow.

This paper is organised into four sections. The second section discusses a few contributions to lyric generation and how LaaLaLaa differs from the rest of the lyric generation systems. The third section presents the LaaLaLaa lyric analysis and generation framework and in detail discusses the various components of the system. The section also presents suggestions for three scoring mechanisms to
compare generated lyrics. The fourth section discusses results and concludes with open questions, ongoing research and future work.

2. Background

Several poetry generation systems have been developed in the past which broadly fall into the following categories - template based, generate and test approaches, evolutionary approaches and case-based reasoning approaches[6].

The Tra-la-Lyrics system[7] generates Portuguese lyrics given a MIDI by calculating syllabic division and syllabic stress identification of words and the strength of each beat. It however doesn’t handle the semantic aspect and produces completely random words and also doesn’t use metric patterns that can serve as an exact template for the words.

The Automatic Generation of Tamil Lyrics for Melodies[12] identifies the required syllable pattern for the lyric and passes this to a sentence generation module which generates meaningful phrases that match the pattern. A corpus of poems and stories is used as a source of phrases. This system is constrained by the size of the phrases/patterns and generates phrases that are independent of the previous phrases which leads to lyrics that are meaningful in parts, but meaningless on the whole. Secondly, the system presented in [12] generates rhyme based on maximum substring match and fails to make use of edhugai, moanai and iyaibu, the three rhyming patterns that are specific to Tamil language.

Other remarkable works in this domain include poetry generation in COLIBRI[3] and An expert system for the composition of formal Spanish poetry[5] which translate a user specified prose message into formal Spanish poetry, Hisar Manurung’s McGonnagall[10] where the poem generation process is formulated as a state space search problem where a state in the search space is a possible text with all its underlying representation, and a move can occur at any level of representation, from semantics to phonetics. The POEVOLVE system which implements the architecture proposed by Levy[9] creates texts that satisfy the form specifications of limericks. The WASP[4] system splits a given block of text into shorter fragments to identify reference patterns and uses the words in the text to produce verses that match these patterns.

In his thesis, Manurung[10], defines a poem to be a text that meets three properties: meaningfulness, grammaticality and poeticness. Our system currently handles the grammaticality and meaningfulness aspects to a certain degree while work is going on to enhance these factors, to meet the poeticness factor (LaaLaLaa framework proposed in this paper uses edhugai monai and iyaibu properties to achieve this) and handle other limitations of existing systems.

3. LaaLaLaa Framework

The LaaLaLaa framework presented in figure 1, can be divided into four major subsystems. Music-To-Template, Lyric Analysis, Lyric Generation and Lyric Scoring are the four subsystems that constitute the LaaLaLaa framework. The following sections describe the subsystems in detail.

3.1 Music To Template

MIDI (Musical Instrument Digital Interface) is a format for representing music in digital devices. A MIDI file is the input to the system. The MIDI file is processed by a MIDI-To-ABC[1] converter. The ABC notes[8] are processed by ABC-To-Template converter which transforms the ABC notes to Tamil Textual Place Holders, we call them Templates. A template can be formed using combinations of a short vowel(thā|na), a long vowel(thaa|naa) and a consonant(n). An example line template is provided below.
A template such as the one provided in the example can be split into numerous combinations of smaller chunks. The objective of splitting the line template into smaller chunks is to treat the chunks as placeholders for fitting words. We solve the problem of minimising the number of templates based on results from the analysis subsystem. A Lyric Stats DB provides information on average word lengths and average number of words per sentence for a template of given length. This statistic is used to split the given template into smaller chunks of placeholders for words.

The line template example provided above may be split into one of the following options based on statistics available in Lyric Stats DB for line of length 9.

Template Selector, selects templates that can be matched with the previous line to maximise rhyme score, which will be explained in the Lyric Scoring in section 3.4. The selected template will then be tagged for patterns. Pattern of a line will the part-of-speech tags associated with each word of the line. This POS tag for each word of the template is obtained from Pattern DB, populated as a result of Pattern Mining in the Lyric Analysis system. Pattern DB & Stats DB are explained in next section. The following tagged template is example of a selected pattern.
It is also to be noted that a single template may match multiple patterns and a pattern is selected at random. This pattern tagged template forms the input to the Lyric Generation subsystem.

3.2 Lyric Analysis

Lyric Analysis is an offline system where a Tamil lyric corpus is analysed for various statistics and patterns. Lyrics collected from various sources are formatted by tagging them with appropriate headers and section tags (pallavi, anupallavi, saranam...) The formatted lyrics are fed as input to a statistics miner and a pattern miner. A statistics miner analyses lyrics for statistics corresponding to length of sentences and words, length of co-occurring words.

<table>
<thead>
<tr>
<th>n=Length(s)</th>
<th>AvgWPS(n)</th>
<th>AvgWLN(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>40</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>45</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Lyric Statistics

Table 1 provides some statistics obtained from our implementation of lyric miner over 2000 Tamil lyrics collected from multiple sources. The Length(s) denotes the character count for a given sentence. AvgWPS(n) denotes the average words per sentence for sentences with length n. AvgWLN(n) denotes the average character count of words for sentences with length n. It is to be noted that the average is computed over the lyric corpus. The template processor uses this information to efficiently split templates for word fitting.

| wps | | pattern |
|-----|----------------|
| 2   | 4 < Noun > < Pronoun + Sandhi + Clitic > |
| 3   | 10 < Interrogative Noun + Clitic >< Pronoun >< Interrogative Noun + Clitic > |
| 4   | 3 < Infinitive >< Verb + Past Tense + Verbal Participle>< Time >< Noun > |

Table 2: Lyric Pattern

A sample output from our pattern miner is provided in Table 2. wps denote the words per sentence and | pattern | denotes the number of times a particular pattern occurs. The pattern gives information on what
pos and morphological suffixes are used [2]. Pattern Selector module uses these results to tag templates for lyric generation.

### 3.3 Lyric Generation

The Lyric Generation Subsystem uses domain specific WordNets, Morphological Generator and a Rhyme Finder. The domain specific WordNet consists of words and their associations specific to a certain domain such as nature, love, history, geography, religion, zoology and more. Given a root word and appropriate tags, it can generate nouns and verbs. The information obtained from the Pattern Tagger is used to generate words according to the tagged pattern. The Rhyme Finder is used by the Lyric Generator to choose words that match one or more of the three Rhyme properties (edhukai, moanai, iyaibu). The edhugai property states that the second letter of two parallel words match. Two words are considered to be parallel if they occur in the same position of two different sentences. The moanai and iyaibu properties state the same for first letter and last letter to match respectively.

Let the above example be considered as two template patterns for parallel sentences sent to Lyric Generator.

The above given lyrics are two options generated by the lyric generator. The words poovae and theevae are not obtained from the WordNet as the WordNet comprises of root words only. The morphological generator uses the noun+clitic pattern for naanaa word template and chooses noun roots from WordNet to generate morphological suffixes that meter-match the template.

Two words are considered to Meter match if the two words have same character length and every character of a word meter matches with the corresponding character of the other word. The meter matching of two words is again obtained from ancient Tamil grammar definition of maathirai aLavu(meter length). Rhyme finder minimises the search space for generating the second line by restricting the list of word to match edhugai monai or iyaibu properties. Choosing same length roots with same clitic or matching the first second or last characters of words enables choosing rhyming words.

### 3.4 Lyric Scoring

We propose a lyric scoring subsystem that takes an entire lyric as input and gives three different scores for the lyric namely flow score\(f\), rhyme score\(r\) and Meaning Score\(m\). Flow of a lyric can be modelled as a property influenced by the phonetic properties of words in the lyrics. Tamil alphabets are sensibly classified into short vowels, long vowels, soft consonants, hard consonants and mid consonants. We use this classification along with doublings in words to determine the score of a word, a sentence and thus a lyric. Rhyme score \(r\) analyses various rhyming patterns and rhyming properties mentioned in the
previous section determine a score for a lyric. Meaning score (m) is the challenge. We do not have any formal method to analyse if a given sentence is meaningful. If meaning can be a property that can be achieved through associations of words in a sentence and connection between sentences, then domain WordNets can be used to determine a meaning score for a lyric.

4. Conclusion and Future Work

The objective of this paper is to present the LaaLaLaa framework for lyric analysis and generation. Very few parts of the LaaLaLaa framework have been implemented. The word generator has been implemented in parts and a very small WordNet corresponding to nature is used for testing. A basic statistics and pattern miner modules have been implemented. Developing multi domain WordNets and implementing the meaning score will be a challenging task. Improving the speed of generation by building a special index based on meter length of words will be part of our future work.

Acknowledgements

We would like to thank Dr. T. V. Geetha and Dr. Ranjani Parthasarathi for providing valuable inputs to this paper. We also would like to acknowledge the contributions of Gayathri Lakshman, Prathyusha Senthil Kumar towards initial implementations of the LaaLaLaa framework.

References

Language Independent Emotion Recognition System
for Web Articles Using NLP Techniques

D. Mahendran,
PG Scholar | dmahendranme@gmail.com

S. Gunasundari,
Senior Lecturer | gunaannauniv@gmail.com

Prof. B. Rajalakshmi,
HOD | velammalcseraji@yahoo.co.in

Department of CSE, Velammal Engineering College, Chennai, India.

Abstract
In recent years, suicide of college students has been a universal phenomenon in the world. And the phenomenon has become more and more severe because of the complex and drastic competitions. With the popularization of internet and the development of information processing technologies, a lot of people have established their own blog websites to write down their experiences and express their feelings at times. It will be very helpful if the computer can recognize the emotions expressed in blog pages automatically. And then it will be convenient for teachers or psychological consultants to monitor the affective information of college students and take measures for the depression prevention when necessary.

Owing to the advances in Affective Computing and Natural Language Processing, researches have begun to pay more attention to the emotion recognition in NLP all over the world. This leads to construct an Emotion Recognition System. It is based on the lexical contents of words and structural characteristics of blog articles. The emotion recognition system is composed of five modules: data collection, data processing, morphological analyzing, emotion tagging, and emotion computing. This algorithm independently works for web articles in any language. We checked articles in English and Tamil and the response of the system is good.

Keywords: Suicide, Depression, Natural language processing, Emotion classification, Emotion recognition

Introduction
In recent years, suicide of college students has been a universal phenomenon all over the world. Furthermore, the number of college students who suicide themselves has been increasing with amazing speed in a worldwide scope, around 100 million persons every year. In the year of 2003, World Health Organization (WHO) appointed September 10th as “World Suicide Prevention Day (WSPD)” to give a caution to the society. And it has adopted many suicide prevention programs and projects to avoid unnecessary death. At the same time, the problem of mental health among college students has raised great attention from more and more psychologists, educators and people in all the other professions.
In order to intervene in the crisis, psychological consultation centers have been set up almost in every university or college. However, some college students are afraid to talk about their private experiences with other people face to face, and refuse to consult psychological consultants when they meet troubles.

In late decades, the popularization of internet and the development of information processing technologies have greatly changed the communication ways of mankind. More and more people have established their own blog websites over the internet where they write down their experiences, put forward their opinions, and express their feelings at times. While some people feel embarrassed to confide their troubles to others face to face, they may feel free to express their emotions in blogs without any pressures.

It can be anticipated that in the near future, almost everyone in the world especially every college student will possess his or her own blog website. At that time, it will be very helpful if the computer can recognize the emotions expressed in blog articles automatically.

In this case, an Emotion Recognition System is needed. If the system detects that some college student has been in a blue mood for continuous days (for example, for continuous three days), it is recommended that the teacher should pay more attention lately and have a talk with him or her regularly so that the depression can be treated in time. That’s why this research has been taken on.

In this paper, we outline a new approach to recognize the author’s emotion from his or her blog articles. Based on the approach, an Emotion Recognition System has been constructed. The results of our experiments on blogs prove the feasibility of the means.

**System Structure**

The emotion recognition system is composed of six modules: data collection, data preprocessing, dictionary creation, morphological analysis, emotion tagging, and emotion computing shown in figure 1.

In the following we will introduce the flow of how the system works step by step. Firstly, the blog articles are collected from various authors with their emotion category mentioned. The collected articles are preprocessed to remove unwanted data from the blog articles. Then the preprocessed articles are subjected to morphological analysis. Here in morphological analysis the articles are split into sentences then the sentences are split into words for comparing with the emotion dictionary to find the emotion category. In emotion tagging, emotion category for each words and sentences are assigned.

Finally, with the emotion category of each sentence, the Emotion Computing module will compute the weight value of emotion categories of all sentences in a blog article according to the blog structure rules, and output the emotion result for the whole article.

**Data Collection**

Our emotion recognition system will work effectively on group of sentences. That’s the reason we go for blog articles where people express their emotions in blogs without any pressures. In our country the blogs are not that much famous. This is the issue we faced on collecting the data. The personal blogs are popular in US, China, Japan which is not that much popular in our country. What we did is, we created a proxy to access the US websites for English blogs. But in Tamil, we got only few web articles.
Around **300 English blog articles** were collected from the internet among various authors.

**Data Preprocessing**

The collected blog articles contain images, videos etc. Raw data cannot be used directly in our emotion recognition system. So that, the data has to be preprocessed. Here in data preprocessing the unwanted elements are removed from the articles like HTML tags, images, videos and other multimedia files are removed. HTML Parser is used for data preprocessing. The HTML Parser module analyzes the blog webpage that has been downloaded, and extracts each blog article in the webpage. Then it divides the blog article into paragraphs and sentences which are stored for further processing. Delimiters used for separating the sentences are .(dot), ?(question mark) and !(exclamatory mark). Other special characters like ~, @,#,$ etc are removed from the articles. And for Tamil articles, the format is changed to Unicode for easier implementation.

**Morphological Analysis**

Morphology is the identification, analysis and description of the structure of words (words as units in the lexicon are the subject matter of lexicology). While words are generally accepted as being the smallest units of syntax, it is clear that in most languages, words can be related to other words by rules. Here in our system we have created the own morphological analyzer. It contains Heuristic rules, English and Tamil dictionaries (noun dictionary, verb dictionary, adverb dictionary). User interface of one of the dictionary is shown in figure 2. Binary search tree algorithm is a node based binary tree data structure which we have used in these dictionaries. Therefore the searching efficiency will be more. The heuristic

---

**Figure 1: Framework for the emotion recognition system**

![Figure 1](image-url)
rules are rules to extract the root words from the sentence. For example: The word ‘happily’ is an adverb. Root word for ‘happily’ is ‘happy’. Likewise ‘anju’ is a root word for Tamil word ‘anjinen’.

For each natural language sentence, the Morphological analyzing module will analyze the lexical characteristics of the sentence and extract words from it based on the rule database according to analyzer rules. The words are used for further processing. Morphological analysis part has been made of automatic.

![Figure 2: User interface of the Noun dictionary](image)

**Emotion Dictionary**

Emotion dictionary is constructed which is to be used in emotion tagging module. In emotion dictionary, the data structure, Binary search tree algorithm is used to improve the searching efficiency. Emotion dictionary is made as semi-automatic. User interface of the emotion dictionary is shown in figure 3. Here two separate emotion dictionaries are used for English and Tamil languages. Numerous words have been inserted in the emotion dictionary with their appropriate category. It contains three fields, word, category and weightage. The weightage field stores the weight values for each word. Weight values of each word play a vital role in emotion recognition system. Emotion category is found out based on weight values.

In the past, emotion has been simply divided into two categories “pleasure” and “displeasure” which are too ambiguous to assess rich emotions of human. Ekman has defined 7 universal affective categories [2] based on unique facial expressions which seem still less while applied into a practical system. In contemporary Chinese, 39 emotional categories are specified for vocabulary [3], nevertheless part of which is seldom used in daily communications.

*Table 1: Emotion vocabularies for Tamil*

<table>
<thead>
<tr>
<th>Emotion category</th>
<th>Happy</th>
<th>Angry</th>
<th>Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>vocabulary</td>
<td>Santhosham</td>
<td>Sinam</td>
<td>Thukkam</td>
</tr>
<tr>
<td></td>
<td>Subam</td>
<td>Veruppu</td>
<td>Thuyaram</td>
</tr>
<tr>
<td></td>
<td>Sugam</td>
<td>Padhatram</td>
<td>Varutham</td>
</tr>
<tr>
<td></td>
<td>...............</td>
<td>...............</td>
<td>...............</td>
</tr>
</tbody>
</table>

In our Emotion Recognition System, we classified emotion based on the affective categories defined by Ekman plus some important categories found in blogs with high frequency of use. They are totally 26 categories: happy, sad, fearful, disgusted, angry, surprised, love, expectant, nervous, regretful, praiseful, shy, respectful, proud, impatient, doubtful, hateful, grievance, critical, depressed, exited, thankful, annoyed, scornful, haughty, envious. When there is no emotion expressed in a blog, we named the mental state as “neutral”. Some examples of the emotional vocabularies for Tamil are listed in Table 1.
**Emotion Tagging**

In morphological analysis the article is split into sentences and words. Those extracted words are passed to emotion tagging module. The words are compared with the emotion dictionary. If the corresponding word is present in the emotion dictionary then the weight values and emotion category is assigned to that word.

![Figure 3: User interface of the emotion dictionary](image)

**Emotion Computing**

Finally, with the emotion category of each sentence, the Emotion Computing module will compute the weight value of emotion categories of all sentences in a blog article according to the blog structure rules, and output the emotion result for the whole article.

Let us suppose that there are \(j (j=1,2,3,...,m)\) sentences in an article. For each sentence \(j\), there is a emotion weight value \(W_j\) assigned. \(E_a\) is one of the 26 emotion categories \((a=1,2,...,26)\). \(E_a\) with maximum sum of all the corresponding \(W_j\) is the emotion of the whole article.

**Experimental Results**

Emotion recognition system is implemented in Java language. And we have attained the expected results. It is still an issue for the evaluation. By now there is not a good evaluation method generally accepted. Since different people may have different opinions even on the same text, it is common for them to give different evaluations of emotion manually. In our experiments, we carried close test based on the manual evaluation of emotions as a standard set which is judged by evaluators in advance.

**Close Test:** The close test is carried out on several blog articles. Our algorithm of emotion computing is realized in our system. And the evaluation is to count the number of correct prediction by the system compared with the standard set. We checked articles in English and Tamil and the response of the system is good.

**Conclusion and Future Work**

Recently, the number of college students who suicide themselves has been increasing with amazing speed in the world. The problem of mental health among college students has raised greater attention from a lot of people in many professions. With the popularization of internet and the development of information processing technologies, more and more people have established their own blog websites. It will be convenient for teachers or psychological consultants to monitor the affective information expressed in blog articles in order to prevent the depression of students. The advances in Affective
Computing and Natural Language Processing enable the emotion recognition from blog articles automatically.

In this paper, we have outlined the approach to develop an Emotion Recognition System. Firstly we decided the classification of emotions, and introduced the model of the System. Based on the lexical contents of words and structural characteristics of web articles, a method has been proposed for emotion computing. The experiments for testing have been carried out. The approach was proved to be feasible and pointed out the future's direction of the research. We find that some parts need to be improved in the future: The first task to improve the emotion dictionary by expanding the vocabularies and arranging them better, for it plays a very important role in the emotion sensing. Since different people have different styles in writing blog articles, in the future we will choose more web articles of different authors for the algorithm analysis and system test, which are supposed to achieve a better performance. Currently we are developing a system to find emotions from Hindi language blogs. We are in research of developing a language independent emotion recognition system for all languages. That is, identifying an emotion from all language blog articles by the help of Unicode.

References
An Integrated Intelligent Framework for automatic story generation in Tamil

G.V. Uma
Assistant professor
Dept of computer science Engineering
Anna University Chennai, Chennai.
gvuma@annauniv.edu

Abstract
Story is a description of a chain of events told or written in prose or verse. It is an interesting way to transfer knowledge from one person to other in the form of narrated sequence of events. The sequence of events is arranged in a chronological order to convey the message to other. An intelligent framework is developed using ontology to benefit the system in order to generate the story automatically and to reason the system generated-stories based on their conceptual consistency and validity. Ontology has the significant role in constructing semantic stories by the system. Ontology is a formal explicit shared conceptualization, which contains domain knowledge for the story construction. This research work developed a framework for automatic story generation in Tamil.

Introduction
Everyone in the world have their own interest to read the stories and like them very much. Stories are naturally has their own way of attraction from children to old age people. Children learn their moral and social obligations in the form of stories narrated to them by their guardians and peers [1]. The basic characteristics of human beings can be explained through stories to the youngsters to inspire them. Hence, they play a vital role in everyone’s life and their traits can be shaped well, based on the characters in the story. For example, Mahatma Gandhiji was influenced by the story “Harichandra” for speaking the truth in any kind of difficulties, and the “shravana story” for obedience to his parents. People can gather a lot of information from the story based on their perceptions.

The ASG system tries to make the computer and Artificial author’ to construct new stories dynamically. The story is a natural verbal description of objects / human beings, their attributes, relationships, beliefs, behaviors, motivations and interaction. It is a message that tells the particulars of an act, and an occurrence or course of events are presented in the form of writing. It can also be described as follows:

➢ It is a description of the sequence of events
➢ It is a piece of fiction that narrates a chain of related events
➢ It focuses on only one incident, has a single plot, a single setting, a limited small number of characters, and covers a short period of time.

The automatic story generation system helps to generate a variety of new stories dynamically as per the reader’s interest. The user has the choice to select any characters, locations, settings for generating a story. The theme may be conceived either automatically or by the user; similarly, existing stories can be revised to remake new ones. Any kind of story can be restructured and reasoned based on the author’s
requirement and specification. Ontology provides full support for story generation by preserving the concepts, attributes and the relations among them, which help for semantic reasoning. The language generator helps to generate suitable sentences with beautiful words for the construction of new stories.

This story generation framework enables the user, to act as reader as well as the author of the story. They can generate stories based on their wish and with necessary ingredients like characters, settings, location, food, etc. A simple story is started with initial situation and which has an active element to precede the story interestingly and finally comes to final situation. The generated stories are helpful for kids to know about the world knowledge.

Story generation framework comprises of three levels of computation such as theme conception, story generation and semantic reasoning. Ontology plays a major role in each of these three phases. Ontology is a formal explicit specification of shared conceptualization [2]. It can be expressed as the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary for a specific domain. The ontology is constructed based on the components of story domain. It involves characters, location settings like forest, palace, home and etc. The ontology posses the events and the order of events/ functions /activities leads the story generation system. Even though constructed ontology is domain specific, it can be updated, modified, reengineered, reused for other purposes also. The main reasons [3] for using ontology for story generation system is

- To share common understanding of the structure of information among people or Software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyze domain knowledge

The structure of the paper organized as follows: Section 2 discusses the related work and Section 3 provides the framework for automatic story generation. Section 4 devotes in the role of ontology in automatic story generation. Section 5 focuses on the role of ontology in story generation and reasoning. Section 6 discusses the conclusion and future works.

**Related works**

There are different types of story generators available for the purpose of automatic story generation and their evolution is described below.

Propp [4] discussed the story generation as; a tale is a whole that may be composed of thirty-one moves. A move is a type of development proceeding from villainy or a lack, through intermediary functions to marriage, or to other functions employed as a denouement (ending). One tale may be composed of several moves that are related between them. One move may directly follow another, but they may also interweave; a development, which has begun pauses and a new move, is inserted. Bailey [5] described an approach towards automatic story generation based on the twin assumptions that it is possible for the generation of a story to be driven by modeling of the responses to the story of an imagined target reader, and that doing so allows the essence of what makes a story work (its ‘storiness’) to be encapsulated in a simple and general way.
Charles, F et al [6], presented results from a first version of a fully implemented storytelling prototype, which illustrates the generation of variants of a generic storyline. These variants result from the interaction of autonomous characters with one another, with environment resources or from user intervention.

Dimitrios N. Konstantinou et al [7] discussed the story generation model HOMER. It receives natural language input in the form of a sentence or an icon corresponding to a scene from a story and it generates a text-only narrative apart from a story line and it includes a plot, characters, settings, the user’s stylistic preferences and also their point-of-view.

Riedl et al [8] had provided planning algorithm for story generation. The story planners are limited by the fact that they can only operate on the story world provided, which impacts the ability of the planner to find a solution story plan and the quality and structure of the story plan if one is found, but which lacks semantics.

George miller [9] provided a wonderful environment to have the collection of words and their synonyms and they are put together to form a lexical database. It helps to retrieve the meaning for any kind of word in any language.

This framework helps to generate stories automatically and to overcome the semantic lacking of stories which can be resolved by reasoning the stories in a systematic and efficient way using the efficiencies of ontology.

**Framework Automatic Story Generation**

This framework for automatic story generation shown in Figure 1 and it helps to provide a systematic approach to generate and analyze stories both syntactically and systematically. This framework mainly divided into five phases such as Theme conception, Sentence generator, Parser, Syntactic checker and Semantic checker.

**Theme Conception** -

There are two main approaches for conceiving theme such as:

- **Static conception** - is a predefined order of events to describe the flow of a story.
- **Dynamic conception** - refers to the order of events which are selected and organized during the phase of story generation.

Set of built-in themes are available in the theme repository; the user can select the necessary theme for the story based on the characters involved. Also, a new theme can be conceived using events from the repository.

![Framework for automatic story generation](image-url)

*Figure 1 – Framework for automatic story generation*
Figure 2 – Detailed Framework for Sentence Generator

Figure 2 depicts the detailed framework for generating simple sentences. The inputs of the Sentence generator are sentence type, sentence grammar and necessary terminal values. Using the sentence type, the corresponding sentence grammars and their production rules are retrieved from ontology. The necessary terminal values are checked with given production rules. There is no standard sentence structure for Tamil. The following grammar rules were framed and based on the rule sentence structures is obtained [11] [12].

1. NC --- > adj N / N / ADJC N / NNC
2. VC --- > adv V / adv rpl / ADVC V / vpl / V
3. NNC --- > S con
4. ADJC --- > NC VC
5. ADVC --- > (NC)* vpl
6. S --- > (NC)* (VC)

Figure 3 – sample grammar for sentence generation

Figure 3 depicts the sample grammar for sentence generation in Tamil which is used for stories. By applying if then – rules, suffixes are updated with root words. Parser contains the two phases namely separator and analyzer. Separator helps to separate the story into story segments called sentences. The separation of sentences helps to check the sentences by its form of sentence structure and the meaning of the story. Analyzer used to identify the noun, verb, settings, location and etc from the story which helps in semantic validation.
Figure 4 depicts the detailed description about the semantic reasoner which divides into two such as syntactic validation framework and semantic validation framework. Syntactic validation framework helps to check the sentence structure of the sentences based on the sentence grammar whereas semantic checker detects the meaning of sentences whether it is valid or invalid.

Role of ontology in ASG

Nowadays ontologies have significant role in information processing. They are important for story generation, which holds the various concepts that are relevant to the story domain. Initially, the ontology is built with minimal knowledge, in later stages; ontology can be extended whenever new concepts are introduced. One of the primary purposes of constructing ontology is to provide a standard, unambiguous representation of a particular domain of knowledge.

OWL has the language expressive representative formalism and reasoning power. Because OWL was derived from DAML+OIL, it can take advantage of the existing reasoning algorithms in Description Logics (DL). The semantic of OWL allows us to define a ranking function that distinguishes multiple degrees of matching. There are three types of matches: exact match where the concept to be found is found, plug-in match where the concept to be found is more specific than the concept in ontology, and subsume match where the concept to be found is more general than the concept in ontology. The scoring function of matching degree is given below [10]:

- Exact Match > Plug in Match > Subsume Match

This matching degree helps to identify the semanticness of the generated story. The basic properties of the Lion is

- Living being → animals → wild → Lion (king, legs, anger, roar, kill)
- Domestic → Rat (small, legs, frightens)
- Bird → crow (black, legs, fly)

If the generated sentence states that ‘(Lion flew) means the semantic checker detects that ‘fly’ action cannot be performed by animal. So the basic properties of Lion are preserved. It helps to improve the quality of the story. Consider another sentence, ‘(in this sentence, the domestic animal ‘mouse’ killed the wild animal ‘Lion’. In reality, it is not possible and then the system detects it, based on the strength of the animals which are categorized in ontology. The semantic reasoner detects and corrects the sentence as, So, the basic properties of Lion and rat are preserved. The sample story generated by the system is given in the Figure 5.

Figure 5 – Sample story generated by the system.
Results

The above framework has been implemented in Java and uses the Java JDBC to connect to MySQL server and retrieve the relevant values. GUI is developed using Java Swing. The necessary basic concepts and their attributes are retrieved for semantic checking. The framework has been built and tested for various stories. The system parses the sentence to the framework for the reasoning purpose. Charles [6] proposed a set of factors that are considered to check the quality of the story. These factors are utilized to check the quality of stories by reasoning. Based on the above factors, the generated stories are given to a group of people to give their opinions about the generated stories with following scaling factors and with above said features. Excellent – 5; V.Good – 4; Good -3; Fair – 2; needs improvement – 1;

**Table1: Factors for Assessment of Story**

<table>
<thead>
<tr>
<th>S.no</th>
<th>Factor</th>
<th>Describes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Overall</td>
<td>How is the story as an archetypal fairy tale?</td>
</tr>
<tr>
<td>2</td>
<td>Style</td>
<td>Did the author use an appropriate writing style?</td>
</tr>
<tr>
<td>3</td>
<td>Grammaticality</td>
<td>How would you rate the syntactic quality?</td>
</tr>
<tr>
<td>4</td>
<td>Flow</td>
<td>Did the sentences flow from one to the next?</td>
</tr>
<tr>
<td>5</td>
<td>Diction</td>
<td>How appropriate were the author’s word choices?</td>
</tr>
<tr>
<td>6</td>
<td>Readability</td>
<td>How hard was it to read the prose?</td>
</tr>
<tr>
<td>7</td>
<td>Logicality</td>
<td>Did the story seem out of order?</td>
</tr>
<tr>
<td>8</td>
<td>Believability</td>
<td>Did the story’s characters behave as you would expect?</td>
</tr>
</tbody>
</table>

**Table2: Results of Story Assessment**

<table>
<thead>
<tr>
<th>S.no</th>
<th>Parameters</th>
<th>Before reasoning (max = 5)</th>
<th>After reasoning (max = 5)</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Overall</td>
<td>4.3</td>
<td>4.7</td>
<td>1.09</td>
</tr>
<tr>
<td>2</td>
<td>Style</td>
<td>4.0</td>
<td>4.5</td>
<td>1.13</td>
</tr>
<tr>
<td>3</td>
<td>Grammaticality</td>
<td>4.1</td>
<td>4.8</td>
<td>1.17</td>
</tr>
<tr>
<td>4</td>
<td>Flow</td>
<td>4.2</td>
<td>4.5</td>
<td>1.07</td>
</tr>
<tr>
<td>5</td>
<td>Diction</td>
<td>2.8</td>
<td>3.4</td>
<td>1.21</td>
</tr>
<tr>
<td>6</td>
<td>Readability</td>
<td>3.7</td>
<td>4.0</td>
<td>1.08</td>
</tr>
<tr>
<td>7</td>
<td>Logicality</td>
<td>3.6</td>
<td>4.1</td>
<td>1.14</td>
</tr>
<tr>
<td>8</td>
<td>Believability</td>
<td>4.4</td>
<td>4.7</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>3.885</td>
<td>4.33</td>
<td>1.12</td>
</tr>
</tbody>
</table>
The Table-1 shows the assessment value of story. The Table - 2 depicts the calculated values for the stories before reasoning and after reasoning. After reasoning, the quality of the story is improved on an average of 22.4 percentage better than the before reasoning. The believability factor and overall content have very good feedback among other factors. The other factors like style of the story, grammar content in the story, flow of the story are the good factors in the next level and also the other good factors in the next level are diction, readability, and legibility. Figure 6 depicts results of story assessment of stories.

![Figure 6 - Results of Story Assessment](image)

Conclusion

Thus the framework helps to generate a simple short story in Tamil with simple sentences and semantic reasoning. It also proves that the framework is efficient, by comparing the results of the generated stories with before and after reasoning. In future, the project work can be extended to generate medium size stories, novels and picture based stories. Also, ontology can be extended with more number of concepts, attributes and their relation. Similarly, Reasoning can be extended to complex type sentences too.

References


Evaluation of Tamil Descriptive Passages using Concept Maps

Mahalakshmi G.S
mahalakshmi@cs.annauniv.edu
Sendhilkumar S
thamaraikumar@cs.annauniv.edu
Shankar B
shanloyola@yahoo.co.in
Department of Computer Science and Engineering,
Anna University, Chennai 600025, India

Abstract

E-learning has changed the way of education. With the wide spread applications of E-Learning technologies to education at all levels, we need to focus on learner-centric knowledge management in order to complement the conventional learning system. This paper examines the pioneering application of concept mapping as a follow-up study strategy for learning from text. With the great advancement in language processing tools in Tamil, we have proposed a novel mechanism which will evaluate Tamil descriptive answer passages based on the pre-stated prose. The idea is achieved by generating concept maps equally from prose and descriptive answer passages, and thereby evaluating the contents via generation of concept maps.

1. Introduction

With the wide spread applications of information technology to education at all levels, we need to focus on learner-centric knowledge management in order to complement the conventional learning system. To enhance self-learning and assessment especially in learning subjects that require high knowledge retention, application of concept maps to promote education has been a proven success worldwide.

1.1 Academic Significance

The first difficulty someone who attempts to comprehend a text faces is to understand what it is all about. That is, to grasp the global sense of the communication, understand its elements and the relationships among them. In this context, the student may understand some of the concepts involved in the definition. These concepts are linked by words forming whole sentences that seem to make sense. However, trying to understand the overall conceptual structure is more difficult. It is probably easier for many students to grasp a whole sense of the concept frame of reference when faced with a graph like the structure. This is due to the powerful visual effect that a graph has in order to facilitate understanding of a concept or a conceptual structure.

1.2 Contribution to Knowledge

Concept Maps help to improve understanding of a given subject and facilitate building student's own knowledge, as long as the student has the opportunity to use, criticize, analyze, question or improve expert's maps or Concept Maps generated by his own peers.
The implementation of concept maps in the classroom allows both the teacher and the student discovering and describing meaningful relations among the concepts object matter of the study, making it possible to create connections between them and the context in which activities are developed.

The concept map helps the learners to have a better overview of the course and what aspect he/she should pay attention. Concept maps constructed are very useful for teachers as an evidence of the way as each one of the parties involved in the process assumes his/her own learning. From their follow-up and analysis, experiences can be designed to help their students overcome weaknesses or to reinforce strengths acquired in learning process.

This motivated us to apply technological advancement and research in contributing better methodologies for education with a special focus on students whose medium of instruction is any regional language (for our study: Tamil) other than English. Although concept maps have been proven to be a successful resource for grade improvement in abroad, their use is little explored in Indian education. The detailed literature analysis conducted for the same revealed the fact that almost no work reported the development and use of concept maps to promote education in regional language – Tamil. In this context this paper concentrates on development of concept maps that eliminates the need for memorization and helps the students with active participation, to learn the subjects in their respective regional languages.

With the objective of applying ICT in Regional Language Education, in this paper we express the methodology of developing concept maps for various subjects at the higher secondary/pre-collegiate level.

2. Related Work

As it is known, an essential aspect in the learning process (either electronic or traditional) is the possibilities to evaluate the students. It is very important both for professor and student to test the understanding degree of the course. One of the best possibilities is to ask questions from the studied course. It is tested this way the degree of understanding of each studied material and the integration of new knowledge with the previous ones (that should already be known). These facts will have as a result an in-depth understanding of the learning materials.

Here we discuss the study and research done in connection with the proposed topic, by various experts outside India.

2.1 Question Generation for Learning Evaluation

Taking into consideration the high number of learning material existing in electronic format, the importance of the testing and evaluation systems has increased. The authors [McGough et. Al., 2008] present an interesting solution to the problem of presenting students with dynamically generated browser-based exams with significant engineering mathematics content. Here, the main idea is to generate the questions automatically based on question templates which are created by training on many medical articles. Liana et al [Liana Stanescu et. Al., 2008] tried to design and implement a software instrument (Test Creator) that permits generation of questions based on electronic materials that students have. The solution implies teachers to have a series of tags and templates that they have to manage. These tags can be used to generate questions automatically.
2.2 Concept Maps applied for Question Generation

Concept maps are a result of Novak and Gowin’s (1984) research into human learning and knowledge construction. One of the powerful uses of concept maps is not only as a learning tool but also as an evaluation tool, thus encouraging students to use meaningful-mode learning patterns. Concept mapping may be used as a tool for understanding, collaborating, validating, and integrating curriculum content that is designed to develop specific competencies. Concept mapping, a tool originally developed to facilitate student learning by organizing key and supporting concepts into visual frameworks, can also facilitate communication among faculty and administrators about curricular structures, complex cognitive frameworks, and competency-based learning outcomes.

However, the only issue is that the learner must choose to learn meaningfully. The one condition over which the teacher or mentor has only indirect control is the motivation of students to choose to learn by attempting to incorporate new meanings into their prior knowledge, rather than simply memorizing concept definitions or propositional statements or computational procedures. The indirect control over this choice is primarily in instructional strategies used and the evaluation strategies used. Instructional strategies that emphasize relating new knowledge to the learner’s existing knowledge foster meaningful learning. Evaluation strategies that encourage learners to relate ideas they possess with new ideas also encourage meaningful learning.

2.3 Concept Maps in E-learning

Recent researches have demonstrated the importance of concept map and its versatile applications especially in e-Learning. Concept maps creation for emerging new domains such as e-Learning is even more challenging due to its ongoing development nature. Concept maps can provide a useful reference for researchers, who are new to the e-Learning field, to study related issues, for teachers to design adaptive learning materials, and for learners to understand the whole picture of e-Learning domain knowledge.

2.4 Concept Map Mining

There is yet another approach [Villalon and Calvo, 2009] for automatic concept extraction, using grammatical parsers and Latent Semantic Analysis. Essays, as any other text, represent both the knowledge and the writing skills of its author; hence, an Automatic Concept Map from Essay (ACME) should reflect both. Therefore, the words for the concepts and relations must be extracted literally from the document, and the hierarchy of concepts must reflect the importance of the concepts relative to what was written in the particular document. However, the performance is related to the way concepts are chosen by humans. We believe that understanding this phenomenon and using it for the automatic selection of concepts could lead to big improvements.

3. Evaluation Mechanism

The proposed answer evaluation system creates concept-maps from the prose and the answer passages, that involves the extraction of concept-words from the passages by parsing the input text [Saravanan et al, 2004], creating a dependency model among the concept-words within the passage, visualization of concept-maps that has two levels viz., words level and the sentence pattern level that helps us to analyze what are the key concepts needed in a passage, Comparison of the concept-maps created and finally evaluation of the answer passages. The evaluation of the passage does not handle the content comparison part alone but also analyzes the finer aspects like the structuring
of the passage, repetitions, duplications, and also the additional aspects used by the students like causation, summarization, expansion at instance, etc. These finer aspects have been inspired from ‘nannool’[12], a Tamil prose that presents the needed and irrelevant aspects of a prose.

For drawing light on the evaluation mechanisms followed, we choose a specific prose and answer and evaluate them. The total number of words in the prose above which was considered for the evaluation is 15. The comparison procedure followed is presented below.

3.1. Missing Concepts

Comparing the prose and answer we find that the third sentence in prose is completely missing in answer. These three words will subtract 3 from 15. The last sentence in answer is not present in prose that’s an irrelevant concept and we don’t add any marks for it.

Marks=15-4 = 11

(11/15)*100 = 73.33% (word_match)

3.2. Duplicates

In answer the first two sentences are one and the same. We detect this with the help of the relation table. There are 59 relations generated and only 28 matches. So this module score is (28/59) * 100 = 47.45% (pattern).

3.3. Repetition

Sentence one is repeated again, so we deduct one mark. Therefore one mark should be deducted from the total score.

3.4. Value Addition

The evaluation method, value addition is computed based on causation and summarization, expansion at instance, argument by example. Here, the answer passage has summarization. Therefore one mark is to the total marks.

3.5. Score

From the values calculated above, the final score can be determined. Here, the given answer passage was evaluated to obtain the score 60.39% ((73.33 + 47.45)/2 + 1 - 1). Below the answer evaluation system results for the same prose and answer is provided.
4. Conclusion

If teaching-learning educational process is considered as a goal through which students can get a meaningful learning of stated concepts, which extend and articulate their network of relations and can apply them in different contexts, it is necessary that teachers include tools to speed up act performance of agents involved in the construction of the new knowledge. In our case, applying a concept map tool in the classroom will allow students being themselves more motivated to carry out proposed activities and to participate in the construction of their own knowledge.

The methodology of developing concept maps discussed in the paper shall be (i) Extended to impart concept map based learning in other regional languages; (ii) Applied to generate associated concept animations for enriching automated content development; (iii) Applied to automated answer evaluation thereby taking part in self-assessment activity of pre-collegiate examinations; (iv) Used to dynamically generate questions and further continue the answer evaluation process in an e-learning setting; and (v) Applied for automatic document summarization.

References
A Machine Translation System for Converting Tamil Text –To –Sign Language

D.Narashiman
Student, M.E. Multimedia Technology-CEG
Anna University, Chennai-600 025
dnarashiman@gmail.com

Dr.T.Mala
Senior Lecturer
Anna University, Chennai-600 025
mala@cs.annauniv.edu

Abstract
This scheme enhances a remarkable approach to decipher the given Tamil text into sign language. The system suggested here is cross modal. It gives a wider interaction of textual input and generates corresponding sign language output in one or a number of sign variants. The application receives Tamil text sentences as input and provides output in 3D animated sequence that can be visualized.

Keywords-Sign language, 3D animation, machine translation.

Introduction
Sign language is a universal language and signs are common features in all the countries around the world. This system has been developed for the deaf community to enhance their communicative ability. This system was developed in such a way that it accepts text as input through keyboard(tactile mode) and generates animated sign symbols as output (visual mode). The benefits of this system include (a) maneuvered for many personal and real time confessions. (b) essential and vital tool for educative purpose to people who are interested in knowing the sign language. (c) The 3D animated output, endows the user to visualize the signs effectively for number of purpose such as news broadcasting and picture animation. This paper talks about the interactive system which translates Tamil text to sign language by compiling of sign language video [5], mapping of tokens and sign symbols [10][11] and creating animated output [3][12]. The sections are arranged as follows: Section two presents the literature survey on sign language. Section three describe the system architecture in general and also explain various phase in generating sign language in detail. Section four gives the method for evaluation of the translation output. Last section deals with conclusion and future enhancement of the system.

Literature Survey
Sign language is represented by non gesture, gesture coding. Non-gesture features include hand movements alone whereas gesture features include hand, face, shoulders, head and facial expression combined with hand sequences. Since sign variants are comparatively fewer than Tamil words the system should be able to generate sign for that corresponding word instantly. Sign language is lexically backward, it has become inventible for a sign generator to exhibit some degree of creativity in assigning concept-to-sign correspondence. When converting the Tamil text into sign language the creative method like finger spelling is used. The sign language mainly deals with non-gesture and gesture features of the signer. In practice however it is highly impossible to separate the context from the given text. Instead we prefer the semantic or conceptual representation of the source text. Various works were carried out earlier to solve this problem. Former approaches for converting text sign language are:
Example-Based method: Morrisey and way investigate corpus-based methods for example-based sign language translation from English to sign language of the Netherlands. With the small corpus and no available lexicon, the system is robust for sentences already encountered in the training set but has problems with unseen combinations of corpus chunks as well as corpus parts that it is unable to align [7].

Theoretical Methods: This method deals with the theoretical issues arising during machine translation of written text to sign language for example a notation for sign which use the 3D space around the signer to form complex expression [3] [4].

Rule-based Method: Safar and Marshall propose a decomposition of the translation process in two steps: initially they translate from written text into a semantic representation of the signs. Afterwards graphically oriented representation is done. Both the steps use rule-based techniques for a specific domain in British sign language [6].

Interlingua Method: An interlingua try to capture the generic fact – stating capacity of language using two different strategies: the first attempts to construct a universal grammar that generalize over the semantics nuances of many language, while the second is knowledge intensive allows for the incorporation of heterogeneous common sense into translation process [8].

Statistical Machine Translation method: This method is used to automatically transfer the meaning of the source language sentences into a target sentence by applying the phrase-based statistical machine translation based on morpho-syntactical analysis [1][2].

**System Architecture**

The system architecture design is given in figure 3.1. The system mainly focuses on the construction of the direct knowledge repository and incremental knowledge repository. The finger spelling knowledge repository (F.S K.R), spatial knowledge repository and rule based knowledge repository comprise the incremental knowledge repository. System uses the knowledge repositories to translate given text to sign language. The pre-processor unit fed with parallel corpus. It converts the given input Tamil text and Sign videos into a suitable form for statistical analysis. The input text is divided into tokens using the white space as delimiter. The sign videos are also segmented. These tokens and corresponding signs are stored in a direct knowledge repository. The tokens for which signs are not available directly are generated and stored in incremental knowledge repository.

*Figure 3.1 System Architecture*
A. Creation of Direct Knowledge Repository

The creation of direct knowledge repository is shown in figure 3.2.

Pre-processing is to convert the given input Tamil text and Sign videos into a suitable form. The text is processed by removing unimportant stop words, special characters and splits into words. Then the pre-processed words are stored in a table along with probability of occurrence using N-Gram methods. This is training the system. While training, statistical systems track common N-grams, learn which translations are most frequently used, and apply those meanings when finding the phrases in the future.

They also analyze the position of N-grams in relation to one another within sentences, as well as words grammatical forms, to determine correct syntax. The system uses the training to develop translation models. The system is fed by the parallel corpus. The language model is applied to the sign images. The system tracks bi-grams and builds the language model. The system then develops translation model from the parallel corpus and uses its training to translate new sentences.

The Direct Knowledge repository contains the index image in which segmented videos are indexed and stored in a database, alignment matrix is constructed in order to obtain the matching of word-to-sign coherence, and Language Model Probability. The language model is applied to the target language (sign language) for reordering and arrangement of the signs. N-Gram calculations are applied and probability of the sign occurring is calculated and stored in the table. Where probability of the output sign depends on the given text and probability of that sign generated for the given text.

A. Creation of incremental Knowledge Repository

The word which does not inherit appropriate sign by using direct knowledge repository is exhibited using incremental knowledge repository. The incremental knowledge repository is created by finger spelling, spatial and rule based method as shown in the figure 3.3. The purpose of this to generate sign for the words not present in the direct knowledge base. The proper noun words, descriptive word and ambiguous words come under this category.

The parts of speech tagger is used to group the words according to proper noun, adjective, ambiguous and action words. Depending upon the classification the sign are generated from respective incremental knowledge repository. The word whose sign are not generated by direct knowledge repository are generated by incremental knowledge repository. The preprocessed word is tagged using the pos tagger.
which gives the information about the word whether proper noun, action and descriptive words according to the classification the process is carried out.

![Figure 3.3 Creation of Incremental knowledge Repository](image)

The proper noun words are represented using this knowledge repository. The process is shown in figure 3.4. This repository contains sign for individual alphabet. The word is tagged by POS tagger if the word is proper noun it is again rechecked using the word net dictionary if synonyms is available it is signed using direct knowledge repository otherwise the word is separated into individual characters and signed respectively. Eg: ֠וֹדַהַםְּ , ָהַםֶנֶפֶש. The name is separated into signal characters and they are signed.

![Figure 3.4 Finger Spelling method](image)

Spatial method is used to represent the descriptive words. The method is shown in figure 3.5. Usually these words describe the place, nature, behavior etc. The main purpose of this module is to reduce the number of sign variants as the certain words have common meanings and the same sign is represented for different words. This process begins with the tagging of the word and check whether the word is descriptive (adjectives), the meaning for the word is retrieved using the word net dictionary then the meanings are compared with the words in the spatial knowledge repository if the meaning match the word then corresponding sign is generated. By this method the signs for different words are generated easily with minimum number of signs video in the database. Eg: בָּדָהַהַםְּ , בָּהַםֶנֶפֶש both the words represent the same meaning of spacious , the same sign is represented for both the word.
Rule based method is used to represent the ambiguous word. The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word. For example, in the word 'é£ò¤Á' it may represent either a day of the week or sun depending on the context in which the sign is generated. A word is assumed to have a finite number of discrete senses, often given by a dictionary, and the task of the forecast is to make a forced choice between these senses for the meaning of each usage of an ambiguous word based on the context of use. This procedure is termed as context-based disambiguation. The word is tagged if the tag is ambiguous and the previous word tag is considered depending on the previous word. The sign is generated. The alternative method is using n-gram model. An n-gram is a type of phrase within a document that has a set of n number of words. N-grams are the basic linguistic unit with which a statistical machine translation system works. The n-gram gives the maximum probability of word occurring together. This is used to check the word occurrence in the context and along with the other words in sequence. This is used to determine the word that are unable to decode correctly, by checking the context and word occurrence. From this the sign for ambiguous word are generated. The rule-based method is shown in the figure 3.6.

Finally, signs are arranged in the logical sequences and the final output is generated. The final output is in the form of 3D animation. If the sign for the corresponding word is not found in the direct knowledge repository, then the word is tagged using the tagger. According to the tag, if the word is a proper noun then the sign is generated using finger spelling knowledge repository, if the word is an adjective then spatial knowledge repository is used and if the word is ambiguous then the rule-based knowledge repository is used.
Evaluation Criteria

The output of the system is final as it should generate the correct sign for the given text. The generated output should be clear, understandable and correct. The hand and facial movements should clear and various actions are understandable. The correct matching of sign and the word should be done. The output of the system is validated using the confusion matrix. The confusion matrix is plotted by taking into consideration the actual generated sign and how it understood. The first quadrant represent the correct sign and understood correctly, second quadrant represent the correct sign but understood incorrectly, third quadrant represent incorrect sign generated but understood correctly and the last quadrant represent the incorrect sign generated and understood incorrectly.

![Confusion Matrix](image)

The confusion matrix shown in the figure 4.1 about 50 words are given as input and output signs are checked, The value in the first quadrant 30 are the signs generated are correct and understood correctly. The value in the second quadrant is obtained when there is difference in retrieval of sign number from the database. The value in the third quadrant is obtain as there is difference in generating spatial sign. The fourth quadrant value is obtain when there is difference in retrieval of sign for the given word as the word is an ambiguous word.

Conclusion

The system described here is very useful for the deaf community. The process starts by breaking the sentence into tokens, mapping the tokens with the sign symbols, generating the perfect symbol in accordance with the conditions and rearranging of symbols to generate the correct meaning of the sentence. The output is a real visualization of sign language and enables interactive application. The main challenge of the system is mapping the sign , generating the new sign if not available in the database and rearranging the symbols according to the meaning of the text checking the grammar of the sentence , adding to the database for future reference. As it is not the only way to generated sign symbols it can be always impoverished to the native speakers. The animation output is selected as it will be pleasing for the viewers. This is mainly designed as a non-human interpreter for sign language. Future development in this system is to increase the sign symbols for the words that are commonly used in day to day life and also generalizing the symbols irrespective of native speakers, to customize the output image according to their interest. The system can be extended to any language by using a language translator with little modification and it also an be extended to translate speech to text and then to sign language.
References


Lexicon for Machine Translation

As he comes suddenly I become functionless.
1. ‘பலர் தவறு பற்றி’ சொல்லின் முதல் ‘பலர்’ சொல்லும் பின்னர் ‘தவறு பற்றி’ சொல்லும் விளக்கம். பலர் போ வெளிப்படுத்துவதற்கு அதில் செய்யும் வார்த்தை வெளிப்படுத்தும்.

2. ‘பலர் தவறு பற்றி’ சொல்லின் முதல் ‘பலர்’ சொல்லும் பின்னர் ‘தவறு பற்றி’ சொல்லும் விளக்கம். பலர் போ வெளிப்படுத்துவதற்கு அதில் செய்யும் வார்த்தை வெளிப்படுத்தும்.

3. ‘பலர் தவறு பற்றி’ சொல்லின் முதல் ‘பலர்’ சொல்லும் பின்னர் ‘தவறு பற்றி’ சொல்லும் விளக்கம். பலர் போ வெளிப்படுத்துவதற்கு அதில் செய்யும் வார்த்தை வெளிப்படுத்தும்.

‘பலர்’ சொல்லின் முதல் வார்த்தையின் விளக்கம் (word senses) (meaning) அதில் போ வெளிப்படுத்துவதற்கு அதில் செய்யும் வார்த்தை வெளிப்படுத்தும். பலர் போ வெளிப்படுத்துவதற்கு அதில் செய்யும் வார்த்தை வெளிப்படுத்தும்.

அதன் அறிக்கைகளில் ஒன்று தரவு விளக்கத்திற்கு பங்கு கொள்ளத் தீர்மானம். அதன் அறிக்கையின் போ வெளிப்படுத்தும் வார்த்தையின் விளக்கத்திற்கு பங்கு கொள்ளத் தீர்மானம். அதன் அறிக்கையின் போ வெளிப்படுத்தும் வார்த்தையின் விளக்கத்திற்கு பங்கு கொள்ளத் தீர்மானம்.
பல்லியல் அறிவு அறிவியலின் தசாலையின் பல்லியல் முறை என்பது அறிவியல் என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு. பல்லியல் செப்புக்கான அறிவுவர்களின் விளக்கம் என்ற தலைப்புடன் தனியே குறுக்கு என்றது. பல்லியல் செப்புக்கான அறிவுவர்களின் விளக்கம் என்ற தலைப்புடன் தனியே குறுக்கு என்றது.

பல்லியல் அறிவு அறிவியலின் பல்லியல் முறை என்பது எதையும் எண் என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு. பல்லியல் செப்புக்கான அறிவுவர்களின் விளக்கம் என்ற தலைப்புடன் தனியே குறுக்கு என்றது. பல்லியல் செப்புக்கான அறிவுவர்களின் விளக்கம் என்ற தலைப்புடன் தனியே குறுக்கு என்றது.

பல்லியல் (Wordnet)

பல்லியலான அறிவு என்று என்னடைக்க வரும் பல்லியலான அறிவுப் போராட்டம் (Relational Semantics), முதல் பல்லியலக்காரர்களின் கூறுக்கு, பல்லியலான செப்புக்கான பல்லியலான செப்பு என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு. பல்லியலான செப்பு என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு. ‘நூத்’ என்று எடுத்து காட்சிகள் என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு.

பல்லியலையிட்டு அறிவியலை (Lexical Semantics) என்று எடுத்து காட்சிகள் என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு. பல்லியலையிட்டு அறிவியலை (Generative operation) என்று எடுத்து காட்சிகள் என்ற சொல்லின் குறிப்பிட்டலின் ஒரு மையப்படுத்தப்பட்ட தலைப்பு.

581


(Generative Lexicon)

"(Generative Lexicon) என்பது ஒரு விளக்கியலார் பார்வையும் மாற்றுதலும் பரவலானது பார்வைகள் புரத்தையும் பிரித்தையும் பார்வையும் அறிவு பல்வேளப்படுத்தும் குறிச்சீட்டுற்ற முயற்சியும் பார்வையும் அந்தக் குறிச்சீட்டுற்ற குறிச்சீட்டுற்ற "(Pustejovsky J.). (Generative Lexicon) என்பது ஒரு விளக்கியலார் பார்வையும் மாற்றுதலும் பரவலானது பார்வைகள் புரத்தையும் பிரித்தையும் பார்வையும் அந்தக் குறிச்சீட்டுற்ற குறிச்சீட்டுற்ற." 

கதைகளை பார்வையும் பார்வையும் அந்தக் குறிச்சீட்டுற்ற குறிச்சீட்டுற்ற மற்றும் அருகில் குறிச்சீட்டுற்ற குறிச்சீட்டுற்ற மற்றும் பார்வையும் மாற்றுதலும் பரவலானது பார்வைகள் புரத்தையும் பிரித்தையும் பார்வையும் அந்தக் குறிச்சீட்டுற்ற குறிச்சீட்டுற்ற குறிச்சீட்டுற்ற 

582
1. இயற்கையேயானவை (Argument Structure)
2. பிராப்பிட்டு வகையில் (Event Structure)
3. பொருள் வகையில் (Qualia Structure)
4. பராமரிப்பு (Lexical Inheritance)


Abstract

Grammatical descriptions of human languages are the results of efforts in modeling of the design features and the internal organization of the structures and the mechanisms. Therefore, Linguistics is about language modeling, designing and studying their theoretical and practical implications. However the activity of grammatical descriptions itself is molded by the specific needs of aims and the goals such as Teaching and Learning a language, investigating the issues related to the evolutionary biology with regard to discovering the universals of human language and development, philosophical and functional aspects of language and Linguistic Computing. Here, we would like to discuss certain issues towards building a Hyper grammar for a given language.

1. Concept

A Hyper grammar is a non-linearly organized dynamic grammar based on the hypertext format. It is intended to simulate certain functions of a native speaker. It can be used both as learning and teaching tool besides as a reference grammar.

It is comprised of a number of non-linearly arranged texts each with a comprehensive note on various grammatical facts of Tamil, with hyper-links. It can be accessed and retrieved for various purposes involving language, to experience the effect of a native speaker and hearer of the language. Functionally it serves better than any of the existing printed grammars, which are simply flat and linear. In a way the existing printed grammars are non-communicative i.e. passive, hence, they are monologues and do not participate or reciprocate to pass judgments about the linguistic facts of the respective languages. A grammar in order to reciprocate should have some of the computationally implemented tools like a morphological generator, analyzer, chunker, parser, lexical accessor etc.

The Hyper grammar is intended to be a reciprocative grammar, as it involves some of the properties like the native speaker’s ability to make judgments on the grammaticality of the linguistic facts. This single feature makes it distinct from the printed grammars. Hyper grammars are extremely useful from the point of learning, teaching and as reference material.

The design features are borrowed from the hypertext format but conceived as a computationally cognitive model. The contents are being developed from both the published and unpublished sources carefully selected and rewritten in the hypertext format.
2. The Contents:

The content of Tamil Hyper grammar has two main components, viz. 1. the description of grammar in hypertext format and 2. the applicational aspect of the Tamil Language as a language manager.

2.1. The Tamil Grammar:

The grammar part includes a number of comprehensive descriptive notes on certain linguistic facts of the Tamil Language. It is conceived in terms of a Computational Grammar. It deals with the Orthography, the design features of Tamil script, the orthographic syllable, the information on the frequency distribution of written syllables etc.

As part of the Tamil morphology, we have information on Tamil categories viz. nouns, adjectives, verbs, adverbs, numerals, pronouns etc. In each of these, there is information regarding the setting up of paradigm types and a list of paradigmatic forms under each category. One can access information regarding the most frequent hundred words, five thousand words and ten thousand words in terms of their frequencies, and communicative contribution to the coverage in Tamil Texts. As regards to the frequency of Tamil characters and syllables as they occur in the 3 million-word corpus, one can find the relevant information. One of the most important and crucial is the lexical component. A number of bilingual dictionaries like Tamil-Hindi, Tamil-Kannada, Tamil-Tamil, Tamil-Oriya, Tamil-Marathi, Tamil-English and English-Tamil – are included. Originally these dictionaries are conceived as bilingual and bi-directional dictionaries initially created using the most frequently occurring words ensuring the coverage.

2.2. The Tamil language manager:

This is the most crucial component of Tamil Hyper grammar. It involves the actual functions of the practical aspect of the grammar outlined above. As said earlier, the grammatical description is only a statement about the competence of a native speaker – about his/her language. In order to make it to simulate the grammar, it should involve a working generator, analyzer, parser and lexical accessor, etc. Currently the Tamil language manager includes a word form generator, a morphological analyzer and lexical accessor among others.

a. The Morphological Analyzer:

The word analyzer incorporated here is intended to analyze the Tamil words in terms of the lexical root/stem, its category, the paradigm type and the inflectional or derivational affixes attached to it. A morphological analyzer (Morph) engine essentially learns from a morphological lexical database of a particular language. The functional coverage and efficacy of the engine is greatly dependent on the structure and the organization of the database. The database of Tamil Morphological Analyzer comprises of inflectional database and the root dictionary. These data comprise purely linguistic information of the language, which are processed subsequently to enable using it in morphological analysis. It uses the Word and Paradigm Model of analysis.
The Organization of the Linguistic data for Morph:

(i) The paradigmatic-data

The term Paradigm refers to an exhaustive set of morpho-syntactically related word forms of a given lexeme. Based on the inflection, six distinct morphological categories are identified and the paradigms are created. They include the major and minor categories of words.

(a) The major word classes which are productive and open class categories (new members are added from time to time) can inflect with distinct but characteristic suffixes which explicit morpho-syntactic functions. The major word categories are listed as below,

1. Nouns
2. Verbs
3. Adjectives

(b) The distinct minor categories which are productive but considered as closed class categories (no new members are added) are listed below,

4. Pronouns
5. Numerals
6. Locative Nouns

The other class of words which are not fallen under the above categories are a list of idiosyncratic word forms. They cannot inflect for any functional categories. They come under functional categories of language with defective morphology. The following words are usually known as indeclinable and have no morphology to process.

1. Postpositions
2. Adverbs
3. Conjunctions
4. Interjections
5. Particles

The above words are listed as 'Avy' (avyayas) in the dictionary.

(ii) Root Dictionary

Root Dictionary is a vast collection of lexemes which contains words, their categorical information and their suitable paradigms. It includes a certain number of minimally distinct words in the semantic system of a language. This is typically called as lexicon without semantics.

Input : a valid word form
Output : 1. Root
         2. Lexical Category
         3. Paradigm type
         4. Morphological Category
         (The output may be one or more analysis)
Input and Output Specifications in Tamil:

Input:
1 koyampuwUr
2 iraNtAvawu
3 mikappeVriya
4 wamilYYaka
5 mAnakarakam
6 Akum

Output:
1 koyampuwUr unk <fs af='koyampuwUr,n,n,sg,3,d,0,0'>
2 iraNtAvawu unk <fs af='iraNtu,num,,sg,,o,Avawu,Avawu'>
3 mikappeVriya unk <fs af='mikappeVriya,adj,any,any,any,,0,0'>
4 wamilYYaka unk <fs af='wamilYYakam,n,n,sg,3,o,0,0'>
5 mAnakarakam unk <fs af='mAnakaram,n,n,sg,3,d,0,0'>
6 Akum unk <fs af='Aku,v,n,sg,3,pp,pp'> | <fs af='Aku,v,n,pl,3,pp,pp'>

b. Word form Generator:

A Tamil word-form synthesizer enables a user to generate Tamil word forms. The user is prompted to select some choices leading to the generation of the desired word. This is extremely useful to the learners of Tamil as second language. Such uses can interactively generate the requested word in Tamil.

The Morphological Generator of Tamil is based on Word and Paradigm Method. It is built using the feature values, suffix informations with add or delete rules and the root word dictionary with its category and paradigm. It uses the Machine Learning techniques to generate the word form from the given input. The basic resources required for present word synthesizer:

Feature Value:
It contains the category, its possible morpho-syntactic properties. It has five values, each viz., category, gender, number, person and the affix. For instance,

“v m sg 3 nw”
The above is an example for the verb for generating third person singular masculine past tense verb as such vanwAnY ‘he came’.

Suffix information and synthesis rule set:
This is generated from the paradigms and its feature values. It contains the rules for words based on their morpho-phonemic processes. It has four columns delimited by comma. For instance, to generate

‘marafkalYE’

maram + kalY +E
Eng: tree + plural+ Accusative case
the suffix information table consists,

“Eylakf,m,maram,89”
Whereas the first is an inverted suffix of 'fkalYE' which is to be added, the second is the word which has to be deleted from root and the third is the name of the paradigm as such the word behaves in its morphophonemic process and finally the row number of the feature value file.

Lexicon:

Lexicon consists of the root words of Tamil, its category and the name of the paradigm based on its phonological behaviour in its inflection. For instance,

'aNi,v,varE' Eng: put on, verb, draw
Here aNi, the verb act morpho-phonemically as varE.

\[ aNi-nw-AnY \quad as \quad varE-nw-AnY \quad 'root-PST-3p.sg.m' \]
\[ aNi-kirY-AlY \quad as \quad varE-kirY-AlY \quad 'root-PRS-3p.sg.f' \]
\[ aNi-v-ArkalY \quad as \quad varE-v-ArkalY \quad 'root-FUT-3p.sg.m' \]

Input :
1. Root
2. Lexical Category
3. Morphological Category

Output :
a valid word form

Input and Output Specifications in Tamil:

Input:
1. kampar NNP  \(<fs \ af='kampar,n,m,sg,3,0,0,0'>\)
2. irAmAyaNam NN  \(<fs \ af='irAmayaNam,n,n,sg,3,o,E,E'>\)
3. iyarYrYu VM  \(<fs \ af='iyarYrYu,v,m,sgh,3,,nw,nw'>\)

Output:
1. kampar  NNP  \(<fs \ af='kampar,n,m,sg,3,0,0,0'>\)
2. irAmAyaNawwE  NN  \(<fs \ af='irAmayaNam,n,n,sg,3,o,E,E'>\)
3. iyarYrYinYAr  VM  \(<fs \ af='iyarYrYu,v,m,sgh,3,,nw,nw'>\)

c. Dictionary:

The Tamil-Telugu bilingual dictionary is built based on the concepts available in the language. It differs from the conventional dictionary which lists words but not concepts. Here, the lexeme(s) are related to each other on the basis of the concept i.e. the idea of ontological entity. The dictionary which is based on concepts is a better one for obtaining a concise and effective lexicon which can be used in many NLP applications.

d. Machine Translation System:

The development of Machine Translation is one of the most challenging tasks of the Natural Language Processing Applications. The development of Machine Translation (MT) System which translates texts from Telugu to Tamil and vice-versa (Bi-directional) are incorporated here. This MT system was developed as part of IL-ILMT consortium project funded by the Government of India at CALTS, University of Hyderabad. This Machine Translation system uses Transfer Based Approach. The System's Architecture is divided into three stages i.e. Source language Analysis module (SLA), Source language to Target language Transfer module (SL-TL) and Target language generation module (TLG).
(i) Telugu-Tamil Machine Translation system:

The crucial tools used in Telugu-Tamil Machine Translation system includes,

a. Source Language Analysis
   Telugu Sandhi Splitter
   Telugu Morphological Analyzer
   Telugu POS Tagger
   Telugu Chunker
   Telugu NER (Named Entity Recognizer)
   Telugu Parser

b. Source Language- Target Language Analysis
   Telugu-Tamil Transfer Grammar Module
   Telugu-Tamil Multi Word Expression Module
   Telugu-Tamil Lexical Transfer Module

c. Target Language Analysis
   Tamil Agreement Module
   Tamil Word form Generator

(ii) Tamil-Telugu Machine Translation:

The crucial tools used in Tamil-Telugu Machine Translation system includes,

a. Source Language Analysis
   1) Tamil Sandhi Splitter
   2) Tamil Morphological Analyzer
   3) Tamil POS Tagger
   4) Tamil Chunker
   5) Tamil NER (Named Entity Recognizer)
   6) Tamil Parser

b. Source Language- Target Language Analysis
   1) Tamil-Telugu Transfer Grammar Module
   2) Tamil-Telugu Multi Word Expression Module
   3) Tamil-Telugu Lexical Transfer Module

c. Target Language Analysis
   4) Telugu Agreement Module
   5) Telugu Word form Generator

3. Conclusion:

The Tamil Hyper Grammar thus described here is the most significant development in the recent applications of Natural language processing of the Tamil language to be used as teaching, learning as well as reference grammar for all kinds of language users.

1 Transliteration Scheme using wx-notation:
   Tamil Orthography :
   a A i I u U e E o O H
   k f c F t N w n p m y r l v lYY lY rY nY j s h R
Telugu Orthography:
a A i I u U q Q e E o O M H
k K g G f c C j J f t T d D N w W x X n p P b B m y r r Y l l Y Y v S R s h

References:
A Tamil - Telugu Bi-directional
Machine Translation System

Christopher M | ethachris@gmail.com
Krupanandam N | nityakrupa@yahoo.co.in
Parameshwari K | parameshkrishaa@gmail.com
Uma Maheshwar Rao G | guraohyd@gmail.com
Vijaya Bharathi D | vijaya.anthony@gmail.com
Centre for Applied Linguistics & Translation Studies
University of Hyderabad, Hyderabad-500046.

Abstract
We present the development of Machine Translation (MT) System which translates texts from Telugu to Tamil and vice-versa (Bi-directional). This MT system was developed as part of IL-ILMT consortium project funded by Govt. of India at CALTS, University of Hyderabad. This Machine Translation system uses Transfer Based Approach. System’s Architecture is divided into three stages i.e. Source language Analysis module (SL), Source language to Target language Transfer module (SL-TL) and Target language generation module (TL). The computational Modules that are used in the building of this system were developed mainly by CALTS-UoH, IIIT-H and AUKBC research teams. We also use the statistical open source engine i.e. CRF++ for POS-Tagging, Chunking and Named Entity Recognizer (NER). Hence the architecture is a hybrid one.

1. Introduction: The development of Machine Translation (MT) is one of the most challenging tasks of Natural Language Processing Applications. In MT there are a number of methods that are being practiced all over the world, chiefly, they are Direct Methods, Interlingual Methods, Transfer Based Approach and a combination of these beside the statistical and corpus based methods. Tamil and Telugu are two closely related languages, which belong to the same i.e. the Dravidian language family. Even though they belong to the same language family, still they exhibit a considerable amount of diversity at every level viz. morphological, syntactic, semantic and lexical levels. Keeping these in mind, building a Machine Translation System for this language pair using Transfer based Method can be non-trivial and challenging. The present paper discusses the successful implementation of the Transfer Based Approach to the Machine Translation (MT) System for the Telugu-Tamil pair. This bi-directional Telugu-Tamil MT system is one of the nine pairs of Indian Language to Indian Language Machine Translation Systems (IL-ILMT) planned to be developed by the Consortium of IL-ILMT constituted by the DIT, MIT, Govt. of India. The system is an assembly of various linguistic modules run on specific engines whose output is sequentially maneuvered and modified by a series of modules till the output is generated. The most crucial linguistic modules include, a Morphological Analyzer (MA), Parts of Speech Tagger (POS-T), a
Simple Parser (SP), the Transfer Grammar Component (TG), a Lexical Transfer module consisting of a Bilingual Dictionary and a Conceptual Dictionary, an Agreement module (AGR) and a Morphological Generator (Wordgen). The system is already built and is now being tested and evaluated. The presentation would involve the demonstration of a randomly selected text from the internet.

2. Module Level details of the System:

2.1. The Format:

The entire system works on a unique standard format called the Shakti Standard Format (SSF). The multi-column format vividly represents input and output of in each module throughout the system. This is especially designed to represent the different kinds of linguistic analyses, as well as different levels of analysis. The two kinds of analyses are: 1. Constituent level analysis and 2. Relational-Structure level analysis. The former is used to store simple phrase level analysis and the latter for storing relations between the simple parses. Feature structures are used to store attribute-value pairs for a phrasal node as well as for a word or a token. Attribute value pairs store relations in different columns. The following is a description of the column format in SSF:

Column 1 stores the node address, mainly for human readability. Column 2 stores the word or word-group input. The symbol “((“ represents the start of the word or word-group and the symbol “))” to represent the end of the word or word-group. Column 3 stores the chunk name or the POS tag of the words occurred in the sentence. Column 4 stores the Morphological information (feature structures) of the words. Column 5, 6, and 7 store the gender, number and person feature values respectively. Column 8 stores the oblique or direct nature of the stem in case of nouns and 9 the case marker in case of nouns and tense in case of verbs. Column 10 store the exact suffix representing the features represented in 5-9.

2.2. Source Analysis:

Tokenizer: The tokenizer converts a text into a sequence of tokens (words, punctuation marks, etc.) within the Shakti Standard Format.

a. Morphological analyzer (MA): A Morphological Analyzer analyzes and identifies the root and the grammatical features of the word. Word and Paradigm based approaches have given good success rates for Indian languages (Rao et. al. 2007). The computational module used in this MT system provides about 20-30 features values for each word, in which 8 are mandatory viz. root, lexical category, gender, number, person, case, case marker or tam and suffix. The lexical categories are divided into 9, they are noun (n), verb (v), adjective (adj), pronoun (pn), adverb (adv), postposition (psp), number (num), nouns of space and time (NST) and indeclinable (Avy) (Rao et.al. 2007).

1  himAlayAlu unk  \textless{}fs af='himAlayaM,n,pl,,d,0,0'>|\textless{}fs af='himAlayaM,n,pl,d,vu,vu'>
2  sahaja unk  \textless{}fs af='sahajaM,n,sg,,o,ti,ti'>|\textless{}fs af='sahajudu,n,sg,,o,ti,ti'>
3  sixXaMgA unk  \textless{}fs af='sixXaMn,sg,,gA,gA'>
4  erpadAyAi unk  \textless{}fs af='erpadu,v,n,pl,3,,A,A'>
5  . unk  \textless{}fs af='punc,,,,,'>

b. Parts of speech tagger (POS-T): Part of speech tagging is the process of assigning a unique part of speech to each word (token) in the sentence. This process helps in identifying the role of each word
(token) in a sentence. There are number of approaches, such as rule-based, statistics based, transformation-based etc. which use for POS tagging. Here we propose to use statistical techniques on a Gold standard manually developed tagged text (follows ILMT Tagset, ILMT 2007).

c. **Chunker:** Chunking involves identifying non-recursive combinations of word groups involving nouns (NP), verbs (VGF/VGNF), adjectives (JJP) and adverbs (RBP) etc. in a given sentence. Here we use statistical methods to identify and chunk tags in a sentence (following ILMT Tagset, ILMT 2007).

1.1 himAlayAlu NN <fs af='himAlayaM,n,,pl,,d,0,0'> | <fs af='himAlayaM,n,,pl,,d,]v,u,vu'>

2.1 sahaja NN <fs af='sahajaM,n,,sg,,o,ti,ti'> | <fs af='sahajudu,n,,sg,,o,ti,ti'>

3.1 sixXaMgA RB <fs af='sixXaM,n,,sg,,gA,gA'>

4.1 erpaddAyi VM <fs af='erpadu,v,n,pl,3,,A,A'>

4.2 . SYM <fs af='.,punc,,,,,,'>

d. **Named Entity Recognizer (NER):** The identification, recognition and tagging of proper nouns such as names of persons and organizations (ILMT 2007) is achieved by this module.

1.1 himAlayAlu NN <fs af='himAlayaM,n,,pl,,d,0,0' head="himAlayAlu" ENAMEX TYPE="LOCATION" SUBTYPE_1="LANDSCAPES">

2.1 sahaja NN <fs af='sahajaM,n,,sg,,o,ti,ti' head="sahaja" vpos="vib1">

3.1 sixXaMgA RB <fs af='sixXaM,n,,sg,,gA,gA' poslcat="NM" head="sixXaMgA">

4.1 erpaddAyi VM <fs af='erpadu,v,n,pl,3,,A,A' head="erpaddAyi">

4.2 . SYM <fs af='.,punc,,,,,,'>
e. Simple parser (SP): Identifies and names Thematic relations between a verb and its participant noun in the sentence, based on the Computational Paninian Grammar framework (Bharathi et.al 1995).

```
1 NP <fs af='himAlayAluM,n,,pl,,d,0,0' head="himAlayAlu" drel=k1:4 name=1 ENAMEX TYPE="LOCATION" SUBTYPE_1="LANDSCAPES">
   1.1 himAlayAlu NN <fs af='himAlayAluM,n,,pl,,d,0,0' name="himAlayAlu">
   )
2 NP <fs af='sahajaM,n,,sg,,o,ti,ti' head="sahaja" vpos="vib1" name=2>
   2.1 sahaja NN <fs af='sahajaM,n,,sg,,o,ti,ti' name="sahaja">
   )
3 RBP <fs af='sixXaMgA,n,,sg,,gA,gA' head="sixXaMgA" name=3 poslcat="NM">
   3.1 sixXaMgA RB <fs af='sixXaMgA,n,,sg,,gA,gA' poslcat="NM" name="sixXaMgA">
   )
4 VGF <fs af='erpaddAyi,v,n,pl,3,,A,A' head="erpaddAyi" name=4>
   4.1 erpaddAyi VM <fs af='erpaddAyi,v,n,pl,3,,A,A' name="erpaddAyi">
   4.2 SYM <fs af='.',punc,,,,,,'> 
)```

f. Transfer Grammar (TG): Wherever, the source language does not have an equivalent structure in the target language, a structural transformation is required to convert the source language structure into an acceptable target language structure. Such cases can be found at all levels wherever divergence occurs between the source and the target language. TG Module contains rules which convert the parsed structure of the source language into the desired structure in the target language giving the acceptable target structures.

```
1 NP <fs af='boVrrA,n,,sg,,0,0_A' head="boVrrA" drel=k1:9 name=3>
   1.1 boVrrA NNP <fs af='^@boVrrA,n,,sg,,0,0_A' ENAMEX/TYPE="LOCATION" name="boVrrA" SUBTYPE_1="PLACE">
   )
2 NP <fs af='guhalu,n,,pl,,d,0,0' head="guhalu" drel=k2:9 name=4>
   2.1 guhalu NN <fs af='^@kukE,n,,pl,,d,0,0' name="guhalu">
   )
3 NP <fs af='10,n,,num,,,,,,' head="10" name=5>
   3.1 10 QC <fs af='10,n,,num,,,,,,' name="10" NUMEX TYPE="COUNT">
   )
4 NP <fs af='^@ANtu,n,,pl,,o,ti,ti' vpos="vib2" head="elYla" name=6>
   4.1 elYla QC <fs af='elYla,n,,pl,,o,ti,ti' poslcat="NM" NUMEX/TYPE="COUNT">
   4.2 ANtu NN <fs af='^@ANtu,n,,pl,,o,ti,ti' name="elYla">
   )
5 NP <fs af='kriwaMsMnAtivi,n,n,sg,3,,0,0' head="kriwaMsMnAtivi" name=7 poslcat="NM">
```

595
5.2. **Multi-Word Expression Transfer (MWE):** Multi-Word Expression module involves identification and transfer of frequently used non-compositional phrases, compounds, reduplicatives, etc. from the source language to the target language.

- **Lexical transfer (LT):** Root words identified by the morphological analyzer are looked up in a bilingual dictionary for the target language equivalent using concept substitution including function words.
**i. Agreement (Agr):** Performs checking and reconstructing gender-number-person agreement between the subject and the predicate in the target sentence, ensuring proper agreement.

**j. Vibhakti Splitter or Complex Inflection Splitter (VBS):** Separates complex cases of inflections involving postpositions and auxiliary verbs ensuring proper word generation.

```plaintext
1  ( ( NP <fs af='AnwirA,n,n,sg,,d,0,0' head=AMXra drel=k1:5 name='1'>
1.1 0 NNP <fs af='0,n,n,sg,,d,0,0' ENAMEX/TYPE=LOCATION name=AMXra
SUBTYPE_1=PLACE>
   ) )
2  ( ( NP <fs af='pirawecam,n,n,sg,,d,0,0' head=praxeS name='2'>
2.1 Anwirapirawecam NN <fs af='Anwirapirawecam,n,n,sg,,d,0,0' name=praxeS>
   ) )
3  ( ( NP <fs af='hEwarApAw,n,n,sg,,E,ni' vpos=vib1 head=hExarAbAxni drel=k2:5
name='3'>
3.1 hEwarApAw NN <fs af='hEwarApAw,n,n,sg,,E,ni' ENAMEX/TYPE=LOCATION
name=hExarAbAxni SUBTYPE_1=PLACE>
   ) )
4  ( ( RBP <fs af='walEnakaram,n,,sg,,,Aka,gA' head='rAjaXAnigA' name='4'
poslcat='NM'>
4.1 walEnakaram RB <fs af='walEnakaram,n,,sg,,,Aka,gA' name=rAjaXAnigA poslcat=NM>
   ) )
5  ( ( VGF <fs af='peVrYu,v,fn,sg,3,,i_iru+nw,i' vpos='tam1_2' head='kaligi' name='5'>
5.1 peVrYu VM <fs af='peVrYu,v,any,any,any,,i,i' name=kaligi>
5.2 iru+nw VAUX <fs af='iru,v,fn,sg,3,,nw,'>
5.3 . SYM <fs af='.,punc,,,'>
   ) )
```

**k. Word generator (WG):** This module takes root words and their associated grammatical features, selects appropriate suffixes and concatenates them into well formed word forms.

```plaintext
1  ( ( NP <fs af=',' n, n, sg,, d, 0, 0' head='.' drel='k1:5' name='1'>
1.1 NNP <fs af='0,n,n,sg,,d,0,0' ENAMEX/TYPE='LOCATION' name='.'
SUBTYPE_1='PLACE'>
   ) )
2  ( ( NP <fs af='.' n, n, sg,, d, 0, 0' head='.' drel='k2:5' name='2'>
2.1 ') NN <fs af='.' n, n, sg,, d, 0, 0' name='.'
   ) )
3  ( ( NP <fs af='.' n, n, sg,, ni' head='.' vpos='vib1' drel='k2:5'
name='3'>
```

597
l. Post Processing: Enables if any unacceptable sequences of words to be modified in terms of more acceptable structures of the target language.

1 (( NP <fs af='有一定的,n,n,sg,,d,0,0' head='有一定的' drel='k1:5' name='1'>
  1.1 NNP <fs af='0,n,n,sg,,d,0,0' ENAMEX/TYPE='LOCATION' name='有一定的' SUBTYPE_1='PLACE'>
  ))
2 (( NP <fs af='有一定的,n,n,sg,,d,0,0' head='有一定的' name='2'>
  2.1有一定的 NN <fs af='有一定的,n,n,sg,,d,0,0' name='有一定的'>
  ))
3 (( NP <fs af='有一定的,n,n,sg,,ni' head='有一定的' vpos='vib1' drel='k2:5' name='3'>
  3.1有一定的 NN <fs af='有一定的,n,n,sg,,ni' ENAMEX/TYPE='LOCATION' name='有一定的' SUBTYPE_1='PLACE'>
  ))
4 (( RBP <fs af='有一定的,n,n,sg,,gA' head='有一定的' name='4' poslcat='NM'>
  4.1有一定的 RB <fs af='有一定的,n,n,sg,,gA' name='有一定的' poslcat='NM'>
  ))
5 (( VGF <fs af='有一定的,v,fn,sg,3,i' head='有一定的' vpos='tam1_2' name='5'>
  5.1有一定的 VM <fs af='有一定的,v,any,any,any,i' name='有一定的'>
  5.2有一定的 VAUX <fs af='有一定的,v,fn,sg,3,i'>
  5.3 . SYM <fs af='.',punc,,,,,'>
3. Conclusion:

The architecture of this system is based on analyze-transfer-generate paradigm. The flow of the input sentence in the system is given in fig:1.

All the modules have been integrated on the dashboard, a tool, where the data flow in the pipeline is configured. This ensures speed, since it uses shared memory. This MT system demonstrated here is a completely automated translation system without involving human interference for the first time involving Tamil. Though the current system is built for the tourism domain, it can be extended to any other domain. The system can be used to translate web pages or text material from books, magazines, newspapers etc. written in standard language. It runs on Linux platform with Apache-2.0 server. The browser used for the online translation can be Firefox 1.0.4, IE 6.0 or Mozilla 1.7.8.

Sample Input and Output of the System in Dashboard.
References:


Certain issues in the Development of Telugu - Tamil Machine Translation
A view from the lexicon

Parameswari K | parameshkrishnaa@gmail.com
Uma Maheshwar Rao G | guraohyd@gmail.com
Krupanandam N | nityakrupa@yahoo.co.in
Lavanya J | jlavanyacse08@gmail.com
Christopher M | efthachris@gmail.com

Center for Applied Linguistics and Translation Studies
University of Hyderabad, Hyderabad - 500046.

Abstract:

Machine Translation (MT) is one of the interesting and challenging tasks of Natural Language Processing. In any Machine Translation system, understanding the pair of languages involved are vital. The present work focuses on certain issues in the development of Telugu-Tamil Machine Translation from the point of the languages involved and the dictionaries that are used in Telugu-Tamil Machine Translation System which are unique since they are based on concept. The paper deals with the compilation of concept based dictionary for Machine Translation purpose and also deals with the divergences arise due to the differences in the lexemes of Telugu and Tamil.

1. Introduction:

Tamil, the South Dravidian Language and Telugu, the South Central Dravidian language are major languages of South India. The Machine Translation between Telugu-Tamil is a best example case taken for the development of MT since there is a great demand for the Translation of texts of each of these languages. Normally, Machine Translation is a challenging task where computers take over the task of translating one language into another. Though the languages involved viz. Telugu – Tamil are closely related, exhibit a number of dissimilarities in their linguistic behavior thus making the task a non trivial one.

The paper deals with the issues in the development of an automatic Telugu-Tamil Machine Translation System which is being developed under the project of IL-IL MT at CALTS, University of Hyderabad as part of the Consortium of Indian Languages to Indian languages Machine Translation Systems funded by DIT, Ministry of Information Technology, Government of India.

The lexical resources are essential for building any Machine Translation system. The Lexicon used in the building of Telugu-Tamil MT is one of the Machine Readable Dictionary types, which differs from printed conventional dictionaries of everyday use. The conventional dictionary is usually meant for defining and providing description about a lexeme. However, the concept based dictionary which is currently used contains lexemes without any encyclopedic knowledge.
2. Concept based Dictionary or Synset:

A concept is an idea which is language specific and based on the ontology of lexemes in languages. The concept Based dictionary is a component of a multilingual dictionary developed for 11 languages: English, Hindi, Bengali, Marathi, Punjabi, Urdu, Tamil, Kannada, Telugu, Malayalam and Oriya (Cf. Mohanty et.al.) by different Indian Institutions and are used in NLP applications. The greatest advantage of using synset is the conceptually related words are grouped under a single concept and the equivalents in the target language along with linkages provided. Here, Hindi is used as a pivot language and other synsets in other languages are built on the principle of translational equivalence. The Telugu-Tamil Machine Translation System uses Telugu and Tamil synsets which are developed by CALTS' NLP group and AU-KBC NLP group respectively. A lexeme used to express a concept in a language may not have the same meaning in all the contexts. The same lexeme may be found in different contexts expressing different meanings or concepts. For instance, a lexical item in Telugu corresponds to one or more lexical (sense's) items in Tamil.

In Telugu, the word 'kuttu' is translated in Tamil as,

a. 'kati' in the context of clima kuttu 'to bite as an ant',
b. 'we' in the context of battalu kuttu 'to stitch clothes',
c. 'kuwwu' in the context of ceVvulu kuttu 'to pierce ears'.

Here the question of providing an appropriate equivalent for 'kuttu' requires word sense disambiguation. The concept which is the central point of this lexeme can help to avoid this problem. The dictionary which we are proposing as a suitable one for Machine Translation is of concept centered one rather than of one to one lexical matching. A word X in a language is taken as a concept, and the conceptually related words of X are provided as W1,W2,...Wn. The hierarchy of frequency is followed in an ascending order of giving equivalents. The links (L) are created between the source language and the target language lexemes. The concept Dictionary is used to perform a lexical transfer of the following:

(a) Situation (1) One to One :

Here a single lexical item is linked with a corresponding lexical item in Tamil.

Ex: X (sw1/L1 <-> tw1)

(where X is a context with category, sw is a source word, L is link, tw is target word)

Ex:

Telugu :
ID :: 7350
CAT :: NOUN
CONCEPT :: /uni0C0E/uni0C35/uni0C30/uni0C3F/uni0C15/uni0C48/uni0C28/uni0C3E /uni0C07/uni0C1A/uni0C4D/uni0C1A/uni0C3F/uni0C28/uni0C2A/uni0C4D/uni0C2A/uni0C41 /uni0C32/uni0C47/uni0C26/uni0C3E /uni0C2C/uni0C4D/uni0C2F/uni0C3E/uni0C02/uni0C15/uni0C41 /uni0C2E/uni0C4A/uni0C26/uni0C32/uni0C48/uni0C28 /uni0C35/uni0C3E/uni0C1F/uni0C3F/uni0C32/uni0C4B /uni0C15/uni0C42/uni0C21/uni0C2C/uni0C46/uni0C1F/uni0C4D/uni0C1F/uni0C3F/uni0C28 /uni0C21/uni0C2C/uni0C4D/uni0C2C/uni0C41/uni0C32/uni0C41 /uni0C07/uni0C38/uni0C4D/uni0C24/uni0C3E/uni0C21/uni0C41 /

EXAMPLE :: /uni0C36/uni0C4D/uni0C2F/uni0C3E/uni0C02 /uni0C35/uni0C21/uni0C4D/uni0C21/uni0C3F /uni0C21/uni0C2C/uni0C4D/uni0C2C/uni0C41/uni0C32/uni0C41 /uni0C07/uni0C38/uni0C4D/uni0C24/uni0C3E/uni0C21/uni0C41

SYNSET-TELUGU :: /uni0C35/uni0C21/uni0C4D/uni0C21/uni0C3F /TAM1

This is the link to Tamil.
(b) Situation (2) Many to One:

Here multiple lexemes are displayed with linkages with a single lexical item.

**Ex: X(sw1/L1, sw2/L1, sw3/L1, sw4/L1, sw5/L1 <-> tw1)**

**Telugu**

<table>
<thead>
<tr>
<th>ID</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>NOUN</td>
</tr>
</tbody>
</table>
| CONCEPT | : ఒక |]
| EXAMPLE | : "ఒకపై మేని నిర్మాణ శిఖరం మేని నిర్మాణ శిఖరం." |
| SYNSET-TELUGU | :: ఒక |

**Tamil**

<table>
<thead>
<tr>
<th>ID</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>NOUN</td>
</tr>
</tbody>
</table>
| CONCEPT | : ஒன்று |]
| EXAMPLE | : "ஒன்று/மேனிய, ஒன்று/மேனிய, ஒன்று/மேனிய, ஒன்று/மேனிய, ஒன்று/மேனிய, ஒன்று/மேனிய." |
| SYNSET-TAMIL | :: ஒன்று |

(c) Situation (3) One to Many:

Here a single lexeme is linked with multiple lexical items.

**Ex: X(sw1/L1 <-> tw1, tw2,tw3)**

**Telugu**

<table>
<thead>
<tr>
<th>ID</th>
<th>12019</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>NOUN</td>
</tr>
</tbody>
</table>
| CONCEPT | : ఒక |]
| EXAMPLE | : "ఒకపై మేని నిర్మాణ శిఖరం మేని నిర్మాణ శిఖరం." |
| SYNSET-TELUGU | :: ఒక |

**Tamil**

<table>
<thead>
<tr>
<th>ID</th>
<th>12019</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>NOUN</td>
</tr>
<tr>
<td>CONCEPT</td>
<td>:: ஒன்று</td>
</tr>
<tr>
<td>EXAMPLE</td>
<td>::</td>
</tr>
<tr>
<td>SYNSET-TAMIL</td>
<td>:: ஒன்று, ஒன்று, ஒன்று</td>
</tr>
</tbody>
</table>
(d) Situation (4) Many to Many:

Here many lexical items are linked with many in the target side.

Ex: \(X(sw1/L1, sw2/L2, sw3/L3, sw4/L4, sw5/L5, sw6/L6, sw7/L7 \leftrightarrow tw1,tw2,tw3,tw4,tw5,tw6)\)

Telugu:

<table>
<thead>
<tr>
<th>ID</th>
<th>12833</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>VERB</td>
</tr>
<tr>
<td>CONCEPT</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>EXAMPLE</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>SYNSET-TELUGU</td>
<td>(\text{తెలుగు} / \text{TAM1}, \text{తెలుగు} / \text{TAM2}, \text{తెలుగు} / \text{TAM3}, \text{தமிழ்} / \text{TAM4}, \text{தமிழ்} / \text{TAM5}, \text{தமிழ்} / \text{TAM6}, \text{தமிழ்} / \text{TAM7})</td>
</tr>
</tbody>
</table>

Tamil:

<table>
<thead>
<tr>
<th>ID</th>
<th>12833</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>VERB</td>
</tr>
<tr>
<td>CONCEPT</td>
<td>குடும்ப</td>
</tr>
<tr>
<td>EXAMPLE</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>SYNSET-TAMIL</td>
<td>தமிழ், குடும்ப, குடும்ப, குடும்ப, குடும்ப, குடும்ப, குடும்ப</td>
</tr>
</tbody>
</table>

The dictionary uses the categories like Nouns, Verbs, Adjectives, Adverbs, Pronouns, Numerals, NST and Indeclinables such as Classifiers, Quantifiers, Interjections, Quotatives, Particles and Conjunctions. Other than this, a whole list of functional words like Case Markers and Tense, Aspectual and Model markers are also included in the dictionary. A bilingual dictionary, which is also a concept based one is used as a stand-by along with the synset dictionary in case of failing.

3. Divergences of Telugu and Tamil from the point of lexicon:

A translation divergence may occur when the underlying concept or "gist" of a sentence is distributed over different words for different languages. According to Dorr (1990), divergences are cross-linguistic distinctions in which the natural translation of one language into another results in a very different that of the original. She proposes seven types of divergent categories comprising of Thematic, Promotional, Demotional, Structural, Conflational, Categorial and Lexical Divergences. This classification of Dorr (1990) on Machine Translation Divergence is taken as a base and is tried to map it with the Telugu-Tamil Machine Translation System. The paper focuses on three divergences due to lexical aspects of the languages involved in translation.

3.1 Conflational Divergence:

It occurs when the sense conveyed by a single word is expressed by two or more words in one of the languages. For instance, Telugu uses 'snAnaM ceVyyi' for 'bath' whereas it is expressed by 'kulYi' in Tamil.
II.a. TEL: nenu snAnaM ceswAnu.
'I bathing do-FUT-1p.sg'

TAM: nAnY kulYippenY.
'I bath-FUT-1p.sg'

ENG: I will bathe.

The Conflational Divergence is mainly carried out by the Multi Word Expression Module. Here the collocative words are given equivalents in the respective language. Multi-word expressions are a set of collocations of words which are often come with a non compositional semantics which otherwise could not be resolved. These forms are sequences of two or more words generally express a co-occurrence meaning. Telugu-Tamil Multi-Word Expression Module is built up with the database which consummates the words of co-occurrence. The root form of the two or more sequences of words are used in the database. For instance, the following expression of Noun (N) and Verb (V) is carried out during the Telugu-Tamil Translation:

1. N N --> N N
   uwwara praxeS, uwwirap pirawecam
   Since 'uwwaraM' in isolation may mean either 'the north' or 'a letter', but in the context of the word indicating the name of a State, it needs to be listed.

2. N V --> N V
   veru ceVyyi, pirivinYE ceVy
   Here the word 'veru' may mean 'separation' and 'root'. But when it is followed by a verb like 'ceVyyi' it means 'to separate'.

3. N N N --> N N N
   calana ciwra pariSrama, wirEp patac cafkam
   Here the cinema is expressed in Telugu as 'calana ciwraM' i.e, 'motion picture' whereas in Tamil it is 'screen picture'.

4. N N --> 0 N
   sahajaM sixXaM, 0 iyarYkE
   The term 'natural product' is expressed by two words in Telugu whereas in Tamil it is one.

5. N V --> 0 V
   vidixi ceVyyi, 0 wafku
   xAdi ceVyyi, 0 wAkku

   In Telugu, the intensifier compounds involve two words to express a single intensified form of the concept denoted by the nouns of temporal/spatial category whereas in Tamil by the corresponding reduplication of the head noun.

   6. REDUP NST --> REDUP NST
      moVtta moVxata, muwanY muwal

As described above lexemes of special cases such as phrases, idioms which are multi word expressions are taken care by the MWE module before the processing enters into the lexical transfer module.
3.2 Categorical Divergence:

Changes in category creates categorical divergence. It is due to the mismatch between the Parts of Speech Categories of the words involved in the pair of Languages. For instance, the word cAla in Telugu is ambiguously used as an adjective as well as an intensifier. However, in Tamil two distinct categories of words are used as shown below:

III.a. TEL: cAlA puswakAlu unYnYAyi.
    'a lot book-pl being-3.p.pl.n'
TAM: nirYEya puwwakafkalY ulYIYanYa.
    'a lot book-pl being-3.p.pl.n'
ENG: A lot of books are there.

III.b. TEL: cAlA eVwwugA uMxi.
    'very high being'
TAM: mika uyaramAka ulYIYawu.
    'very high being'
ENG: It is very high.
Handling Strategy:
(i) Repair Role:
\[
\begin{align*}
[[cAIA<$cat>][N1.gA]] & \Rightarrow [[cAIA<INT>][N1.gA]] \\
[[cAIA<$cat>][N1]] & \Rightarrow [[cAIA<ADJ>][N1]]
\end{align*}
\]

3.3 Lexical Divergence:
It arises when there is a lack of exact lexical equivalent but structure presents a translational equivalence between the language pair. Here, the literal translation of the source language word is substituted by a corresponding translational equivalent to resolve the problem.

For instance,

IV.a. TEL: nAku Iwa vaccu.
'me-DAT swimming come'

TAM: eVnYakku nlccal weVriyum.
'me-DAT swimming know-fut-3p.sg.n'

ENG: I know to swim.

Handling Strategy:
Transfer Rule:
\[
[[N1ku][N20][vaccu]] \Rightarrow [[N1ku][N20][teri<3p.sg.n]]
\]

IV.b. TEL: AmeVku kadupu vacciMxi.
'She-DAT bellyt come-PST-3p.sg.n'

TAM: avalYukku karppam erYpattawu.
'She pregnancy form-PRS-3p.sg.f'

ENG: She became pregnant

In the above example, the idiomatic sense of 'kadupu' is 'pregnant' which means 'the stomach'. But Tamil uses the term 'karppam' to express the same.

4. Conclusion: The Telugu- Tamil Machine Translation system is built by using the concept based dictionaries discussed above. The concept based dictionaries ensure the resolution of much of the disambiguation presented by the words in the lexical substitution in translation. The system is tested continuously by the native speaker of Tamil in order to validate its performance in the translation. The five scale Evaluation method of IL-IL MT is adopted for this purpose. The current comprehension of the outputs fall between 85-90%.

1 Transliteration Scheme using wx-notation:
Tamil Orthography:
\[ a A l l u U e V e E o V o O H \]
\[ k f c F t N w n p m y r l v l Y Y l Y r Y n Y j s h R \]
Telugu Orthography:
\[ a A l l u U q Q e V e E o V o O M H \]
\[ k K g G f c C j F t T d D N w W x X n p P b B m y r Y l Y l Y Y v S R s h \]
References:


EILMT: A Pan-Indian Perspective in Machine Translation

Hemant Darbari,
Executive Director, C-DAC, Pune, darbari@cdac.in
Anuradha Lele,
Group Co-ordinator, C-DAC, Pune, lele@cdac.in
Aparupa Dasgupta,
Team Co-ordinator, C-DAC, Pune, aparupa@cdac.in
Priyanka Jain,
Project Leader, C-DAC, Pune, priyankaj@cdac.in
Sarvanan, Amrita University, sarvanster@gmail.com

Abstract

To cut-across the language barrier and to encourage the language pluralism of morphologically complex languages [Sproat 1991], especially South-Asian languages [Krishnamurti et al. 1986] in India, a consortium mode robust Machine Translation system (MTS) that is able to raise the accuracy of generation is developed jointly by C-DAC, Pune and DIT, GOI. In Natural Language Processing (NLP) and Natural Language Understanding (NLU), Machine Translation plays a vital role in today’s India for any sort of e-language processing and understanding by machine. In each of the quarter of electronic era of a multi-lingual community machine translation, information retrieval or speech processing becomes obligatory. This paper proposes to describe a hybrid based machine translation system from English to Indian languages. This paper also proposes the TAG based memory managed Machine Translation System [Joshi et al. 1981] aligning with other rule based, example based and statistical based Machine Translation System for English-Hindi, English-Urdu, English-Oriya, English-Bangla, English-Marathi and English-Tamil. EILMT has especially been designed to translate in platform independent modules. This is a proposed hybrid based thin-client/thick-server design; where users (clients) of this system use a standard browser to access the translation services of the server. We call this as a Pan-Indian perspective on Machine Translation. In this paper, we will explain the challenges faced and solution drawn at the various levels of architecture, language and linguistic computation. While building the Machine Translation System, we have taken care of the speed and accuracy of syntactically and morphologically diversified languages at modular and phases of EILMT system.

1.0 Introduction to Machine Translation

In present paper, we explain the challenges encountered to cope with the speed and accuracy of syntactically and morphologically diversified languages tested and developed for Machine Translation system based on consortium mode for English to Indian Languages in collaboration with C-DAC, Pune and DIT, Govt. of India. In 1629 the idea of Machine Translation evolved, when Rene Descartes proposed Universal Language. In 1954, the Georgetown experiment (1954) involved fully-automatic translation of sixty Russian sentences into English. In late 1980s, machine translation inclined to statistical models and example based models evolved gradually. And the Machine Translation system like Systran used by AltaVista search engine, METEO used at the Canadian Meteorological Centre, Example-based machine
translation proposed by Makoto Nagao and several other Hybrid based Machine Translation system came into existence. During the year 1990-91, DIT (Department of Information Technology) of Government of India initiated the TDIL (Technology for Development of Indian languages) project to encourage the Indian language processing in the area of IT. The institutions namely, C-DAC, Pune (MANTRA); NCST (now C-DAC, Mumbai; MATRA); IIIT-Hyderabad (Anusaaraka, and SHAkTI) and IIT-Kanpur (Anglabharati) have taken the Machine Translation System from English to Hindi to greater height by developing applications using cutting edge technology.

2.0 Introduction to EILMT

To overcome the language barrier and to encourage the language pluralism of morphologically complex languages [Sproat 1991], especially South-Asian languages [Krishnamurti et al. 1986] in India, a consortium mode robust Machine Translation system (MTS) that is able to raise the accuracy of generation is developed jointly by C-DAC, Pune and DIT, Govt. of India. It is domain specific Machine Translation system from the domain of tourism. This project is developed by 10 consortium institutes: C-DAC, Mumbai, IIIT-Hyderabad, IISc-Bangalore, IIT-Bombay, Jadavpur University – Kolkata, Amrita University – Coimbatore, IIIT-Allahabad, Banasthali Vidyapeeth – Banasthali, Utkal University – Bhubaneshwar and C-DAC, Pune being the consortium leader. EILMT is a hybrid based Machine Translation system with TAG formalism (Tree Adjoining Grammar based MT developed by C-DAC, Pune), SMT (statistical based MT developed by C-DAC, Mumbai), ANALGEN (Rule based MT by IIIT-Hyderabad) and EBMT (Example based system developed by IISc, Bangalore). To measure the performance of aforementioned translation engines and evaluate the language pair wise translation accuracy, we represent here the internal testing carried out by consortium. The translation output accuracy on English-Hindi EILMT system of each of these aforementioned translation engines are given below. Following table data is the average score of engine for each sentence structure type:

<table>
<thead>
<tr>
<th>Sentence Structure type</th>
<th>AnalGen (%)</th>
<th>EBMT (%)</th>
<th>SMT (%)</th>
<th>TAG (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copula</td>
<td>88.75</td>
<td>42.50</td>
<td>66.25</td>
<td>87.50</td>
</tr>
<tr>
<td>Simple</td>
<td>58.75</td>
<td>60.00</td>
<td>57.50</td>
<td>95.00</td>
</tr>
<tr>
<td>Appositional</td>
<td>71.50</td>
<td>47.50</td>
<td>67.50</td>
<td>91.25</td>
</tr>
<tr>
<td>Relative Clause</td>
<td>75.00</td>
<td>53.75</td>
<td>55.00</td>
<td>95.00</td>
</tr>
<tr>
<td>That-Clause</td>
<td>62.50</td>
<td>51.25</td>
<td>60.00</td>
<td>92.50</td>
</tr>
<tr>
<td>Wh-Clause</td>
<td>56.66</td>
<td>38.33</td>
<td>53.75</td>
<td>63.33</td>
</tr>
<tr>
<td>Co-ordinate</td>
<td>65.00</td>
<td>32.50</td>
<td>78.75</td>
<td>95.00</td>
</tr>
<tr>
<td>Conditional</td>
<td>48.50</td>
<td>56.25</td>
<td>63.75</td>
<td>77.50</td>
</tr>
<tr>
<td>PP Initial</td>
<td>60.00</td>
<td>43.75</td>
<td>57.50</td>
<td>93.75</td>
</tr>
<tr>
<td>Adverb Initial</td>
<td>80.00</td>
<td>53.33</td>
<td>53.33</td>
<td>95.00</td>
</tr>
<tr>
<td>Gerundial</td>
<td>36.00</td>
<td>31.60</td>
<td>75.00</td>
<td>83.33</td>
</tr>
<tr>
<td>Participle</td>
<td>81.25</td>
<td>25.00</td>
<td>81.25</td>
<td>91.25</td>
</tr>
<tr>
<td>Infinitive</td>
<td>50.00</td>
<td>46.25</td>
<td>60.00</td>
<td>93.75</td>
</tr>
<tr>
<td>Discourse Connector</td>
<td>70.00</td>
<td>55.00</td>
<td>75.00</td>
<td>75.00</td>
</tr>
</tbody>
</table>

Table 1: Engine wise Translation output accuracy for English -> Hindi pair
Similarly, language pair wise translation accuracy on TAG translation engine was evaluated, whose approximate translation accuracy is as follow: for English-Hindi pair the translation accuracy is approximately 85%; for English-Urdu is approximately 75%; for English-Oriya is approximately 80%; for English-Bangla is approximately 70%; for English-Marathi is approximately 65%; and for English-Tamil is approximately 70%.

3.0 Introduction to EILMT Architecture: The Challenges

EILMT is a web-based Machine Translation system solution with a hybrid approach across six language-pairs from English to Hindi, Urdu, Oriya, Bangla, Marathi and Tamil. Along with four different machine translation engines, the Named Entity Recognizer [NER] and Word Sense Disambiguation [WSD] modules are developed by IIT, Mumbai. EILMT system architecture is represented in the following diagram:

Basic system components of EILMT consortium are: User Log module; Pre-Processing module; four Translation Engines: AnalGen, EBMT, SMT & TAG for six language pairs; Post-Processing module; Collation and Ranking module; a compatible system with W3C; and Browser compatibility for IE, Mozilla, Firefox, Google Chrome, Apple Safari & Opera. (See Annexure 1 B for detailed EILMT system specifications).
3.1 Overall Architecture of EILMT

EILMT is a web based translation system accessed simultaneously with multiple users and requests. JBoss is the application server with robust database I/O file exe and for rapid, transactional, secure and portable application EILMT is supported by EJB (Enterprise Java Bean). EILMT is designed on the line of centralized design where Internet clients submit their documents to a multi-core server where the parsing and generation is a spawning of multi-threaded embedding. Significantly, the outer layer thread connects to ANALGEN engine (implemented in PERL on Linux platform), another thread with SMT with server and the other with EBMT engine. And the Ranking module collates and rank the translation from the above mentioned translation engines.

EILMT system has been tested on multi core (8 core machine) machine for execution that has raised the system processing speed upto three times. Initially NER which is used in SMT system developed by C-DAC, Mumbai followed a Maximum Entropy Based Approach. This system had an accuracy of 81.33% on ConLL-2003 dataset. (Precision: 83.85%, Recall: 78.95% F-Measure: 81.33%). The current system uses two stages: SVMs followed by MEMMs. Using 2 phases, improved the accuracy to 93% (Precision: 92.56% Recall: 93.48% F-Measure: 93%).

3.2 Implementation of TAG Formalism

Tree Adjoining Grammar [Kroch and Joshi, 1985] is implemented for all 6 language-pairs in EILMT on TAG translation engine. The JAVA based TAG parser translates English documents to Hindi, Urdu, Oriya, Bangla, Marathi and Tamil. The significant feature of this parser is incremental parser that identifies the

(a) clause or phrase on the basis of probable declarative clause boundary and,

(b) after identifying clause boundary the TAG tree derivation structure identifies probable parent derivation to the nearest child derivation structure to give the final integrated derivational tree to the TAG Generator.

The TAG engine is enriched in such a way that it can process the parsing and generation for interrogative sentences, negation, gerundial construction, relative clause construction, and past & progressive participle etc. The pre-processing is controlled by supervised modules such as - syntaxic TAG tree disambiguator module with optimized code and database-design written in regular expressions. Consider the following description of the incremental parser that has given modularity, extensionality and speed in the translation process of TAG engine. Probability of adjoining the parent derivations to a nearest probable child derivation is given by the following equation:

\[ Y = c(X) \] ; Where, \( X \) = Number of Child derivations, \( Y \) = Number of Parser Derivations, \( c \) = Combination,

Consider the sentence “The 18th century Bharatpur-Bird-Sanctuary, which is also known as the Keoladeo-Ghana-National Park, is famous as the most important bird breeding and feeding habitat of the world.”
The English corpus of 15,200 sentences from tourism domain were collected, organized, vetted and aligned [Sinclair, J. 1991 and 2004] for all 6 language-pairs. India being a Linguistic Area [see Krishnamurti et al, 1986] in South-Asian sub-continent, both Indo-Aryan (eastern and western Indo-Aryan) and Dravidian language families with rich morphological heritage have its separate distinct linguistic identity at source-target TAG grammar, transfer grammar (a source-target link grammar), rule-normalizer, rules for morphological analysis and synthesis and transliteration and typing-tool rule. The stylistic trend observed in EILMT tourism corpus is: simple sentence (14.94% frequency of occurrence), copulative construction (3.49% frequency of occurrence), co-ordinate sentences (20.10% frequency of occurrence), appositional sentences (11.33% frequency of occurrence), various declarative clause structures (22% frequency of occurrence), gerundial constructions (.35% frequency of occurrence), conditional sentences (1% frequency of occurrence), discourse connector (0.77% frequency of occurrence) and infinitival sentences (9.03% frequency of occurrence). Thus, the parallel corpus created for all 6 language-pairs and features such as intelligibility, comprehensibility and fluency in translation are maintained to set a reference to the machine output, as E.M. Enquest has said very correctly that “Proper words in proper places creates styles”.

In Natural Language Processing (NLP) and Natural Language Understanding (NLU) [Terry Patten, 1985], Machine Translation plays a vital role in Indian sub-context for e-language processing by machine. In Eng-Hindi EILMT system, the localization of linguistic peculiarities of Hindi such as oblique formation, ergativity, marked-gender system, case-marking, direct-oblique pluralization etc. are handled in a controlled environment through morph-synthesizer, finite and non-finite generators, POS conversion rule...
etc. Similarly, for other Indo-Aryan language pairs i.e., Eng-Urdu, Eng-Oriya, Eng-Bangla and Eng-Marathi the linguistic features such as, Perso-Arabic and Indic pluralization system, lexico-semantic peculiarities, copula drop, dropping of existential subject, post-position synthesis, synthesis of case-marking, emphatic clitic formation, usage of classifier, verb root alteration, strong and three level gender system, gender based noun synthesis, and compounding etc., [Bhattacharya, T et. al 1996; Krishnamurti, Bh. et al 1986; Selkirk 1982; Williams 1981] are incorporated through feature-based lexicon, ordered rule-based normalizer etc. Approximately 37,000 bilingual lexicon, 2000 phrasal lexicon, 97 TAG tree disambiguation rule, 125-150 source TAG trees, 215-230 target TAG trees, 800 transfer grammar mapping and 70-75 morph-synthesis rule are developed for each Indo-Aryan language-pairs. Following section will explain the linguistic challenges faced and language computing solutions drawn in EILMT system for syntactically and morphologically diversified and complex languages to raise the speed and translation accuracy:

4.1 Raising Translation Speed and Accuracy: Intermediate Solutions

a) Rectifying wrong POS tagging of Stanford tagger (version 1.6) through rule based POS tagging (See Computational Linguistics, volume 19, number 2, pp313-330.). Consider the following examples that states the internal POS conversion rule that rectifies the erroneous tagging output of Stanford tagger,

“Visit the Sheesh-Mahal or the Hall of Victory glittering with mirrors and ascend the Fort on elephant's back”

Stanford Output: [Visit@@NN, the@@DT, Sheesh@@NNP, Mahal@@NNP, or@@CC, the@@DT, Hall@@NNP, of@@IN, Victory@@NNP, glittering@@VBG, with@@IN, mirrors@@VBZ, and@@CC, ascend@@VB, the@@DT, Fort@@NNP, on@@IN, elephant@@NN, zxt@zxtdzxt@zxt@zxt@NN, back@@RB]

Internal Pos Category String: [Visit@@VERB the-Sheesh-Mahal@@NOUN or@@CONJ the-Hall@@ NOUN of@@PREP Victory@@NOUN glittering@@PrPART with@@PREP mirrors@@NOUN and@@ CONJ back@@ADV ascend@@TYPE_APPOINT the-Fort@@NOUN on@@PREP elephant@@NOUN zxt@zxtdzxt@zxt@AS]

Apart from POS tagging, emotion and sense tagging is necessary in Machine Translation to capture the semantic anomaly of the natural language.

b) Chunking is an important part of shallow parsing level. It minimizes the number of tokens to be sent to the core parser, thus reducing the number of possible adjunctions and effected the translation time as well as the translation quality. We perform noun phrase chunking and verb group collation. Consider the following example of Chunking at level-1stage,

[The-Prince/NNP of/IN Wales/NNP Museum]/NNP ,/, [the-Jahangir-art-Gallery]/NNP ,/, [the-various-churches]/NNS ,/, temples/NNS and/CC shrines/NNS including/VBG [the/DT one/CD]/NNP of/IN [Haji-Ali]/NNP out/IN on/IN [an-island]/NN linked/VBN by/IN [a-causeway]/NN ,/, are/VBP worth/JJ [a-glimpse]/NN

Chunking at the level-2 stage:

[The-Prince-of-Wales-Museum]/NNP ,/, [the-Jahangir-Art-Gallery]/NNP ,/, [the-various- churches]/NNS ,/, temples/NNS and/CC shrines/NNS including/VBG [the-one]/NNP of/IN [Haji-Ali]/NNP
out/IN on/IN [an-island]/NN linked/VBN by/IN [a-causeway]/NN ,/, are/VBP worth/JJ [a-glimpse]/NN

c) We use a TAG (Tree Adjoining Grammar) [Joshi et al., 1975] parser, and for that we have created a number of trees to represent structure of source and target languages. In this formalism each token is tagged with a POS tag/category, on the basis of which a set of possible tree tags are assigned to the token. This process is called tree tagging. A sentence as a string of tree-tagged tokens, are then sent to the parser. When a token in a sentence is tagged with a number of trees the parser is liable to produce multiple derivations, most of them being inappropriate. This reduces accuracy and speed. To eliminate this spurious derivation, or at least minimize them, we adopted the technique of TAG tree pruning. Our pruning module further disambiguates according to the syntactic context and helps in selection of TAG tree in more precise way. Accuracy and speed of the system, thus, was substantially improved.

d) To handle the synthesis of constructions in Indian Languages, morphological complexities of nouns and verbs and their inter-relationships, and the kaaraka formalism [Gangopadhyay, M. 1990] in a defined context plays a major role in noun or verb synthesis. Henceforth various categories at adjoining positions as an adpositional words like adjectives, post-positions (parasargas), and various particles like the avyayas etc. are also within this defined context. Basically post-positions, avyayas have modified-modifier [Aronoff, M. 1976] function adding more linguistic information to the end-users of the target language. The feature embedded morphological rules (and also sometimes gender agreement) written for the synthesizer can be seen through the synthesized output. Verbs in the language demand the karaka identities and the nouns fulfil the demands according to the yogyataa. And, in a defined context, nouns demand parasargas or post-position on a semantic account. Following diagram explains the synthesis process in EILMT system:

![Diagram 4: Morph-synthesis process of EILMT system](image)

Above mentioned points from 5 a) to d), the linguistic variations and complexities that are handled through pre-processing or post-processing generative modules have escalated translation accuracy and speed in a considerable way. Following graph represents the comparison of translation speed between old and new version of EILMT system (i.e., speed of translation before and after pruning and context disambiguation of the POS tagsets, TAG tree tagging and noun-verb synthesis). The above development at parsing and generation stages has raised the speed of translation in the latter (or new) version of EILMT:
5.0 English-Tamil EILMT System: an Overview

In English-Tamil EILMT system, special attention to Tamil morphological system has been given. As Tamil roots to Dravidian language family, being agglutinative language, the synthesis of finite and non-finite forms, synthesis of noun or noun group and gender based system has been catered through feature based lexicon, and noun and verb morph-synthesizer. In modern Tamil three types of words – noun, verb and itaiccol or particles are found. The noun indicates animate and inanimate categories (tInai, is classified into uyartiNai and akRiNai). There are three genders in Tamil - masculine and feminine and neuter where masculine and feminine indicates singular number and neuter gender indicates plural number.

There are three persons in Tamil (first, second and third person). Case inflexion is prominent with suffixes in Tamil. Tamil being agglutinative in nature [See Varadarajan, Mu. 1988] is found to be different to parse and generate than the manner in which Indo-Aryan languages are generating in EILMT system. Approximately 35,000 bilingual lexicon, 92 phrasal lexicon, 97 TAG tree disambiguation rule, 125 source TAG trees, 127 target TAG trees, 147 transfer grammar mapping and 100 morph-synthesis rule are developed for English-Tamil version. Consider the following example from tourism domain with EILMT TAG output in Tamil:

**English:** Mother Earth' is kind in return. / **Tamil:** பெப்பர்க்குட்போக்கு பொருட்களை வைத்து வெள்ளி பெற்று கைக்கொண்டு
Following diagram represents the English-Tamil User Interface (with Tamil output):

Diagram 6: English-Tamil User Interface output

5.1 Translation Accuracy of English-Tamil EILMT System

To evaluate the translation accuracy of English-Tamil system the score was evaluated through Subjective/Human Evaluation. The parameters for testing the translation accuracy of EILMT system for subjective/human evaluation are: POS tagging, P-Syntax, G-Syntax, Morph-Synthesis, Lexicon availability and phrase marking. We represent here the internal testing carried out by consortium on the test-report provided by the Testing Agency on EILMT alpha version 5.1 in the following bar-chart. [See Appendix I for English-Tamil output].

Diagram 7: Bar-Chart of Eng-Tamil translation output Evaluation
5.2 Scope of improvement for Eng-Tamil TAG EILMT system

Future improvement on English-Tamil EILMT system on the basis of development done for alpha version 5.1 is as follows:

a) Re-framing and enhancing the Noun Collation module on the basis of Phrase Tagging and agglutinating character of Tamil.

b) The process of new Tree set creation for the following structures: interrogative, imperative, negative sentences, handling of objects other than adverbial-synthesis.

c) Enhancement of feature based linguistic rule-set in synthesis of synonym, the verb generator module for all the tenses and Bilingual lexicon correction.

6.0 Conclusion

All these above findings, research and implementations to EILMT system give a more productive and evolutionary ground. And this ground will definitely raise some critical questioning not only on machine translation but also text mining, data pruning, information extraction and retrieval, speech technology and in IL to IL information exchange and access. Thus the research and study on EILMT for Indian languages should be guided and formalized as following:

a) Standardization of Indian tagset and considering the factor of morphologically rich language families and formal tagging, sense tagging and emotion tagging of the e-corpora available in Indian languages.

b) Memory based parsing management to organize the multiple language with multiple domain. Further, memory managed MT will increase the system efficiency 15-20% more.

c) And, feature-and-morphology based modules for morphologically rich Indian languages so that scope of this analysis and synthesis can be extended for reverse translation as-well.

7.0 Reference


ANNEXURE I

(Evaluation of EILMT system Version 5.1 Translation output for English-Tamil language pair)

<table>
<thead>
<tr>
<th>English Sentence</th>
<th>Translated output (E-T)</th>
<th>Analysis of Translation</th>
<th>Type of Structure</th>
</tr>
</thead>
</table>
| The Ganga Golden Jubilee Museum has a large collection of pottery, paintings,   | கஞ்சா கலோரணம் துறையில் பல்வேறு பொருட்கள், மர்ப்புற்ற வைப்பை, தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், கூல்லான்கள், பல்லுயர் வைப்பை, பல்குட்டுகள், பல்குட்டுகள் போன்றவை பல்வேறு பொருட்கள். | POS= 100%  
P syntax = 100%  
G-syntax= 100%  
Lexicon= 100%  
Phrase Marking= 100%  
Synthesizer= 100% | Simple + Copula (Possessive form)                                                     |
| The shops are full with colorful items which include handicraft items, precious | வைப்பைகளாக்கிகள், மர்ப்புற்ற வைப்பை, தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், பல்குட்டுகள், பல்குட்டுகள், தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், பல்குட்டுகள் போன்றவை பல்வேறு பொருட்கள். | POS= 100%  
P syntax = 100%  
G-syntax= 100%  
Lexicon= 100%  
Phrase Marking= 100%  
Synthesizer= 100% | Relative Clause (subordinate clause)                                                  |
| Balsamand Lake & Palace, an artificial lake, is a splendid spot and was built in | பல்லுயர் மாவை, பொருள்வற்ற மாவை, தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், பல்குட்டுகள், பல்குட்டுகள், தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், பல்குட்டுகள் போன்றவை பல்வேறு பொருட்கள். | POS= 100%  
P syntax = 100%  
G-syntax= 100%  
Lexicon=95%  
Phrase Marking= 100%  
Synthesizer= 100% | Appositional(complement + initial)                                                    |
| The picturesque Kangra valley has several spots that offer mahaseer river carp. | பல்லுயர் மாவை, பொருள்வற்ற மாவை, தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், பல்குட்டுகள், பல்குட்டுகள், தண்டாண்டனாக்காக்கிகள், பல்குட்டுகள், பல்குட்டுகள் போன்றவை பல்வேறு பொருட்கள். | POS= 100%  
P syntax = 100%  
G-syntax= 100%  
Lexicon=80%  
Phrase Marking= 100%  
Synthesizer= 100% | (Co-ordinate) + That Complement                                                      |
Visitors found the villagers dancing to the tune of folk music.

| Visitors found the villagers dancing to the tune of folk music. |
| POS= 100% |
| P syntax = 100% |
| G-syntax= 100% |
| Lexicon= 80% |
| Phrase Marking= 100% |
| Synthesizer= 100% |

Udaipur is known for its beautiful lakes, well structured palaces, lush green gardens and temples but the major attractions of this place are the Lake Palace and the City Palace.

| Udaipur is known for its beautiful lakes, well structured palaces, lush green gardens and temples but the major attractions of this place are the Lake Palace and the City Palace. |
| POS= 80% |
| P syntax = 70% |
| G-syntax= 70% |
| Lexicon= 70% |
| Phrase Marking= 100% |
| Synthesizer= 70% |

Unless you are accustomed to horse riding, a daylong camel ride will be tiring.

| Unless you are accustomed to horse riding, a daylong camel ride will be tiring. |
| POS= 80% |
| P syntax = 70% |
| G-syntax= 70% |
| Lexicon= 70% |
| Phrase Marking= 100% |
| Synthesizer= 70% |

Today, Mumbai is the country's financial and cultural centre, it is also home to a thriving film industry.

| Today, Mumbai is the country's financial and cultural centre, it is also home to a thriving film industry. |
| POS= 100% |
| P syntax = 100% |
| G-syntax= 100% |
| Lexicon= 80% |
| Phrase Marking= 100% |
| Synthesizer= 90% |

Soon enough, prince Jahangir was born to his Hindu wife Jodha Bai.

| Soon enough, prince Jahangir was born to his Hindu wife Jodha Bai. |
| POS= 100% |
| P syntax = 90% |
| G-syntax= 90% |
| Lexicon= 80% |
| Phrase Marking= 100% |
| Synthesizer= 100% |
10

கலாச்சாரின் சுற்றுக்கு செல்லுதல்
துறையில் ஆங்கிலத்தில் கருத்துடைய
பதில் விளக்கப்பட்டது அளவில்

அரசு, ஆங்கில பதிவு பி.மெ.சி., மாநிலச்
Email: amaithianantha@yahoo.co.in

அழைக்கும் எண்ணிக்கை, கல்லூரியை வளமன்
அழைக்கும் எண்ணிக்கை, துவாரம் இரண்டாம் இரண்டு,

கல்லூரியை அல்லது எண்ணிக்கை அளவில் (inci) குறியை, திறந்துரைகளுக்காக, நோக்குரை நோக்கு

முக்க தொடர்நாள், 2/7, 11 மாத தொடர், ராஜ்நிதிக், அறிவுபெற்றவர் நிறுவனமான - 600088.

முக்கோளில், துறை குறிப்பிடுவதற்காக, முக்கோ தொடர்நாள் நோக்கு அறிக்குரை (finger pressure) பதிவு பிடிக்கும், நோக்கு அறிக்குரை குறிப்பிடும் (finger pressure) குறிப்பிட்டு குறிப்பிட்டு ஒரு குறிப்பிட்டு, துறையில் அளவில் குறிப்பிட்டு பதிவு பிடிக்கும். குறிப்பிட்டு துறை

முக்கேற்குருக்கு நோக்கு அறிக்குரை (finger pressure) குறிப்பிட்டு குறிப்பிட்டு ஒரு குறிப்பிட்டு, துறையில் அளவில் குறிப்பிட்டு பதிவு பிடிக்கும். துறை

முக்கேற்குருக்கு நோக்கு அறிக்குரை (finger pressure) குறிப்பிட்டு குறிப்பிட்டு, துறையில் அளவில் குறிப்பிட்டு பதிவு 


1. பதில் விளக்கப்பட்டது அளவில் (ABCDEFG விளக்கப்பட்டது – தொடர்நாள்)

<table>
<thead>
<tr>
<th>ABCDEFGHIJ</th>
<th>10</th>
<th>t</th>
<th>தொடர்நாள் தொடர்</th>
</tr>
</thead>
<tbody>
<tr>
<td>abcdedghij</td>
<td>10</td>
<td>t</td>
<td>தொடர்நாள் தொடர்</td>
</tr>
<tr>
<td>KLMNOPQRS</td>
<td>9</td>
<td>t</td>
<td>தொடர்நாள் தொடர்</td>
</tr>
<tr>
<td>Klmnopqrs</td>
<td>9</td>
<td>t</td>
<td>தொடர்நாள் தொடர்</td>
</tr>
<tr>
<td>TUVWXYZ</td>
<td>7</td>
<td>t</td>
<td>தொடர்நாள் தொடர்</td>
</tr>
<tr>
<td>Tuvwxyz</td>
<td>7</td>
<td>t</td>
<td>தொடர்நாள் தொடர்</td>
</tr>
<tr>
<td>இலக்கக்கூடர்</td>
<td>52</td>
<td>அறிவுநுக்குரை</td>
<td></td>
</tr>
</tbody>
</table>
1. புள்ளியியல் விளிம்புண்ணக அடி (QWERT விளிம்புண்ணக – கிரி)

<table>
<thead>
<tr>
<th>விளிம்புண்ணக அடி</th>
<th>புள்ளியியல் விளிம்புண்ணக அடி</th>
<th>2</th>
<th>விளிம்புண்ணக அடி (தரும்)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUERTYUIOP</td>
<td>Qwertyuiope</td>
<td>10</td>
<td>குணாகராக வரும்</td>
</tr>
<tr>
<td>ADFGJKL</td>
<td>asdfghjkl</td>
<td>9</td>
<td>இரண்டு வரும்</td>
</tr>
<tr>
<td>ZXCVBNM</td>
<td>zxcvbnm</td>
<td>7</td>
<td>52 வரும்</td>
</tr>
<tr>
<td>கிரியூரூர்</td>
<td>கிரியூரூர்</td>
<td>52</td>
<td>வரும்</td>
</tr>
</tbody>
</table>

2. புள்ளியியல் விளிம்புண்ணக அடி (களாங்கோ விளிம்புண்ணக – கிரி)

<table>
<thead>
<tr>
<th>விளிம்புண்ணக அடி</th>
<th>புள்ளியியல் விளிம்புண்ணக அடி</th>
<th>2</th>
<th>விளிம்புண்ணக அடி (தரும்)</th>
</tr>
</thead>
<tbody>
<tr>
<td>கை எது எம் எம்</td>
<td>கை எது எம் எம்</td>
<td>12</td>
<td>குணாகராக வரும்</td>
</tr>
<tr>
<td>மெம்</td>
<td>மெம்</td>
<td>12</td>
<td>இரண்டு வரும்</td>
</tr>
<tr>
<td>எம் எம் எம்</td>
<td>எம் எம் எம்</td>
<td>9</td>
<td>வரும்</td>
</tr>
<tr>
<td>புருஷ யார் யார்</td>
<td>புருஷ யார் யார்</td>
<td>35</td>
<td>வரும்</td>
</tr>
</tbody>
</table>

3. கனவியல் விளிம்புண்ணக அடி (களாங்கோ 99 விளிம்புண்ணக – கிரி)

<table>
<thead>
<tr>
<th>விளிம்புண்ணக அடி</th>
<th>புள்ளியியல் விளிம்புண்ணக அடி</th>
<th>2</th>
<th>விளிம்புண்ணக அடி (தரும்)</th>
</tr>
</thead>
<tbody>
<tr>
<td>கை எது எம் எம்</td>
<td>கை எது எம் எம்</td>
<td>12</td>
<td>குணாகராக வரும்</td>
</tr>
<tr>
<td>மெம்</td>
<td>மெம்</td>
<td>1</td>
<td>இரண்டு வரும்</td>
</tr>
<tr>
<td>எம் எம் எம்</td>
<td>எம் எம் எம்</td>
<td>11</td>
<td>வரும்</td>
</tr>
<tr>
<td>புருஷ யார் யார்</td>
<td>புருஷ யார் யார்</td>
<td>8</td>
<td>வரும்</td>
</tr>
<tr>
<td>புருஷ யார் யார்</td>
<td>புருஷ யார் யார்</td>
<td>35</td>
<td>வரும்</td>
</tr>
</tbody>
</table>
அறா சிற்று பின்பிட்டு எவ்வாறு குறிக்கவும்: “அ” கலை அறாம்பகிராம அறாம் குறிக்க “அ” நிலையிலும் முறை முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதக்கால புதிய முதல்வாதப் புதிய முதல்வாதப் புதிய முதல்வாதப்...
<table>
<thead>
<tr>
<th>நிறுவனம்</th>
<th>தகவல்</th>
<th>காலம்</th>
<th>பணிகர்த்தைகள்</th>
<th>மறைவு நேரம்</th>
<th>பணம்</th>
<th>புணர்பயிற்சியானது</th>
<th>புணர்பயிற்சியானது</th>
</tr>
</thead>
<tbody>
<tr>
<td>கோவை நாட்டுக்குழும்</td>
<td>குழுவாள்</td>
<td>குழுவின் குழுவாள்</td>
<td>14</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>குழுவின் குழுவாள்</td>
<td>14</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ஆண்டு கல்லூரிகள்</td>
<td>குழுவாள்</td>
<td>A முதல் J வரை 10 முதல் 20 வரை, பொது பாடசைகள்</td>
<td>12</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>A முதல் J வரை 10 முதல் 20 வரை, பொது பாடசைகள்</td>
<td>12</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>குறிப்பிட்டு நேர் கல்லூரிகள்</td>
<td>குழுவாள்</td>
<td>K முதல் S வரை 9 முதல் 19 வரை, பொது பாடசைகள்</td>
<td>11</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>K முதல் S வரை 9 முதல் 19 வரை, பொது பாடசைகள்</td>
<td>11</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>நேர் கல்லூரிகள்</td>
<td>குழுவாள்</td>
<td>T முதல் Z வரை 7 முதல் 17 வரை, பொது பாடசைகள்</td>
<td>10</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>T முதல் Z வரை 7 முதல் 17 வரை, பொது பாடசைகள்</td>
<td>10</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ஆ = ஆழ்வாசம், த = தரக்கு. நல்குறிக்காண்ட 94 நேர்கல்லூரிகள் முதல் 14 வரை பொது பாடசை கிளைகள்

<table>
<thead>
<tr>
<th>நிறுவனம்</th>
<th>தகவல்</th>
<th>காலம்</th>
<th>பணிகர்த்தைகள்</th>
<th>மறைவு நேரம்</th>
<th>பணம்</th>
<th>புணர்பயிற்சியானது</th>
<th>புணர்பயிற்சியானது</th>
</tr>
</thead>
<tbody>
<tr>
<td>கோவை நாட்டுக்குழும்</td>
<td>குழுவாள்</td>
<td>குழுவின் குழுவாள்</td>
<td>14</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>குழுவின் குழுவாள்</td>
<td>14</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ஆண்டு கல்லூரிகள்</td>
<td>குழுவாள்</td>
<td>முதல் 12 - வ, அதுக்குநேரக்கு</td>
<td>12</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>முதல் 12 - வ, அதுக்குநேரக்கு</td>
<td>12</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>குறிப்பிட்டு நேர் கல்லூரிகள்</td>
<td>குழுவாள்</td>
<td>பொதுவான வரியாளர் “உ, வ, த, ப, ா” முதல் வரையாளர், அதிகாரி,</td>
<td>14</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>தரக்கு தரக்கு</td>
<td>குழுவாள்</td>
<td>பொதுவான வரியாளர் “உ, வ, த, ப, ா” முதல் வரையாளர், அதிகாரி</td>
<td>14</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>முறைப்படி</td>
<td>முதற் வரியாளர்</td>
<td>முதலாம் வரியாளர்</td>
<td>முறையே</td>
<td>மறுமுறையே</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| துடுவ தெளிவட்டம்
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>நாளாம்</td>
</tr>
<tr>
<td>14</td>
</tr>
</tbody>
</table>

| பின்னர்
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>நாளாம்</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

| ராமர்
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>நாளாம்</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

627
1. புதிய மையில் விளையாட்டுகள் அடையும் - அடையாக்க / QWERT

<table>
<thead>
<tr>
<th>குறுங்கள்</th>
<th>குறுங்கள்</th>
<th>வாயில் தளப்பட வேளை</th>
<th>வரையறுக்கப்பட்ட வேளை</th>
<th>வேறுபாடு</th>
<th>கணினி நாள்</th>
</tr>
</thead>
<tbody>
<tr>
<td>இழுவரிகள்</td>
<td>இழுவரிகள்</td>
<td>வரையறுக்கப்பட்ட வேளை</td>
<td>வரையறுக்கப்பட்ட வேளை</td>
<td>வேறுபாடு</td>
<td>கணினி நாள்</td>
</tr>
<tr>
<td>இடம்பெறும்</td>
<td>இடம்பெறும்</td>
<td>வரையறுக்கப்பட்ட வேளை</td>
<td>வரையறுக்கப்பட்ட வேளை</td>
<td>வேறுபாடு</td>
<td>கணினி நாள்</td>
</tr>
<tr>
<td>குறுங்கள்</td>
<td>குறுங்கள்</td>
<td>வரையறுக்கப்பட்ட வேளை</td>
<td>வரையறுக்கப்பட்ட வேளை</td>
<td>வேறுபாடு</td>
<td>கணினி நாள்</td>
</tr>
</tbody>
</table>

10 94

10 84

10 74

10 63
2. புத்தாண்டு முல்லைக்கொண்டால் அதையும் – QWERT / QWERT
<table>
<thead>
<tr>
<th>பாடல் பக்தம்</th>
<th>பாடல் விளயுவுத் தொடர் தொகுப்பு</th>
<th>மின்னலால் விளயுவுத் தொடர் தொகுப்பு</th>
<th>மீண்டுக்கோருதல்</th>
</tr>
</thead>
<tbody>
<tr>
<td>பாடல் பக்தம்</td>
<td>பாடல் விளயுவுத் தொடர் தொகுப்பு</td>
<td>மின்னலால் விளயுவுத் தொடர் தொகுப்பு</td>
<td>மீண்டுக்கோருதல்</td>
</tr>
<tr>
<td>3. 3. 3. 3. 3. 3. 3. 3. 3. 3. 3.</td>
<td>11 – 37</td>
<td>10 – 94</td>
<td>11 – 37</td>
</tr>
</tbody>
</table>

3. மின்னலால் விளயுவுத் தொடர் தொகுப்பு - குறிப்பிட்டு குறிப்பிட்டு 99
<table>
<thead>
<tr>
<th>புத்துச்செல்வு</th>
<th>வகையில்</th>
<th>குறிப்பிட்டிடும்</th>
<th>முறையீடு குறிப்பிட்டிடும்</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 – 61</td>
<td>14 – 1 4</td>
<td>12 – 73</td>
<td>12 – 26</td>
</tr>
<tr>
<td>11 – 84</td>
<td>11 – 37</td>
<td>10 – 94</td>
<td>10 – 47</td>
</tr>
</tbody>
</table>

கணினியியல் - அறிவியல்
(பட்டியல் 99 - கணினியியல் விளையாட்டக்குறிப்பிட்டிடும்)

கணினியியல் - அறிவியல் - இந்துரை பாதுகாக்குமாறு, தொகுப்பு போல, பொறியியல் போரும், பொறியியல் போரும், பொறியியல் போரும், ஒலிஸ்கைநியம் ஒலிஸ்கைநியம் ஒலிஸ்கைநியம் ஒலிஸ்கைநியம் ஒலிஸ்கைநியம் (All Computer Operating Systems like Linux, Microsoft, Macintosh), போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரும், போரு...
(1) சூடுகிறது சொக்கி (spell checker), பாக்கொம்பெடுத்து சொக்கி (grammar checker) ஆகியவற்றின் மூலக் ஒளிப்படும் வழியில் வசதிக்கொள்ள பட்டியல் உசையுடன் மிக்கது. அத்துடன், சூடுகிறது சொக்கி (spell checker), பாக்கொம்பெடுத்து சொக்கி (grammar checker) ஆகியவற்றின் “இருப்பு ஒளிப்படும்” (open source) பட்டியல் சொக்கிகளின் சொக்கியனின் பயனைப் பாக்கிப்படுத்த வேண்டியது என்பது. 

(2) மூன்று கட்டடக்கல்லுக்குக் கடுமையான அசையகத்தை பிரித்துதல் புருத்தப்பட்டது. இதுத் தொன்நமது கண்டுபிடித்து வரும் சொல்லிருக்கும் இடைவெளியில் அத்துறையான சிறுவர்களுக்கு அறிய வேண்டிய விளக்கம் செய்யப்பட்டது. (Unicode Consortium). உள்ளே கண்டுபிடிக்கப்பட்டது. அத்துடன், ஒவ்வொருவரும் கண்டுபிடித்து எழுத்திற்கு குறிப்பிட்டது. இந்த கண்டுபிடிக்கும் விளக்கம் என்றவற்றில் அறியவும் அறியவும் (symbols) செய்யப் பயன்படுத்தப் பயிற்சிக்கு ஆக்கமுதல் (Unicode Consortium) உள்ளே கண்டுபிடிக்கப்பட்டது.

தக்கசின்: (1) அரக்காய – அரக்க, மீனாராயச்சு அரக்காய

ஒருங்கிணி – அரக்க (தொகுப்பு 99 – இடைவெளியில் விளக்கம்)

<table>
<thead>
<tr>
<th>தொகுப்பு</th>
<th>ஏண்ணும்</th>
<th>விளக்கம்</th>
<th>பிரித்தல்</th>
<th>இழையல்</th>
</tr>
</thead>
<tbody>
<tr>
<td>பிரித்தல் நூற்றாண்டு</td>
<td>மலர் வனவிதம் பல்வேறு பார்வை பாதுகாப்பு அமாவாசை போன்றவை அவாகையாகவே யாரும் மாற்றாக நூற்றாண்டு முழுந்தையும் அரக்கின் அறிக்கைகள்</td>
<td>14</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>பிரித்தல் கருத்து</td>
<td>பட்டியல் வனவிதம் “வல்லு”, மலர் வனவிதம் “ல்லா”, அமாவாசை போன்றவை அவாகையாகவே யாரும் மாற்றாக</td>
<td>14</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>பிரித்தல் வல்லு</td>
<td>தொன்றுள்ள வனவிதம் 12 - ம் அரக்காயச்சு</td>
<td>12</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>பிரித்தல் புனையல்</td>
<td>முன்னூர்த்தம் வருவது “வல்லு, ம, ச, ப, எரு”, அவாகையாகவே அரக்க வேறுபாடு, அவாகையாகவே புனையல் மூலம் அரக்க அரக்காயச்சு</td>
<td>12</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>இருபது பாகங்கள்</td>
<td>குறிப்பிட்டு சொல்லும் படைப்புகள்</td>
<td>உச்சரிக்கை, மண்காள், ஆல்பை, குடநெல், தேர் இக்குறிப்பிட்டு படைப்புகள் “பதா, மூலை மண்காள் அதுபோல், இரண்டு பாரி ஒருவருடன் என்ன அடங்கிருந்தது, ஒருவரை பாரி ஒருவருடன், என்றாலும் கால்வரை கூறுவில்லை. அனுபவத்துடன் 11 63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------</td>
<td>----------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ஒழுங்க பாகங்கள்</td>
<td>குறிப்பிட்டு சொல்லும் படைப்புகள்</td>
<td>உச்சரிக்கை, மண்காள், ஆல்பை, குடநெல், தேர் இக்குறிப்பிட்டு படைப்புகள் “பதா, மூலை மண்காள் அதுபோல், இரண்டு பாரி ஒருவருடன் என்ன அடங்கிருந்தது, ஒருவரை பாரி ஒருவருடன், என்றாலும் கால்வரை கூறுவில்லை. அனுபவத்துடன் 11 74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ரெபால் பாகங்கள்</td>
<td>குறிப்பிட்டு சொல்லும் படைப்புகள்</td>
<td>உச்சரிக்கை, மண்காள், ஆல்பை, குடநெல், தேர் இக்குறிப்பிட்டு படைப்புகள் “பதா, மூலை மண்காள் அதுபோல், இரண்டு பாரி ஒருவருடன் என்ன அடங்கிருந்தது, ஒருவரை பாரி ஒருவருடன், என்றாலும் கால்வரை கூறுவில்லை. அனுபவத்துடன் 10 84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>முன்னூர் பாகங்கள்</td>
<td>குறிப்பிட்டு சொல்லும் படைப்புகள்</td>
<td>உச்சரிக்கை, மண்காள், ஆல்பை, குடநெல், தேர் இக்குறிப்பிட்டு படைப்புகள் “பதா, மூலை மண்காள் அதுபோல், இரண்டு பாரி ஒருவருடன் என்ன அடங்கிருந்தது, ஒருவரை பாரி ஒருவருடன், என்றாலும் கால்வரை கூறுவில்லை. அனுபவத்துடன் 10 94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Free Tamil Keyboard Interface for Business and Personal Use

Panmozhi Vaayil - A Multilingual Indic Keyboard Interface
Abhinava Shivakumar, Akshay Rao, Arun S, A. G. Ramakrishnan
MILE Lab, Department of Electrical Engineering, Indian Institute of Science
Abhinav.zoso, u.akshay, sarun87, agrkrish@gmail.com

Abstract
A multilingual indic keyboard interface is an Input Method [1] that can be used to input text in various Indic languages like Tamil, Kannada, Telugu, Hindi, Gujarati, Marathi and Bengali. The input can follow the phonetic style making use of the standard QWERTY layout along with support for popular keyboard and typewriter layouts [1] (also known as soft layouts) of Indic languages using overlays. Indic-keyboards provides a simple and clean interface supporting multiple languages and multiple styles of input working on multiple platforms. XML based processing makes it possible to add new layouts or new languages on the fly. These features, along with the provision to change key maps in real time make this input method suitable for most, if not all text editing purposes. Since Unicode is used to represent text, the input method works with most applications. This is available for free download and free use by individuals or commercial organizations, on code.google.com under Apache 2.0 license.

Keywords-Input Method, IM, Indic, Localization, Internationalization (i18n), FOSS, Unicode, panmozhi vaayil, vishwa vaangnukha

Introduction
Input method editors or IMEs provide a way in which text can be input in a desired language. Traditionally, IMEs are used to input text in a language other than English. Latin based languages (English, German, French, Spanish, etc.) are represented by the combination of a limited set of characters. Because this set is relatively small, most languages have a one-to-one correspondence of a single character in the set to a given key on a keyboard. When it comes to East Asian languages (Chinese, Japanese, Korean, Vietnamese etc.) and Indic languages (Tamil, Hindi, Kannada, Bangla etc.), the number of key strokes to represent an akshara can be more than one, which makes using one-to-one character to key mapping impractical. To allow for users to input these characters, several input methods have been devised to create Input Method Editors. The term input method generally refers to a particular way to use the keyboard to input a particular language. The term input method editor refers to the actual program that allows an input method to be used.

Objective
The focus has been to develop a multilingual input method editor for Indic languages. The interface should be minimalistic in nature providing options to configure and select various language layouts. Configurability is inclusive of addition of new layouts or languages and option to enable or disable the . Inputs can be based on popular keyboard layouts or using a phonetic style [2]. We call it Panmozhi
Vaayil in Tamil and Vishwa Vaangmukha: in Sanskrit, both meaning entrance for many languages. It is known by the generic name, Indic Keyboard IM in English and is available for download from http://code.google.com/p/indic-keyboards/

**Motivation**

Some of the main reasons for developing indic-keyboards are as follows:
1. To ease inputting of any Indian language under any platform.
2. To facilitate increased use and presence of Indian languages on the computer and internet.
3. To provide a free with an unrestricted license.
4. Phonetic as well as popular layout support in a single package.
5. Need for a unified multiplatform input method.
6. Ease of configurability and customizability.

Figure 1 shows the system architecture of Panmozhi Vaayil showing the various modules and their interaction.

**Existing works and what they offer**

A good amount of effort has already gone into the development of easy, flexible input methods. Some popular ones are:

- **Baraha IME** – Provides phonetic support for a fixed number of languages and designed for use on Microsoft Windows platform [3].
- **Aksharamala** – Similar to BarahaIME with support for Microsoft Windows [4].
- **Smart Common Input Method (SCIM)** – Designed to work on Linux with phonetic style of input [5].

**What indic-keyboards (Panmozhi Vaayil) offers**

- No installation hassles.
- Phonetic as well as popular keyboard layouts.
- Dynamic module enabling the addition of new keyboard layouts by even users.
- Both on Linux platform and Microsoft Windows.
- Phonetic key maps can be changed to meet user's requirements.
- Open source.
- Available under Apache 2.0 License, which means even commercial companies can use our code to develop products, after acknowledging us.

Design

The design can be broadly categorized into the following modules:

- User interface and the shell extension.
- Capturing the keyboard events.
- XML based Unicode processing.
- Rendering the Unicode.

The following block diagram shows the architecture:

User Interface and Shell Extension

The User interface is a shell extension which sits in the system tray/Notification area. The main purpose of this is to allow users to interact with the input method. This mainly involves selection of the language and the particular keyboard layout. It also helps in enabling and disabling the input method, accessing help and to display the image of the keyboard layout currently selected. Apart from these, the menu also has provision for addition of new keyboard layouts.

Capturing the Keyboard Events

The input method is designed to operate globally. That is, once the input method is enabled, further key strokes will result in characters of the particular language selected being rendered system wide. This requires the capture of the key presses system wide across all processes. A keyboard hook installed in the kernel space will enable this. This module is, therefore, platform specific.

XML Based Unicode Processing

Finite Automata exists for each language (for every keyboard layout). The finite automata has been designed as XML files, where every XML file corresponds to a keyboard layout. XML based processing makes it possible to add new layouts or new languages dynamically. The XML file has a pattern, which corresponds to the input key(s) pressed. The input pattern is matched and required processing is carried out to see if the pattern matched is a vowel or a consonant. For the input pattern, a sequence of Unicode(s) is returned. The structure of the XML file is as shown below:

```xml
<pattern>
  <char>A</char>
  <unicode>0C86</unicode>
  <consonant>0</consonant>
  <uni2>0CBE</uni2>
</pattern>
```

The above XML block indicates that for the key press “A”, the corresponding Unicode is 0C86. The consonant tag tells us whether the key is a vowel or a consonant. In case it is a vowel, a second tag indicates the Unicode of the associated dependent vowel (if any).
Two algorithms have been designed, one for phonetic style of input and the other for keyboard layouts. Both the algorithms are generic, i.e. same algorithm is used for keyboard layouts of all languages and one algorithm for phonetic input in any language.

The XML key maps can be changed on the fly and the changes are reflected instantly.

**Unicode Rendering.**

Once the key is pressed, simple grammar rules are applied to determine whether the output has to be a consonant, an independent vowel or a dependent vowel. The XML file is parsed and the corresponding Unicode is fetched. The Unicode is sent back to the process, where the keypress event took place and is rendered if any editable text area is present. The rendering of Unicode is platform specific.

**Implementation**

The following tools and languages have been used to implement the input method:

- **Java SE** - *The main language processing module has been implemented using Java SE. This has enabled easy portability and up to 80% of the code has remained common across platforms.*

- **Eclipse SWT** - *Used to implement the user interface. Eclipse SWT uses Java SWT is preferred over other toolkits to get a native look and feel.*

- **XML** - *As described previously, finite automata exists for every language (layout) and XML has been used to design it. Simple API for XML Parsing (SAX) has been used to parse the XML.*

- **Win32 libraries** - *The Windows API, is Microsoft’s core set of application programming interfaces (APIs) available for the Microsoft Windows platform. Almost all Windows programs interact with the Windows API. Some examples are SAPI, Tablet PC SDK, Microsoft Surface etc. Platform specific portions of the input method has been implemented to run on Microsoft Windows variants using the Microsoft Win32 libraries. Both keystroke capturing and Unicode rendering have been accomplished using Win32 libraries.*

**Steps:**

a. **Syshook.dll** : Install a keyboard hook in the operating system. The hook is set up for the keyboard to listen to key presses. The Windows API used is `SetWindowsHookEx()` and the library accessed is `user32.dll`. (See Fig. 2)

b. **opChars.dll** : Responsible for putting the character on to the current active window. Sends a message to the input event queue using the Windows API `SendInput()`. The library accessed is `user32.dll`

Fig. 2. JNI-Native code and keyboard hook procedure
Evdev - Also known as the user input device event interface/Linux USB subsystem. Used to capture keystrokes in GNU/Linux. This provides a clean interface to handle input devices like keyboard, mouse, joystick etc. in the userspace in Linux. This involves the following things:

a. Open file descriptor to the input device using the `open()` API with suitable parameters. Use `ioctl()` to identify the device.

b. Using the open file descriptor, continuously read bytes from the input device (can be key press/release, mouse movements) using the `read()` API.

Xlib - Used for Unicode rendering in GNU/Linux.

Steps:

a. Identify the current active window using `XGetInputFocus()`.

b. Make the window listen to all keypress events using `XSelectInput()`.

c. Using the keycodes obtained for every keypress/release event from evdev, using a mapping table to map the keycode to the keysym. Output the Unicode to the active window using `XSendEvent()` API.

Java Native Interface – Also known as JNI in short. The JNI enables the integration of code written in the Java programming language with code written in other languages such as C and C++. The write once, run anywhere concept arises from the fact that Java acts as an abstraction layer on top of the native implementation. All the API java provides have been natively implemented and the Java code allows the same APIs to be used across platforms.

The native code is usually packaged as a DLL or a Shared Object. The Java method which accesses the native code is created with a keyword “native”. Header files need to be created for the classes which contain these methods. At run-time, java code interacts with the native libraries using predefined interfaces. The native methods can also call Java methods. This mechanism is known as JNI callback.

Languages and Layouts Supported

<table>
<thead>
<tr>
<th>Language</th>
<th>Phonetic</th>
<th>Layouts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hindi</td>
<td>Yes</td>
<td>Inscript</td>
</tr>
<tr>
<td>Remington</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kannada</td>
<td>Yes</td>
<td>Inscript</td>
</tr>
<tr>
<td>KaGaPa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tamil</td>
<td>Yes</td>
<td>Inscript</td>
</tr>
<tr>
<td>Tamil99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remington</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telugu</td>
<td>Yes</td>
<td>Inscript</td>
</tr>
<tr>
<td>Gujarati</td>
<td>Yes</td>
<td>Inscript</td>
</tr>
<tr>
<td>Marathi</td>
<td>Yes</td>
<td>Inscript</td>
</tr>
<tr>
<td>Remington</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bengali</td>
<td>No</td>
<td>Inscript</td>
</tr>
<tr>
<td>Malayalam</td>
<td>No</td>
<td>Inscript</td>
</tr>
<tr>
<td>Oriya</td>
<td>No</td>
<td>Inscript</td>
</tr>
<tr>
<td>Gurmukhi</td>
<td>No</td>
<td>Inscript</td>
</tr>
</tbody>
</table>
Currently, the input method supports 10 Indian languages namely Tamil, Telugu, Kannada, Malayalam, Hindi, Bengali, Gurmukhi, Oriya, Gujarati and Marathi. The different keyboard layouts currently supported in these languages are listed in Table I.

An easy-to-use user interface has been provided to add new layouts which are Inscript like. Additional phonetic or other layouts can be added based on the existing layouts by creating new XML files and following the prescribed structure. Existing layouts can be changed/customized to suit the users' needs. In phonetic layouts, a single key press to vowel mapping is used to ensure lesser key presses for the completion of the CV combination. Ex : k (keyCode1) + Y (keyCode2) = keyCode3 instead of k + ae/ai.

**Performance**

The input method is multithreaded and the following runtime statistics have been obtained. Java Monitoring and Management console has been used to profile.

- Average Heap Memory usage : 4.0 MB (maximum : 5.0 MB)
- CPU usage : 0.2% – 0.3%
- Garbage Collector: Average time for one sweep = 0.05s. Average heap space freed up = 1 MB
- Number of threads: Peak = 15. Average live threads = 13 (2 threads are spawned by the input method)

**Conclusion**

In conclusion, the project has been an attempt to providing a very dynamic input method editor. The flexibility of adding new Indic languages on the fly, modification of the existing layouts, changing the keypress - Unicode input combination for phonetic input and a host of many other features makes for very easy to use. The main focus has been on flexibility; ease of use and to keep things to a minimum. Keeping that in mind, we have abstained from touching or modifying any system files as well as relieving the user of all installation hassles. Performance wise, the input method is fast and very light on system resources. From the user’s perspective, all one needs to do is run it. This also means the user can run the through a pen drive, CD, DVD, hard disk or any portable media. The being open source and licensed under the Apache 2.0 License, developers and users alike can modify, recompile, or rewrite the entire source and can also make these appendages closed source. Apache 2.0 license also allows developers to sell the modified code. All in all, a very dynamic, flexible, easy to use, clean, unrestrictive input method has been designed, which is multiplatform and multilingual.

**References**


Authors

Abhinava Shivakumar, Project Staff, MILE Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore. abhinav@mile.ee.iisc.ernet.in

Akshay Rao, Project Staff, MILE Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore. u.akshay@gmail.com

Arun S., Project Staff, MILE Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore. sarun87@gmail.com

A. G. Ramakrishnan, Professor, Department of Electrical Engineering, Indian Institute of Science, Bangalore. ramkiag@ee.iisc.ernet.in
Development and Evolution of Tamil Keyboard Input Systems

Ravindran K. Paul
Malaysia (ravi@tsc2.com)

There have been major developments in design of Tamil Keyboard input systems over the past 30 years. A gradual evolution has taken place in the available input systems with changes in computer technology and Operating Systems. Other than the input designs there have also been changes in the philosophies behind the input systems.

This paper will trace the evolution of the input methods and philosophies in line with the changes in computer technology. There were several significant evolutionary steps that took place in this development. These little evolutionary steps were only possible because of the contributions of several individuals and organizations.

This overview is based on personal observations and involvement in keyboard design over the past 25 years. This is not a historical record, but an overview based on information and personal interaction with key players in this field. As such other developments that I have had not access to are no included in this paper.

This evolution in keyboard design was only made possible by the contributions of several key individuals. This overview will show that some of the current thinking on Tamil Keyboard input systems are clearly outdated and need to be changed to take into account current technological and social developments.

Brief History

Tamil Computer software first started to appear in the mid 1980's in DOS based computers. During this time, Tamil software had to written by programmers as all Tamil characters need to be displayed and printed by the program. There were very few keyboard layouts and programs during this period.

In the late 1980's Windows based software began to appear. This allowed non-programmers to create Tamil fonts to type Tamil. This introduced more keyboard designs as more people were able to create Tamil input systems.

In the mid 1990's the internet revolution started and international discussion on Keyboard designs began. Based on these discussions and the Tamil Nadu Government became involved. This led to the introduction of the Tamil99 keyboard.

In the late 1990's Romanised keyboards began to emerge. Software developers like Murasu and Tamizha began to create keyboards that allowed Tamil typing using the English keyboard.

Currently all these keyboard designs are available in many software. This development of Tamil keyboard layouts over the past 25 years has seen great changes in the philosophy of keyboard design. These changes will be looked to in detail in the next few pages.
Tamil Typewriter Keyboard

In the early 1980's the most popular Tamil keyboard layout was the "Remington Typewriter" layout. The typewriter was the only way to type Tamil text and this was the keyboard layout used by journalists and writers. The first few Tamil software that were released used this design. The buyers of these software were mainly professionals and the software followed the design that they were familiar with.

In conclusion, the first keyboard input systems were visually based. The user typed the "image" that was to appear. Not much thinking was involved.

Phonetic Keyboard.

In 1986, Thunaivan introduced the worlds first Tamil Phonetic Keyboard. This was a totally new concept at the time. For the first time what the typist typed and what appeared on the screen was not the same.

A Phonetic Keyboard was defined as follows:

The phonetic keyboard is based on the use of only 13 keys for vowels and 17 keys for the consonants. With just these 30 keys, all the keys in the Tamil alphabet can be typed.
*Note: This is a personal definition of the term coined in 1986. Whether the term is correct or not is not the subject of discussion. I have seen other definitions for this term.

The principle is as follows:

\[
\begin{align*}
\text{uvir eluthu} & = \text{uvir eluthu} \\
\text{mei eluthu} & = \text{mei eluthu} \\
\end{align*}
\]

(Note that the definition of “mei eluthu” has been modified from பி to பළி)

\[
\begin{align*}
\text{uvir eluthu} + \text{mei eluthu} & = \text{uvirmei eluthu} \\
\text{பி} + \text{பி} & = \text{பி} \\
\text{பி} + \text{பி} & = \text{பி} \\
\text{பி} + \text{பி} & = \text{பி} \\
\end{align*}
\]

Because there were less keys to press, it was easier to memorise and type in Tamil. Due to this new innovation, more and more teachers started to type in Tamil. Due to my limited knowledge in Tamil, the keyboard layout was quite inefficient.

World’s first Tamil Phonetic Keyboard.

Refinements to Phonetic Design

In the first Tamil Phonetic keyboard, the arrangement of the keys was not efficient. The next development was the repositioning of keys for optimum typing speeds. In 1987 based on discussions with the late Naa Govindasamy of the Institute of Education in Singapore, both of us decided to look for a more efficient keyboard layout. This resulted in the Thunaivan Advanced Phonetic Keyboard and the I.E. Singapore Keyboard.
These designs were based on the frequency analysis of keys needed to type Tamil. Following were the results of key use analysis:

**Bar chart on frequency of Vowel Column**

Note the relatively high frequency of the vowel keystrokes over consonants below.

**Bar chart on frequency of Consonant Column**

The 3 most frequent consonants in use are ச, ட and ப. The next few characters varied in frequency depending on type of document. After this, the most frequent characters varied between ர, ம and ட.

The highest frequency for a consonant is 7.3% for ச, while the highest frequency for the vowel ஞ is more than double at 16.7%. Similarly த and ந have relatively high frequencies of 7.2% and 7.8% respectively. It would be very important to make sure that these 3 consonants and these 3 vowels are on the home keys, especially ஞ.

Although the objective was to create the most efficient layout possible, the above layout was chosen to help faster key memorizations. Also characters are from left to write. It was assumed that since we write from left to write, it should be more natural to type from left to write.
Despite this compromise, the keyboard was surprisingly efficient. 56% of all frequently used Tamil words can be typed with just the HOME keys (ASDF and JKL;)

<table>
<thead>
<tr>
<th>Thunaivan Phonetic Keyboard Keystroke Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of Keystrokes per Finger, excluding Home Keys</td>
</tr>
<tr>
<td>Q W E R T Y U I O P [ ]</td>
</tr>
<tr>
<td>2.2% 3.8% 3.2% 3.4% 0.6% 1.9% 1.6% 1.5% 1.1% 1.1% 0.4% 0.4%</td>
</tr>
<tr>
<td>Left Home Keys</td>
</tr>
<tr>
<td>20.6% 7.3% 2.3% 4.1% 6.8% 4.6% 1.8% 16.7% 7.8% 7.2% 4.5% 1.3% 36.3%</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>% 53.7% 0.7% 0.1% 1.1% 1.9% 4.0% 3.7% 2.7%</td>
</tr>
</tbody>
</table>

I.E. Singapore Keyboard
The I.E. Singapore keyboard worked on a slightly different philosophy. The primary goal was for optimum keyboard placement. Secondly, the keystroke was right to left. The assumption was that the right hand was stronger and should be faster. Also, some ideas were borrowed based on the Phonetic system in the Hindi language Keyboard developed by Mohan Thambi in India in 1983.

Unfortunately there was a mistake in the design. The "pulli" which should have been placed on the "F" key was instead placed on the "G" key. This was because the author had wrongly assumed that the HOME keys were "SDFG" and "JKL;". This was corrected in other keyboard layouts that were introduced later like the Nalinam, Tamil97 and Tamil99 keyboard layouts.

Influence of Microsoft Windows

The above developments are all pre-Windows. All these ideas were implemented in DOS based computers. With the introduction of Windows more keyboard layouts started to come out. Many of these new keyboard layouts were font based.

The first few fonts designed for Windows followed the Typewriter Keyboard layout, but with one significant difference.

\[ \text{lT} = \text{L} + \text{g} \]  

and not \[ \text{lT} = \text{g} + \text{L} \]

Basically this was the Tamil Typewriter keyboard in reverse. For the first time, non-programmers started to create Tamil input systems. By creating Tamil fonts they were able to type in Tamil. Popular fonts like Tharagai, Baamini and Mylai fonts came into existence.

Romanised Fonts

From these many fonts, one font took a slightly different approach and became very popular. This was the Mylai font by Dr. K.Kalyanasundaram. This popular font tried to match Tamil with English sounds. As shown below most of the Tamil characters were mapped to the English Alphabet.
All the developments above are pre-1995. With the start of online discussions, the need for standardisation became more important. After the first Tamil Internet conference held in Singapore in 1997 discussions on a common keyboard design began. In 1999, the Tamil Nadu government released the Tamil99 Keyboard layout as well as the TAM and TAB font encoding.

Phonetic Keyboard enhancements of Tamil 99

While previous phonetic keyboard designs had concentrated on optimum key locations, the Tamil 99 keyboard took it one step further. One of the design goals was to decrease the number of keystrokes. The Tamil 99 keyboard achieved this in the following way.

Normal Phonetic Keyboard

Note the use of க + க above. This function allows the user to save keystrokes, by not having to type an extra key for the "pulli". If there is a need for the sequence கக type the sequence க + அ + க. This applies to all consonants and is designed to save keystrokes.
This combination technique also applies to the following characters:

This new enhancement managed to reduce the number of keystrokes required and made the keyboard layout more efficient. Some of these techniques were adapted for use in later Tamil Romanized Keyboards.

**Romanized Keyboards**

Around this time Romanized keyboards began to appear. One of the first implementations of this keyboard method was developed by Muthu Nedumaran of Malaysia as part of Murasu Anjal. Other than typing directly in English, there is another interesting change here. This is the use of "mei" (_letter) instead of "uyirmei" (_letter) for consonant key strokes. For example typing the "K" key produced "_letter" instead of "_letter" as was normal in previous Keyboard Layouts. This was a significant change.

![Romanized Keyboard Diagram]

The Romanized keyboard is the last of the innovations in Tamil Keyboard development. It incorporates all the innovations that have taken place in the past 20 years. This keyboard design has had a major impact in the number of people typing Tamil. For the first time, users could start typing in Tamil at reasonable speeds within 20 minutes. Speeds that took 2 to 3 weeks to achieve could now be attained in an hour or less.

There is still some debate on which keyboard is the best or which is the best keyboard should be taught to children. Currently adults and children learn the English keyboard first when they start using computers. Younger children and teenagers reach significant typing speeds English because of high Internet use. Considering the above and the increase use of Tamil as a second language and not as first language in previous generations, the Romanized Tamil Keyboard is very suitable to be taught to adults and children alike.
11

தின்ப மலட்ப புராணம்
உலக மார்க்கில் வடைப்பட்டார்

பூச்சியமார், சுந்தர, பெண் சீலியேன்

பற்றிய தகவல்கள், கவிப்புகள்,

பட்டியலைத் தவறு அறிவிப்புகள் காணவுக்கு (பார்னூசியம் மார்க்குடன் கூறப்பட்டு காணவுக்கு)

இன்றுவரும் திறனியேப்புகள்

ஈலைநந்தம்: mkduraimani@gmail.com

மூலகூறா

21 ஆண்டு புனரத்திற்கு அருட்கரை மார்க்கில் அண்டைய உடன் பிரித்துப் பிரித்து பிரித்து கோரித்தன. கனவு சேர்த்து வந்த தன்னுடைய திட்டம் பிரித்துப் பிரித்து பிரித்து கோரித்தன. பொருந்தும் போராட்டம் வந்த தன்னுடைய திட்டமையும் திட்டமையும் திட்டமையும் கோரித்தன. இதனால் பற்றிய சொல்லான உயிரை உயிரை உயிரை பிரித்து பிரித்து பிரித்து கோரித்தன. ஐரோப்பியனின் காலநடையில் எதிரான உயிரை உயிரை உயிரை பிரித்து பிரித்து பிரித்து கோரித்தன.

மகாபால

உதுவிடமாக இந்தப்படுத்தல் பார்னூசியம் தொடங்க வந்தது. புதிய உறுப்பு பலவையும் கணினியாலேயே விளக்கும் நிகழ்வுகள் காணத்தக்க உயிரை உயிரை உயிரை பிரித்து பிரித்து பிரித்து கோரித்தன. இதை குறிப்பிட்டு பள்ளியாளர் உதுவிடமாக இந்தப்படுத்தல் பலவையும் கணினியாலேயே விளக்கும் நிகழ்வுகள் காணத்தக்க உயிரை உயிரை உயிரை பிரித்து பிரித்து பிரித்து கோரித்தன. இந்த மகாபாலும் பலவையும் கணினியாலேயே விளக்கும் நிகழ்வுகள் காணத்தக்க உயிரை உயிரை உயிரை பிரித்து பிரித்து பிரித்து கோரித்தன. ஐரோப்பியனின் காலநடையில் எதிரான உயிரை உயிரை உயிரை பிரித்து பிரித்து பிரித்து கோரித்தன.
தமிழ் எழுத்து

தமிழ் எழுத்தில் புது வலை அடிப்படையில் தமிழ்மரபு தமிழில் நூறு வரும் நாள்கள் நூற்றாண்டு பட்டியல் பதிவு செய்யவுல்ளன. எனவே கட்டுப்பாட்டுகள் லக்ஷரை நேரடியாக செய்யப்படுவதுடன், தமிழ்மரபு தொடர்புடைய வலைப்பக்கத்தின் மூலம் வலிமையாக செய்யப்படுகிறது. (karthikramas.blogspot.com/archive/21.html)

என்ன எழுத்து

என்ன எழுத்தில் புது வலை அடிப்படையில் தமிழ்மரபு தமிழில் நூறு வரும் நாள்கள் நூற்றாண்டு பட்டியல் பதிவு செய்யவுல்ளன. எனவே கட்டுப்பாட்டுகள் லக்ஷரை நேரடியாக செய்யப்படுவதுடன், தமிழ்மரபு தொடர்புடைய வலைப்பக்கத்தின் மூலம் வலிமையாக செய்யப்படுகிறது. (www.navan.name/blog/?p=18)

என்ன எழுத்து

என்ன எழுத்தில் புது வலை அடிப்படையில் தமிழ்மரபு தமிழில் நூறு வரும் நாள்கள் நூற்றாண்டு பட்டியல் பதிவு செய்யவுல்ளன. எனவே கட்டுப்பாட்டுகள் லக்ஷரை நேரடியாக செய்யப்படுவதுடன், தமிழ்மரபு தொடர்புடைய வலைப்பக்கத்தின் மூலம் வலிமையாக செய்யப்படுகிறது. (karthikramas.blogdrive.com/archive/21.html)
1. பல வாசவப்படுத்திகள் அறிவிக் கால்நிலைக்கான அவசியமாகவும், நவீனத்துறுத்துணர்வு பலகை அளிக்கிறது. நல்ல பணியுடன் திறமையான உயர்நிலைப்படுத்தும் கட்டுறக்கள் கவனத்தை ஆராய்ச்சியாக வலியுறுத்துகிறது. இதன் நூலின் விளக்கங்கள் விளக்கங்களின் விளக்கங்களையும் குறிப்பிட்டுகிறது. (www.emadal.blogspot.com)

2. சுகாதார கல்விப்படுத்திகள் வலியுறுத்தும் பணியுடன் உயர்நிலையில் உருவாக்கப்பட்டுள்ளன. பெரும்பான்மையான பாவு விளக்கங்கள் விளக்கங்களின் விளக்கங்களின் விளக்கங்களையும் குறிப்பிட்டுகிறது. பொதுமக்கள் பணியுடன் உயர்நிலைக்கான அமைப்புகள் பலகை ஆராய்ச்சியாக வலியுறுத்துகிறது. நல்ல பணியுடன் திறமையான உயர்நிலைப்படுத்தும் கட்டுறக்கள் கவனத்தை ஆராய்ச்சியாக வலியுறுத்துகிறது. இதன் நூலின் விளக்கங்கள் விளக்கங்களின் விளக்கங்களையும் குறிப்பிட்டுகிறது. (www.mandidar.blogspot.com)

3. முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. மூலமாக வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. (www.mvelangovan.blogspot.com)

4. முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. (www.tamiltools.blogspot.com)

5. தொல்லியல் அறிவியல் அறிவியல் முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. முற்பதானம் வலியுறுத்தும் பணியுடன் உருவாக்கப்பட்டுள்ள மூலமாக பல வாசவப்படுத்திகள் அனுப்புகிறது. (www.murugaperuman.blogspot.com)
செயலியல் துறையிலுள்ள எந்தகூற்றுகளும் மூலம் பயின்று வைத்தற்கின்ற போன்ற செயலியல் துறையிலுள்ள எந்தகூற்றுகளும் மூலம் பயின்று வைத்தது. மேலும் பல வராயனங்கள் செயலியல் துறையிலுள்ள எந்தகூற்றுகளும் மூலம் பயின்று வைத்தது. பல வராயனங்கள் செயலியல் துறையிலுள்ள எந்தகூற்றுகளும் மூலம் பயின்று வைத்தது. 

6. மேலும் கூற்றுகளின் செயலியல் என்னும் செயலியல் துறையிலுள்ள எந்தகூற்றுகளும் மூலம் பயின்று வைத்தது. 

III. செயலியல் துறையிலுள்ள எந்தகூற்றுகளும் மூலம் பயின்று வைத்தது. 

IV. இந்தகூற்றுகள் மூலம் பயின்று வைத்தது. 

V. இந்தகூற்றுகள் மூலம் பயின்று வைத்தது. 

VI. இந்தகூற்றுகள் மூலம் பயின்று வைத்தது. 

VII. இந்தகூற்றுகள் மூலம் பயின்று வைத்தது. 

VIII. இந்தகூற்றுகள் மூலம் பயின்று வைத்தது.
IX. எல்லாப்பகுதிகளும் இருந்தும் தொழிலாளர்களின் நிலை மற்றும் தனிப்பாட்டில் விளக்கம் வாசிக்கவுடன்.

X. எல்லாப்பகுதிகளிலும் நூற்றட்டங்களும், நாகர் குடும்பங்களும், மற்றும் வருஷத்தில் தன்மையான பிரார்த்தனை வெளியேற்றவுடன் மலர் இருக்கின்ற விளக்கம் வெளியேற்றவுடன் மாற்றார். இது மற்றும் பலரும் கீழ்வரை விளக்கம் வெளியேற்றவுடன் மாற்றார். இப்போது பலவகையான வகையான விளக்கங்களும் பிரிவுகளும் வாசிக்கப்பட்டுள்ளன.
ஆசிரியர் சமாதானத்தழை மாணவபூர்வ்வாகத்து (Blogs)

nagalango@gmail.com

இதுவண்டி பயன்படும், பட்டமுறைப்படும் வழி பக்திக்குது.

ஆசிரியர் சமாதானத்தழை

முன்னெச்சுட்டித் வரவேற்குப்பட்டு மதிப்புற்றுள்ளனர். மன்னர் சான்றத்தை தொடுகள் மீது மாரபிற்றுக்கொண்டு பெறும் மற்றை காணமலை தமிழ்நாடு குழுவில் கைத்திருந்தன. இவ்விளக்கு மீண்டும் மன்னர் சமாதானத்தழை மாணவபூர்வ்வாகத்து பாதுகாப்பு பதவி பெற்று அன்று வேலூர் மாநிலத்தில் இருந்தது. மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக. மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக. மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக. மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக. மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக. மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.

மாநிலத்தில் சுயங்கு மாணவபூர்வ்வாகத்து துனையாளராக.
தொலைவுக் கூட்டத்தில் குறிப்பிட்டுள்ள ஆற்றல்களுடன் உள்ளதும் தினநேர்முறையில் தொடர்ந்து வந்துருவது குறிப்பிட்டுள்ளது. ஒவ்வொரு குறிப்பிட்டுள்ள ஆற்றல்களும் உள்ளதும் தினநேர்முறையில் குறிப்பிட்டுள்ளது. அது குறிப்பிட்டுள்ள ஆற்றல்கள் உள்ளதும் தினநேர்முறையில் குறிப்பிட்டுள்ளது.
Blogging

Archives

- - -

HTML

Archives (archive)
வெளியில் செய்யப்பட்டுள்ள HTML ஆறியறிவாள். வெளியில் செய்யப்பட்டுள்ள HTML ஆறியறிவாள் ஆண்டவர்கள் ஒன்றிணைந்து பேசுகின்றனர். ஒருவரிடையே ஒருவர் பேசுகின்று. மேலும், பொருள் பெற்றுள்ள மூன்று போட்டியில் பேசுகின்றது. பொதுவாக பதிவு வழங்குவதற்கான ஆண்டவர்கள் ஒன்றிணைந்து இயங்குகின்றனர். பதிப்புடன் (Templates) செய்யப்பட்டுள்ள பதிகளின் படியாகப் பதிவு வழங்குவது.
பின் நாள்வருட்காக தமிழ் வாரணாசியில் செல்லவும் வழங்கத் தொகை வெளியிட்டு வந்தால் எப்படி இன்றுக்கு நான் முன்னிட்டு வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும். என்பதோடு வெளியிட்டு வந்தால் எப்படி இன்றுக்கு நான் முன்னிட்டு வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.

1. நான்கு வகுப்புகள் உள்ளன. விளக்கமின்னும் குறிப்பிட்டு வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.

2. நடைமுறையில் கருத்துடன் அளித்து நான்கு வகுப்புகள் உள்ளன. விளக்கம் வெளிப்படுத்துவதற்கு கருத்துடன் தொடர்ந்து அளித்து வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.

3. நான்கு வகுப்புகள் உள்ளன. விளக்கம் வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.

4. மாணவர்கள் வகுப்புகள் உள்ளன. விளக்கம் வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.

5. மாணவர்களுக்கு முன்னிட்டு வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.

அவர்களுக்கு ஆதாரமாக தேவையான விளக்கத்தை நுழைவு வந்து விளக்கம் வெளிப்படுத்துவதற்கு தொடர்ந்து அறிவுகள் தெரியும்.
Tamil Blogs – Tools, Aggregators and Beyond

Kasi Arumugam

Abstract:

Evolution of Blogs in Tamil started in 2003 and is growing steadily. This article describes technical challenges faced by Tamil Bloggers at the early phase mainly caused by encoding related issues, and how Unicode got established as a standard for Tamil web. As founder-developer of the first Tamil blog aggregator Tamilmanam, the author explains the features and issues of tools and services dedicated for the Tamil Web Content Management with particular relevance to Blogs. By recording events, the Article attempts to help future web development efforts targeted for Tamil community. A few thoughts on the current trends and future possibilities are also touched upon.

The Basics of Blogs:

A Blog is a personal website that helps an individual or a team of few people to publish their content on the internet for many others to read. The fundamental differences between a typical ‘website’ and a blog (even though slowly getting blurred these days), can be explained by the following table:

<table>
<thead>
<tr>
<th>Feature</th>
<th>A typical ‘Website’</th>
<th>A Blog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Updating frequency</td>
<td>Less frequent - some of these not updated since launch</td>
<td>Very frequent, sometimes several times a day</td>
</tr>
<tr>
<td>People involved</td>
<td>Author→ Designer→ Developer→ Publisher/ Hoster</td>
<td>Blogger</td>
</tr>
<tr>
<td>Run by</td>
<td>Institution/Business/Government</td>
<td>Individuals</td>
</tr>
<tr>
<td>Participation model</td>
<td>Readers have no direct say</td>
<td>Readers’ participate through comments; bring life into the system</td>
</tr>
</tbody>
</table>

Evolution of Blogs as alternative media around the world

Blog is a child of technology. Unlike other forms of publications that had a classical form and a modern form, blogs cannot be thought of without the ‘wired world’ that is today’s World-Wide-Web. Blogs takes away the need for an author to be at the mercy of an editor or a publisher. Blogs have brought in equality and democracy to the countless minds that aim at the ‘authorship’. Thus a blog makes the voices of weak and less-opportune people be heard – technology’s true gift to the society.

Blogs have started playing vital roles, more visibly so in developed countries, in politics, technology and issues concerning societies. Leading media websites have exclusive pages to show popular blogs.
e.g.

- Guardian -UK (http://www.guardian.co.uk/tone/blog)
- The Indian Express – India (http://www.expressbuzzblogs.com/).
- BBC – UK (http://www.bbc.co.uk/blogs/)

Technological Challenges faced by early Tamil Bloggers

It is understandable that Blog authors need to have access to web-enabled computers. But, Tamil Bloggers were challenged by few more special needs: the authoring tools & displaying technologies. They had three major needs:

1. Tamil typing tools for entering their writings into the computer, with facilities to convert from one encoding to another.
2. Technologies for ensuring that those prospective readers see Tamil text properly at their desktop without any downloads/ modifying browser configurations.
3. Tamil typing tools for readers to enter comments using no new downloads/ learning.

Technology enthusiasts and volunteer teams lent great help to Tamil Bloggers in this area that brought Tamil blogs to the forefront of many Indic languages.

Tamil typing tools:

While there were many tools available for typing Tamil in computers, the Tamil blogging community mainly operated with only few tools viz. Murasu Editor, e-Kalappai and PonguTamil writer. Murasu Systems offered free downloads of Murasu Editor. E-Kalappai was made a free download thanks to a group of donors who had paid on behalf of the community. PonguTamil was an online service from suratha.com. Of these, due to the simplicity and compatibility with multiple applications, e-Kalappai was the most popular among Bloggers, particularly those who entered Tamil computing after the blog era started. PonguTamil (http://www.suratha.com/reader.htm ) suited those without having rights to install anything on their computers.

Display – related issues:

Then came the next challenge. Hitherto, most websites displaying Tamil content had their text created with a variety of encodings. There were proprietary fonts every viewer needs to download and install for each website one visits. Even after this, there were browser settings one had to do in order to see Tamil properly rendered on their screen.

Unicode standard was already out and few websites like ezhilnila.com, thisaigal.com, etc. have been displaying Unicode Tamil content. This demonstrated that Unicode provided the answer to most of the issues concerning displaying text at reader’s computer out-of-the-box – no font downloads, no tricky browser settings. This sealed the growth of other encodings for Tamil Blogs. Even few blogs operating with TSCII text had to eventually convert to Unicode.

With the growth of blogs in Tamil, Unicode became the de-facto encoding standard for the Tamil web. Today the most visited media sites such as Dinamani, Dinamalar, Dinakaran, Kumudam, Vikatan etc., in Tamil are in Unicode.
Windows 98-specific issues:

Still there was a need to support the computers running on Windows 98, where Unicode was not working straight away. Users still needed to download Unicode fonts; People offered how-to help pages for enabling Indic language setting, updating the uniscribe rendering library (USP10.dll). But still it needed a simpler solution. The web embedded font technology (WEFT) of Microsoft (simply called ‘dynamic font’) filled in this gap and Tamil Bloggers readily grabbed it. Of course it required certain paid technology that worked only for a given domain. Thanks to Athirampattinam Umar Thambi’s domain-independent dynamic font file Thenee.eot, Tamil blogs exploited this technology to the fullest extent and all their blogs were fully working in Windows 98. Thenee was such useful tool that it supported both TSCII and Unicode in a single font file, helping many comment-writers who were using TSCII as part of their yahoo group activity.

Tamil typing tools for comment-writers

The last Tamil-specific challenge was to offer means for the blog readers to type-in their comments. Obviously it was utterly impractical to expect a casual reader to learn and install tools/settings for typing in Tamil. Being online tool running on JavaScript that every browser supported, PonguTamil JavaScript code was embedded into many comment boxes enabling Romanized input of Tamil by readers. Gopi improved this further and offered with many options at http://www.higopi.com/ucedit/Reuse.html

Volunteering and Community Efforts

Tamil Blogs started appearing in the scene from early 2003. There were several online resources mainly by early volunteers like Suratha Yazhvanan, Mathy Kandasamy, Umar Thambi, etc. explaining blogging in general and Tamil-specific issues in particular. These offered help and guidance for many early Bloggers.

The ‘Tamil Bloggers List’ at http://tamilblogs.blogspot.com/ originally compiled by Mathy Kandasamy was the launch pad for many new Bloggers. Community blog Valaippoo (http://valaippoo.blogspot.com/) was a a meeting point for discussing blogging related issues while its weekly authors tried to show case interesting posts from Tamil blogs.

An article series by the author himself on essentials of blogging (இந்திய செய்திகள் அங்காளின் பேரிட்டு பல்லாடுதல் பேர்கள்) was originally published in e-sangamam e-zine and is still available at his personal blog at http://kasilingam.com/wiki/doku.php?id=tamil_blogging

Several Bloggers themselves were writing articles on blogging and computing such as one by the author himself on encoding and Unicode: ஐடோ ஓடுதீரை, ஐடோ ஓடுதீரை, ஓடுதீரை ஓடுதீரை, ஓடுதீரை http://kasilingam.com/wiki/doku.php?id=tamil_unicode_for_a_blogger

By mid 2004, there were around 100 blogs written in Tamil. The blogging community needed new initiatives to help and manage growth. News feed a.k.a. rss technology became available for the popular Blogger.com service and there were efforts to network Tamil Bloggers as majority of them are from Diaspora and needed a common platform to showcase and reach out to the readers.
Arrival of Aggregators

Having created and published their blog, the authors now had to overcome the next challenge: how to make themselves visible in the crowd? Unlike the conventional magazines, blogs are written with no fixed release day or time. Blog’s ability to totally break the order of time needs a technological answer too. They need to keep the interested audience to know ‘What is written fresh this minute? How to reach it? Who wrote it and who are all engaged in conversations on such stuff?’

Blog aggregators help answer the above questions through intelligent use of technology. www.tamilmanam.net (formerly www.thamizmanam.com, started in Aug 2004) was the first and foremost of the new genre of websites called blog aggregators. Tamil led the whole of India in the concept of blog aggregators and have pioneering efforts in this front. Blog aggregators resemble large communities on the cyberspace serving both authors and readers. They are the main streets of the blogging village and provide a meeting point; they make this momentum to sustain. Aggregators accelerate the pace of blogging movement thus plays a vital role in the growth of web pages in languages of the people.

Blog portal Tamilmanam

Using open source feed aggregation technology as its starting point for Tamilmanam, the author ventured into developing programs and services tailored to meet the specific needs of Tamil bloggers. Tamilmanam employed the popular open-source tools PHP & MySQL to offer a feature-rich, contemporary service for the benefit of Tamil Bloggers.

Some of the pioneering features of Tamilmanam were:

1. Bloggers List sortable by name, location, start date, etc. updated by auto submission of URL
2. Auto aggregation of posts (updated every 20 minutes)
3. Posts written in the past days
4. Posts written on specific topics/ genre
5. Automatic aggregation of Comment status
6. Facility to convert a post into a PDF document
7. Voting on posts
8. Automatic hiding of Non-Tamil post
9. Intimation of objectionable content
Tamilmanam’s most visited ‘readers’ page w/ star of the week, hot tags, auto-updated thumbnails of cartoons and photos – Nov-2004

Tamilmanam Star of the week showcased prominent Bloggers for a week. Tamilmanam was the first to aggregate comments to blogs long before blogging services offered newsfeeds for comments.

Tamil Blogs grew leaps and bounds with Tamilmanam. Adjoining chart published in one of the studies of blogs in India revealed the position of Tamil Blogs among other Indian language blogs.

More than a year later, Tamilmanam was rewritten and upgraded to version 2 with Pathivu toolbar. This provided for instant aggregation, post-wise categorization, one-click voting and on-the-fly pdf e-book creation was first-of-its-kind technological feature of Tamilmanam.

More Tamil Blog aggregation services

Thenkoodu was another aggregator service started around the time Tamilmanam v2 was launched but had to suspend operations due to sudden demise of the developer-owner, Sakaran. Other leading aggregator services Tamilveli and Thiratti are playing important roles in their own ways to further the growth of Tamil blogs.

Other than aggregators based on automated scripts, there were efforts by teams in showcasing select blog posts by manual recommendations. Gilli was a pioneering effort followed by Maatru. Tamilish merged the automation and manual recommendations using Digg.com model. There is a spurt of new aggregators coming out recently.

Efforts beyond aggregation

Tamilmanam’s efforts towards nurturing the alternative media and supporting Tamil computing was taken further by the current team TMI (Tamil Media International Inc), a US-registered non-profit organization. TMI team also brought out Poonga, a monthly e-zine based on collections entirely out of blog posts. TMI instituted an elaborate awards program for Tamil Blogs and have developed dedicated software for managing the award program that is unique in several ways, like there is no judges panel, etc.

Tamilmanam Today

Tamilmanam currently aggregates over 7000 blogs that update with some 300+ posts everyday. It is feature-rich service and empowers a vibrant community of Tamil Bloggers. Many TMI members are part of INFINIT and are involved in advancement of Tamil in internet and computing. Thanks to its primary technical contributor Sasikumar’s continuous efforts, Tamilmanam is up-to-date in terms of technology. It is one of the few successful team efforts in Tamil computing world.

Tamil Blogs as alternative media

The emergence of blogs has created an independent alternative media that touch upon issues hitherto not given adequate space. The speed, economy, varieties of topics and plural perspectives possible to be written in blogs have no parallels in the conventional media. Along with blogs, the language computing is coming of age and blogs are here to stay.
Many mainstream writers have started to blog. People from other creative fields, like the film world are also attracted towards this interesting and creative pastime. Many magazines have started highlighting interesting blogs. Some of the noted blogs have been printed into books too. All this lead to greater visibility for Tamil blogs that are showing signs of alternative media.

The Trends of Tamil Bloggers:

The micro-blogging service Twitter is an effective short communication-cum-networking medium that many Bloggers are now tweeters too. The companionship they have generated through their blogs helps them get a ready-made following in Twitter. Twitter is available in cell phones, a major push over other web-only networking services. Still, for expressing elaborate thoughts on the net, Tweeters use blogs. There are several automated cross-service applications that lets one makes best use of each from the other. Many Bloggers trans-locate their friendship network into Social Networking such as FaceBook, Orkut, MySpace, etc. and Professional Networking such as LinkedIn. As in micro-blogging, their Blogging continues, albeit less frequently.

Bloggers arrange meets as an extension of their coming together in Blogosphere. Such meetings of Tamil Bloggers are very common in Chennai, as here is the largest congregation of Tamil Bloggers. Meets also happen at locations such as Coimbatore, Puducherry, Bangalore, Erode, Madurai. Bloggers from the Diaspora too meet at Singapore, USA, UAE, Sri Lanka, etc. Bloggers organize workshops to encourage and help more would-be bloggers.

Future scenario

There are blog-driven news/media sites in English, a very popular one being SlashDot (www.slashdot.com ) Tamil blogs have full potential to evolve into such a format which will differ from plain aggregators/bookmarking applications. A possible blog-driven media site with following working arrangement is very much a reality:

1. Enlisted authors write independently and publish their articles at a central site on topical matters on a continuous basis.
2. Editor’s panel is just to monitor and intervene on legal issues or complaints but no formal pre-reading before publication. This ensures timely publication on news and articles, but still moderated for integrity/legality.
3. This online media will have at a given time 10’s or 100s of contributors, who have been selected by the editor’s panel based on their demonstrated authorship. Thus there is no scope for much noise that is common in current aggregators.
4. There will be moderation of comments, voting and abuse-reporting
5. This site will be monetized by advertisements and the revenue will be shared among the contributors after deducting the expenses, proportional to the page hits of each article.

Tamil blog world is waiting for web application developers to bring out interesting relevant applications.
மார்த்தான் மீண்டும் நிச்சயமான அகலம்

கல்வி குடும்ப கணினி

suba.nargunan@gmail.com

http://thirutamil.blogspot.com
தொலுங்கு அப்பா வன்ன கன்னி தெரள்போனது

1989ஆம் ஆண்டு ‘வெல்ல’ என்றால் அரசு தொலுங்கு பெருமையற்ற முறையைப் பயிப்பதற்கு எதிரிக்கும் மதிக்கும் மாற்றம். 1993ஆம் ஆண்டு பையூரி என்றால் தொலுங்கு பெருமையற்ற முறையை இன்றும் பயிப்பதற்கு எதிரிக்கும் மதிக்கும் மாற்றம்.

இன்றும், தொலுங்கு பெருமையற்ற முறையின் தூண்யில் வரும் முறையிலும் அப்பா வன்ன கன்னி தெரள்போனது.
வெளியிட்டுள்ள சொல்லியின் நோக்கம் போற்றுகிறது பிரித்து குறித்திருக்கிறதும் முறையிலும். இவ்வப்படுத்தல் வழியில் பிரித்து, வெளியிட்டுள்ள சொல்லியின் நோக்கம் போற்றுகிறது பிரித்து குறித்திருக்கிறதும் முறையிலும். 2000 ஆம் ஆண்டு வரையறுக்கப்பட்டுள்ள வெளிப்புறம் பார்வையிலி விளக்கம்.

மேலும் நூற்றாண்டு இரண்டு முதல் பிரித்து செய்யப்பட்டுள்ள வேலை, வெளியிட்டுள்ள சொல்லியின் நோக்கம் போற்றுகிறது பிரித்து குறித்திருக்கிறதும் முறையிலும். அவர்கள் வலியேற்றப்பட்டுள்ள வெளிப்புறம் பார்வையிலி விளக்கம்.

மேலும் நூற்றாண்டு இரண்டு முதல் பிரித்து செய்யப்பட்டுள்ள வேலை, வெளியிட்டுள்ள சொல்லியின் நோக்கம் போற்றுகிறது பிரித்து குறித்திருக்கிறதும் முறையிலும். அவர்கள் வலியேற்றப்பட்டுள்ள வெளிப்புறம் பார்வையிலி விளக்கம்.

மேலும் நூற்றாண்டு இரண்டு முதல் பிரித்து செய்யப்பட்டுள்ள வேலை, வெளியிட்டுள்ள சொல்லியின் நோக்கம் போற்றுகிறது பிரித்து குறித்திருக்கிறதும் முறை�ிலும். அவர்கள் வலியேற்றப்பட்டுள்ள வெளிப்புறம் பார்வையிலி விளக்கம்.
உலக தொடர்பு தற்காலத்தை மீண்டும் விளையாட்டுகள் நோக்கிய ரீதியாக இயக்கப்பட்டுள்ளது. முதல் இல்லாமலும் இருந்துள்ளது போக்குகள் என்பதாக விளக்கப்பட்டது. இதனால் முக்கியமான முக்கியத்துவம் மற்றது அல்லது ஒரு வல்லுனரால் என்று மேற்கோள்கள் உள்ளன. உலகத் தொடர்பு தற்காலத்தை மீண்டும் விளையாட்டுகள் நோக்கிய ரீதியாக இயக்கப்பட்டுள்ளது. முதல் இல்லாமலும் இருந்துள்ளது போக்குகள் என்பதாக விளக்கப்பட்டது. இதனால் முக்கியமான முக்கியத்துவம் மற்றது அல்லது ஒரு வல்லுனரால் என்று மேற்கோள்கள் உள்ளன.
உலகப்பகுதியில் என்று. பொதுவாக குழு வியாபாரச்செயலாளர் ஒரு குழுக்களாகவும் உரையாளர் என்றும் உலகப்பகுதிகளின் ஒரு குழுக்களாகவும் அர்க்காமா குழுவாளர், வணக்கத்தார், வண்ணம் வைக்கப்பட்டுள்ளனர். அவ்விருந்து அவர்கள் தனிச் செயல் செய்தவைகள் குறிப்பிட்டு வருகின்றனர். குழுவாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர். உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர். உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர்.

குழு வியாபாரச்செயலாளர் இல்லைத் தொடர்ந்து உரையாளர் இல்லைத் தொடர்ந்து அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர். உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர். உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர்.

தொடர்ந்து உரையாளர் இல்லைத் தொடர்ந்து குழுவாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர். உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர். உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் அலங்கார உரையாளர் என்றும் உழைக்கப்பட்டுள்ளனர்.

வாள்பாண்டியம்

உரையாளர் உரையாளர் உரையாளர் உரையாளர் இல்லைத் தொடர்ந்து வாள்பாண்டியம் இல்லைத் தொடர்ந்து குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் ஆரிய உரையாளர் என்றும் அவர்கள் குழுவாளர் என்றும் உரையாளர் என்றும் அரண்மனை இல்லைத் தொடர்ந்து பொருள் கூறப்பட்டுள்ளனர்.

671
 3.  
 2.  

விளக்கம்:
1.  தமிழ்  வளாயம் 2001 புத்தாக்கம் களப்பதனத்தின் இயற்கைப் புத்தாக்கம்
2.  புதுக்குடி குறிப்பிட்டு, நைக்கரங்கள் குறிப்பிட்டு, விளக்க நோக்கினாள் எழுதல் (2007), எசுமாண்டா, ப. 250 – 264
3.  ஐய்யும்பியா புரசனார் குறிப்பிட்டு, நைக்கரங்கள் எழுதல் (2002), ப. 271 – 273
12

மிகுந்தை தோற்று தோல்வ

நூற்றீண்டுப்போறம்
அந்த குறிப்பிட்டிட்டில் குறிப்பிட்டிட்டில் இருந்து காக்கியும் பிரமட்டா குறிப்பிட்டிட்டில் குறிப்பிட்டிட்டில் அதிகாரத்தில் கானும் முனைவாயுதான். இது ஒளி அமர்ந்து இந்த அனைத்து குறிப்பிட்டிட்டில் இருந்து காக்கியும் பிரமட்டா குறிப்பிட்டிட்டில் இருந்து கானும் முனைவாயுதான்.

முழுக்காக அருகமானவர்கள் கூறுகின்றனர் பின்னர் காரணிகள் பதிவிட்டிட்டில் இருந்து கானும் முனைவாயுதான். மிகு அணுதல் / பிரமட்டாக தரும் தீர்மானம் இவற்றை குறுக்கும் வருடாமல் வாற்றவும் முடியாது.

இன்னும் முழுக்காக கல்லால் விளக்கும் பின்னர் காரணிகள் பதிவிட்டிட்டில் இருந்து கானும் முனைவாயுதான். அதே காரணிகளின் விளக்கம் வாற்றவும் பதிவிட்டிட்டில் முருக்கும் முனைவாயுதான் என்று விளக்கும் முடியாது.

மறிகைகள் 32 வருடந்தவர். பராமரிக் தீர்மானம் அத்துடன் அதே வருடந்த என்றால், குறுக்கு விளக்கும் காரணிகள் பதிவிட்டிட்டில் இருந்து கானும் முனைவாயுதான் "ஒருவர் ஒருவரும் பதிவாக்கம் பதிவிட்டிட்டில் முறையே ஒருவர் ஒருவரும்" என்று விளக்கம். அருகு என்றால் குறிப்பிட்டிட்டில் என்று விளக்கம் பதிவு பயன்படுத்திட்டில் வாற்றவும் முடியாது.

முனைவாயுதான் விளக்கக்காரவின் என்னும் விளக்கம் அதன் விளக்கத் தீர்மானம் ஒருவர் விளக்கானது. அப்படி விளக்கான தீர்மானம் ஒருவர் விளக்கானது. ஆனால் விளக்கம் பதிவிட்டிட்டில் விளக்க உள்ளது. குறிப்பிட்டிட்டு ஒருவர் விளக்கானது.

என்று பதிவு தீர்மானம் ஒருவர் விளக்கானது. என்னும் விளக்கம் பதிவு?

நிரணார் அசத்தல் அளவில் ஒருவர் விளக்கானது. குறிப்பிட்டிட்டு 385 முறைகளில் ஒருவர் விளக்கானது! நிரணார் நிரணாரும் முடிவு 12,618 வருடந்த என்றால். நிரணாரும் பாசுகை என்று தீர்மானம் ஒருவர். குறிப்பிட்டிட்டு ஒருவர் விளக்கானது என்று தீர்மானம் ஒருவர் விளக்கானது. என்னும் விளக்கம் பதிவிட்டிட்டு விளக்கானது. எண் 0 முறைகளில் நிரணார் பதிவு ஒருவர் விளக்கானது. என்னும் விளக்கம் பதிவிட்டிட்டு விளக்கானது. எண் 0 முறைகளில் நிரணார் பதிவு ஒருவர் விளக்கானது.
அம்மா நென்றால் அதுட்கு செய்தியான டெம்பை கொண்ட எழுத்தாளர் பார்க்கச் செய்கிறார். அத்துடன் செய்தியான இடுக்கு அதுன்று பார்க்கச் செய்கிறார். எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுந்தவர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழுத்தாளர் எழு Chattanooga
ஒரு கொழும்பி தன்மையிலேயே. அது என்று தம்முற்குறியான விளக்கம் வந்து வரும் இந்த பொறியியல் அரங்கங்கள் விளக்கப்பட்டுள்ளது. அதன் மூலம் முதல் தொடங்கும் காரணங்கள் எண்ணாக்கலாம். இந்த பொறியியல் அரங்கங்கள் தம்முற்குத்து ஆர்வண காரணங்கள் எண்ணாக்கலாம். 24 மாத இணைப்பாக வசதியை அது அறிவிவாத தரவு விளக்கும் காரணம். 24 மாத இணைப்பாக வசதியை அதற்கு நிறூற்றும் காரணம். என்றால் பொறியியல் அரங்கங்கள் எண்ணாக்கலாம். 300 மாத இணைப்பாக வசதியை அதற்கு நிறூற்றும் காரணம் தொடங்கும் பொறியியல் அரங்கங்கள் எண்ணாக்கலாம். 7,200 மாத இணைப்பாக வசதியை அதற்கு நிறூற்றும் காரணம். 

ஒரு பொறியியல் அரங்கங்கள் வசதியைப் பிரித்துக்காட்டும் பொறியியல் பொறியியல் காரணம் மாடார வெளியை வந்து முடியும் காரணம் 75 மாதமாக அனுமதிக்கப்படும். 

பொறியியல் வசதியைப் பிரித்துக்காட்டும் பொறியியல் பொறியியல் காரணம் மாடார வெளியை வந்து 7,50,000 மாதங்கள். 

32 பொறியியல் வசதியைப் பிரித்துக்காட்டும் பொறியியல் பொறியியல் காரணம் மாடார வெளியை 22,65,000 மாதங்கள். 

தன்மையிலேயே தன்மையிலே! தன்மையிலேயே! தன்மையிலேயே! தன்மையிலேயே!
பருவங்கள் பொருள்பாடு பணிகள் அனுப்பப்பட்டிருக்கும். புதுக்காட்டும் பொருள்பாடு பிளேசங்களைச் சேர்ந்த அநுமதிப்புப் பருவங்கள் பொருள்பாட்டின் முக்கியத்துவம் பெற்றது. அந்தப் பருவங்கள் பொருள்பாடு பொருள்பாடு கொண்ட தலைமைச் சேமாரின் தலைமைத்துறை அவர்களுக்கு அளிக்கவுள்ளது. அவர்கள் தலைமைத்துறை பொருள்பாடு பொருள்பாடு கொண்ட தலைமைச் சேமாரின் தலைமைத்துறை அவர்களுக்கு அளிக்கவுள்ளது. அவர்கள் தலைமைத்துறை பொருள்பாடு பொருள்பாடு கொண்ட தலைமைச் சேமாரின் தலைமைத்துறை அவர்களுக்கு அளிக்கவுள்ளது. அவர்கள் தலைமைத்துறை பொருள்பாடு பொருள்பாடு கொண்ட தலைமைச் சேமாரின் தலைமைத்துறை அவர்களுக்கு அளிக்கவுள்ளது.

- புதுக்காட்டும் பருவங்கள் பொருள்பாடு பொருள்பாடு கொண்ட தலைமைச் சேமாரின் தலைமைத்துறை அவர்களுக்கு அளிக்கவுள்ளது.

http://thoduvaanam.blogspot.com/2010_01_01_archive.html
http://namakkalcollector.net/
http://www.tn.gov.in/districts.html
http://www.tn.gov.in/departments.html
http://panchayat.nic.in/index.do?siteid=101&sitemain=Government%20of%20India%20<br>%20Ministry%20of%20Panchayati%20Raj
http://panchayat.nic.in/viewMore.do?ppid=200&ptltid=375&itemid=1

678
Prototype for E-government issued E-card using RFID technology based on Tamil font database

Vijayalakshmi.S.R
Assistant Professor, School of IT and Science, Dr.GRD College of Science, Coimbatore-14, India.
srvijisiva@gmail.com

Abstract

RFID is not a new technology and has passed through many decades of use in military, airline, library, security, healthcare, sports, animal farms and other areas. Industries use RFID for various applications such as personal/vehicle access control, departmental store security, equipment tracking, baggage, fast food establishments, logistics, etc. The enhancement in RFID technology has brought advantages that are related to resource optimization, increased efficiency within business processes, and enhanced customer care, overall improvements in business operations and healthcare. Passports and other identification documents may be enhanced using the advancements in RFID technology. Various national and international bodies are pursuing machine-readable approaches with biometric information. This paper examines implementation regarding these electronic approaches and developments toward electronic data storage and transmission. Radio-frequency identification (RFID) devices for electronic passports and other existing identity documents are discussed. Our proposed research aim is to produce a model for e-government. All identity information about the citizen is included and database is maintained in Tamil language. Every citizen is having voter ID, PAN card, Credit card, ration card, driving License information, insurance information, Passport and many more. All cards can be replaced with single E-card. In the proposed model, database (all information about the citizen) is stored in tamil font using php mySQL database. This E-card provided by the government could be used by the citizen. The E-government maintains database in Tamil of all citizens. It helps in identifying the person during voting time. It can be used as security. However, the focus of this paper is to explore the main RFID components, i.e. the tag, antenna and reader.

Keywords: RFID technology, Electronic Passport, E-card, Biometrics.

1. Introduction

RFID stands for Radio Frequency Identification and is a term that describes a system of identification. RFID is based on storing and remotely retrieving information or data as it consists of RFID tag, RFID reader and back-end Database. RFID tags store unique identification information of objects and communicate the tags so as to allow remote retrieval of their ID. RFID technology depends on the communication between the RFID tags and RFID readers. The range of the reader is dependent upon its operational frequency. Usually the readers have their own software running on their ROM and also, communicate with other software to manipulate these unique identified tags. Basically, the application which manipulates tag deduction information for the end user, communicates with the RFID reader to get the tag information through antennas. Many researchers have addressed issues that are related to RFID reliability and capability. RFID is continuing to become popular because it increases efficiency and provides better service to stakeholders. RFID technology has been realized as
a performance differentiator for a variety of commercial applications, but its capability is yet to be fully utilized.

2. RFID Evolution

RFID technology has passed through many phases over the last few decades. The technology has been used in tracking delivery of goods, in courier services and in baggage handling. Other applications includes automatic toll payments, departmental access control in large buildings, personal and vehicle control in a particular area, security of items which shouldn’t leave the area, equipment tracking in engineering firms, hospital filing systems, etc.

3. How RFID System Works

Most RFID systems consist of tags that are attached to the objects to be identified. Each tag has its own “read-only” or “rewrite” internal memory depending on the type and application. Typical configuration of this memory is to store product information, such as an object’s unique ID personal details, etc. The RFID reader generates magnetic fields that enable the RFID system to locate objects (via the tags) that are within its range. The high-frequency electromagnetic energy and query signal generated by the reader triggers the tags to reply to the query; the query frequency could be up to 50 times per second. As a result communication between the main components of the system i.e. tags and reader is established.

If the reader is on and the tag arrives in the reader fields, then it automatically wakes-up and decodes the signal and replies to the reader by modulating the reader’s field. All the tags in the reader range may reply at the same time, in this case the reader must detect signal collision (indication of multiple tags). Signal collision is resolved by applying anti-collision algorithm which enables the reader to sort tags and select/handle each tag based on the frequency range (between 50 tags to 200 tags) and the protocol used. In this connection the reader can perform certain operations on the tags such as reading the tag’s identifier number and writing data into a tag. The reader performs these operations one by one on each tag.

4. Components of an RFID System

This RFID system allows to deduct the objects (tag) and perform various operations on it. The integration of RFID components enables the implementation of an RFID solution. The RFID system consists of following five components (as shown in Figure 1):

- Tag (attached with an object, unique identification).
- Antenna (tag detector, creates magnetic field).
- Reader (receiver of tag information, manipulator).
- Communication infrastructure (enable reader/RFID to work through IT infrastructure).
- Application software (user database/application/ interface).

5. Tags

Tags contain microchips that store the unique identification (ID) of each object. The ID is a serial number stored in the RFID memory. The chip is made up of integrated circuit and embedded in a silicon chip. RFID memory chip can be permanent or changeable depending on the read/write characteristics. Read-only and rewrite circuits are different as read-only tag contains fixed data and can not be changed without re-program electronically. On the other hand, re-write tags can be
programmed through the reader at any time without any limit. For example, in the case of the credit cards, small plastic pieces are stuck on various objects, and the labels. Labels are also embedded in a variety of objects such as documents, cloths, manufacturing materials etc.

There are two types of tags (active and passive) are being used by industry and most of the RFID system. The essential characteristics of RFID tags are their function to the RFID system. This is based on their range, frequency, memory, security, type of data and other characteristics. These characteristics are core for RFID performance and differ in usefulness/support to the RFID system operations.

5.1 Tag Frequencies

The range of the RFID tags depends on their frequency. This frequency determines the resistance to interference and other performance attributes. The use/selection of RFID tag depends on the application; different frequencies are used on different RFID tags. The following are the commonly used frequencies:

- **Microwave** works on 2.45 GHz, it has good reader rate even faster than UHF tags. Although at this frequency the reading rate results are not the same on wet surfaces and near metals, the frequency produce better results in applications such as vehicle tracking (in and out with barriers), with approximately 1 meter of tags read range.

- **Ultra High Frequency** works within a range of 860-930 MHz, it can identify large numbers of tags at one time with quick multiple read rate at a given time. So, it has a considerable good reading speed. It has the same limitation as Microwave when is applied on wet surface and near metal. However, it is faster than high frequency data transfer with a reading range of 3 meters.

- **High Frequency** works on 13.56MHz and has less than one meter reading range but is inexpensive and useful for access control, items identifications on sales points etc as it can implanted inside thin things such as paper.

- **Low Frequency** works on 125 kHz, it has approximately half a meter reading range and mostly used for short reading range applications.

![Fig 1 Components of an RFID System](image)
6. Antennas

RFID antennas collect data and are used as a medium for tag reading. It consists of the following: (1) Patch antennas, (2) Gate antennas, (3) Linear polarized, (4) Circular polarized, (5) Di-pole or multi-pole antennas, (6) Stick antennas, (7) Beam-forming or phased-array element antennas, (8) Adaptive antennas, and (9) Omni directional antennas.

7. RFID Reader

RFID reader works as a central place for the RFID system. It reads tags data through the RFID antennas at a certain frequency. Basically, the reader is an electronic apparatus which produce and accept a radio signals. The antennas contains an attached reader, the reader translates the tags radio signals through antenna, depending on the tags capacity. The readers consist of a build-in anti-collision schemes and a single reader can operate on multiple frequencies. As a result, these readers are expected to collect or write data onto tag (in case) and pass to computer systems. For this purpose readers can be connected using RS-232, RS-485, USB cable as a wired options (called serial readers) and connect to the computer system. Also can use WiFi as wireless options which also known as network readers. Readers are electronic devices which can be used as standalone or be integrated with other devices and the following components/hardware into it. (1) Power for running reader, (2) Communication interface, (3) Microprocessor, (4) Channels, (5) Controller, (6) Receiver, (7) Transmitter, (8) Memory.

8. Storing Tamil font using PHP:

By default mysql supports many european languages, Since unicode character (UTF-8) support implemented in mysql it allows us to store many of the indian (Asian) languages. Mysql supports Gujrathi, Hindi, Telugu and TAMIL among too many languages in the subcontinent. Let’s consider TAMIL language and workout:-

To store & search tamil character sets in MySQL table, first of all we need to create a table with character set UTF-8.

```
CREATE TABLE multi_language
(
    id INTEGER NOT NULL AUTO_INCREMENT,
    language VARCHAR(30),
    characters TEXT,
    PRIMARY KEY(id)
) ENGINE=INNODB CHARACTER SET = utf8;
```

```
INSERT INTO multi_language VALUES (NULL, 'English', 'welcome');
INSERT INTO multi_language VALUES (NULL, 'Arabic', 'ﺃﺩﺝﺏﻝﻙﻱﻁﺡﺯﻭﻫ

ﻥﻡ');
INSERT INTO multi_language VALUES (NULL, 'Tamil', 'Tamil character letters');
```

If the result shows like the???? then properly in the system (windows XP) need to install the extra language support tool by enabling the following options,

```
Control Panel -> Regional and Language Option -> Languages -> Install files for complex script and right-to-left languages (including Thai)
```
While fetching the row using PHP, for displaying the Multilanguage content properly it is needed to include the Meta tag like,

```html
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=utf-8"
```
if needed the add `mysql_query('SET character_set_results=utf8')` in the php code before fetching the record. Using this software the database could be created in tamil for the citizen information. i.e. E-card information. E-card information includes personal information, driving license, medical policy, PAN number, Passport number, voter ID, Credit card, Ration card and many more information in Tamil language.

11. Conclusions

This study has identified and explained the nature of RFID technology evolution with respect to RFID applications. RFID technology will open new doors to make organizations, companies more secure, reliable, and accurate. The first part of this paper has explained and described the RFID technology and its components, and the second part has discussed the main considerations of Tamil font in the PHP my SQL. The paper considers RFID technology as a means to provide new capabilities and efficient methods for e-governance. The implementation of e-card has not been without challenges, and some continue to challenge the use of contactless technology and other identity documents. This paper analyzed the major current and potential uses of RFID in identity documents. This paper also gave a model for E-card using RFID technology and E-card information could be stored in the tamil font database.

References

1) Yingjiu Li, Member, Vipin Swarup, and Sushil Jajodia, Fingerprinting Relational Databases:Schemes and SpecialtiesIEEE, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 2, NO. 1, JANUARY-MARCH 2005 pp 34 to 45.


Recommended Approach for e-Governance in Tamil Nadu using MPMLA.IN

Syed Hussain
(admin@auroinfotech.com)

Core Objective
The topic of this presentation is based on my experiences with MPMLA.IN website http://mpmla.in/. The objective is to present this working model as a sample way to implement eGovernance using a mixture of technologies related to Tamil language to help the elected representatives to interact with general public and also to use it to effectively improve the quality of life of common man.

Current Implementation and Limitations
Currently, MPMLA.IN covers all the MPs and MLAs in every single parliamentary and assembly constituency in India. The website is an unbiased effort to facilitate communication between people and their respective MPs/MLAs in India. The people can put across their grievance/appreciation to their MP/MLA and the MP/MLA can choose to respond and address the issue. There is no bias or political affiliation.

Key Features:
Some of the key features available for the visitors to MPMLA.IN today are:

- Easy identification of MPs and MLAs from the website by State, Parliamentary Constituency and Assembly Constituency.
- Search by MP, MLA, Party, Constituency or Politician.
- Easy entry of problems using English language as well as typing in English automatically converted to Tamil.
- Every feedback and problem reported is validated by moderators before publishing on the site.
- Top Five Issues List: A Top Five Issues list is available across each constituency thus helping the elected representatives to focus on the core problems reported in their constituency.

Major Limitations
Some of the main limitations we identified are:

- People who are not comfortable in English are forced to express their thoughts in English.
- People who know English and Tamil try to type in Tamil language using English alphabets.
- People who only know Tamil cannot express their concerns on the website
- People who do not have a computer cannot express their concerns

The idea behind this paper is to recommend an enhanced version of MPMLA.IN as a viable model for e-Governance using various technology advances in Tamil. This model also addresses the major limitations mentioned above, along with many additional features. The following page has a screen
shot of how MPMLA.IN looks today. This is followed by sections explaining the new technology advancements recommended in the near future and also the long term plan for the same.

Screen shot of MPMLA.IN website as of 14-Apr-2010

Planned Implementation for eGovernance using MPMLA.IN

The following section gives a full scope of how the existing model can be adopted in Tamil Nadu by leveraging technology advancements in Tamil.

Overview

MPMLA.IN can be adopted by Government of Tamil Nadu as the central system for general public to report and track their day to day non-emergency issues. This system will facilitate better tracking and monitoring of any problem across the state.

The following are some of the core features to be integrated in the immediate future in MPMLA.IN:
Elected Official Login Management

- **Main Login Module:** Every elected official will be able to access the problems reported by the visitors using various methods such as computer, mobile devices, etc.

- **Communication Module:** The elected official can use this method to communicate their thoughts and ideas over a period of time.

- **Delegation:** Elected officials can also delegate their access to their subordinates – this way they do not have to always be present to type all the responses. The responses can be typed up by the delegates and submitted. This also makes it easy for the elected representative to quickly give follow ups based on each issue.

Problem Management Process

- Feedback and Problem Submission

  - Computer – Typing in Tamil or English – Any one can access the MPMLA.IN website and type their problem in Tamil or English and report the issue.

  - SMS – in Tamil or English – Any person can SMS their issues to a number from their mobile device.

  - Voice Mail – in Tamil or English – The system allows users to call a toll free number and record their messages. This message will be automatically converted into text using a voice to text software.

  - Picture – Users of the system can also choose to upload any photo that represents a problem better. For eg. If a road is in a poor condition, a visitor can take a photo and upload it from their mobile device or from a computer. This image will be associated with the problem mentioned by the visitor.

Problem Command Centre

A Problem Command Centre has to be established in order to review, analyse and also route the problem to its right officials. The Problem Command Centre also takes care of the following items:

- **Feedback Parsing:** Every feedback is automatically parsed and allocated based on the following parameters:
  - MP or MLA under whom the problem was reported
  - Area under which the problem was reported
  - Nature of the issue (eg. Road, Water, etc)

- Problem Number Allocation: Every feedback identified as a problem is allocated a unique Problem Number and tracked on the website. This Number helps track every issue from initiation to closure.

- **Service Level Agreement (SLA) with Escalation Procedure**
  - SLA Time commitment: Every elected representative will commit to a specific time within which an issue will be replied to. This commitment will be officially displayed on the site and
communicated to the problem reporter. Whenever a reply is made by any official, it will include the following:

- The current status of the issue
- Description of what has been done so far and what is planned
- When will the next update be provided on this issue
- Escalation Procedure: A well defined escalation procedure will be implemented in place which says who will be contacted if a person does not respond on time. This will make sure that if the issue crosses its SLA time, it will be escalated to the next higher official up in the chain of command. This will continue until the problem is addressed and resolved.

- **Periodic Analysis and Reporting**
  - **Dashboards:** The plan is to provide extensive dashboard for each constituency so that the entire details about each issue are available at any time to any one visiting the constituency page. This also helps each official to understand key issues, their categories, etc.
  - **Reports:** Various reports will be generated to show problems that are unattended for a long time, repeated patterns, etc. This can be generated for any given time period, thus helping us to understand the history of an issue and identify repeatable pattern of issues, so that we can find a permanent fix for the same.
  - **Comparison Charts:** Comparison Charts help us to understand how one constituency is handled with respect to another and how each elected official has performed over a period of time. This kind of metrics provide competitive analysis, based on which we can even arrive at the Constituency Index which will help us determine how a constituency fares in comparison to a similar constituency anywhere else in India.

**Benefits for Elected Representatives**

- Additional Communication Medium: This platform provides the elected representative an additional communication medium to interact with the citizens.
- Transparent View of Status: This platform provides the responsible officials with a full view of all the issues, their status and also how they are performing when compared to their peers.

**Benefits for Citizens**

- Suggestion Box at Home: People do not have to leave their homes and can directly report the feedback or issues using their computer or their cell phones.
- Transparency and Accountability: This model gives complete transparency on each issue reported by the people. This also improves the accountability on the part of the officials as they are aware that their performance is tracked publicly by every one.
- Social Awareness and Involvement: This form of interaction will automatically improve social awareness and improve the level of involvement for citizens in contributing towards improving their areas.
Future Enhancements

All Elected Representatives

One way of expanding MPMLA.IN would be to include the elected representatives at all levels starting from local municipality elections. This way every single elected representative will be able to have a full scope of their local issues as well as how it rolls up under the larger platform. This also provides full transparency and accountability at all levels.

All Government Officials

A second way of expanding MPMLA.IN would be to include all the Government officials belonging to various departments such as Public Works, Transportation, Industries and Commerce, etc. This will make sure there is complete transparency across all the departments that handle the day to day affairs of the people.

Conclusion

This approach is recommended to provide the simplest approach for any one in Tamil Nadu to leverage this system in order to have better reach to their elected representatives. A proper execution of this plan will ensure that there is better progress in every single area where there is a lack of attention today – thus leveraging various technologies to solve day to day problems for the common public.

References

- Election Commission of India Website
- Google Maps
- Google Transliteration Tool
- Vlingo Speech to Text Software
- Elgg Social Networking Software
சிறந்த நாள் அடுத்துக்கும் 
குறிப்பிட்டியும் எங்குறு

சிறந்த நாள் அடுத்துக்கும்
குறிப்பிட்டியும் எங்குறு

வாழ்க்கையில் பதவி அமைந்து
வாழ்க்கையில் பதவி அமைந்து
1990ஆம் இலக்கில் அரசு அறிவித்தது 24 வருட இரண்டு மாதங்களுக்கு முன்னர் புதுச்சூழலான அரசியல் வடிவங்கள், எழுதிய விளக்கை அரசியல் (விளக்கை அரசியல்) அனுப்புகின்றது. அகத்தியாக அரசியல் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. ஆனால் அத்தோறு, அரசியலின் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. ஆனால் அத்தோறு, அரசியலின் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. 

1990ஆம் இலக்கில் அரசு அறிவித்தது 24 வருட இரண்டு மாதங்களுக்கு முன்னர் புதுச்சூழலான அரசியல் வடிவங்கள், எழுதிய விளக்கை அரசியல் (விளக்கை அரசியல்) அனுப்புகின்றது. அகத்தியாக அரசியல் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. ஆனால் அத்தோறு, அரசியலின் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. 

1990ஆம் இலக்கில் அரசு அறிவித்தது 24 வருட இரண்டு மாதங்களுக்கு முன்னர் புதுச்சூழலான அரசியல் வடிவங்கள், எழுதிய விளக்கை அரசியல் (விளக்கை அரசியல்) அனுப்புகின்றது. அகத்தியாக அரசியல் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. ஆனால் அத்தோறு, அரசியலின் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. 

1990ஆம் இலக்கில் அரசு அறிவித்தது 24 வருட இரண்டு மாதங்களுக்கு முன்னர் புதுச்சூழலான அரசியல் வடிவங்கள், எழுதிய விளக்கை அரசியல் (விளக்கை அரசியல்) அனுப்புகின்றது. அகத்தியாக அரசியல் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. ஆனால் அத்தோறு, அரசியலின் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. 

1990ஆம் இலக்கில் அரசு அறிவித்தது 24 வருட இரண்டு மாதங்களுக்கு முன்னர் புதுச்சூழலான அரசியல் வடிவங்கள், எழுதிய விளக்கை அரசியல் (விளக்கை அரசியல்) அனுப்புகின்றது. அகத்தியாக அரசியல் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது. ஆனால் அத்தோறு, அரசியலின் வடிவங்கள் தொடர்ந்து அறிவித்திட்டுதலுக்கு வரும் வகுப்பிடப்பட்டது.
நான் தெரிந்தது அப்படியானது... 

தெரியும் போது வாசிக்கிறீர்கள் பொருளியல் பண்டைய தமிழ் தமிழ் என்று பண்டைய தமிழ் என்று. பொருளியல் பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று. பண்டைய தமிழ் என்று பண்டைய தமிழ் என்று.


1) பதில் பற்றிய விளக்கத்தை வைத்து?

2) பதில் பற்றிய விளக்கத்தை வைத்து?

3) பதில் பற்றிய விளக்கத்தை வைத்து?

நேர்ப் பற்றிய விளக்கத்தை வைத்து...
நல்ல மாணவரையிட்டு எழுதியவர்கள். ஆண்டு ஆண்டும் மேற்கொண்டு இருந்தனர் ஆண்டு மேற்கொண்டு இருந்தனர். முதல் ஆண்டு பதிலிட்டினான் முதல் ஆண்டு பதிலிட்டினான். முதல் ஆண்டு பதிலிட்டினான் முதல் ஆண்டு பதிலிட்டினான்.

2008-ஆம் ஆண்டு பெற்றிய இரண்டு பாதுகை

குதிரை


--- இடைவனவும் குறியீடு வாழ்த்துப்பட்டு பாதுகை

அவர்கள் வரும்போது சொல்லாய்வானது பிறந்தவன் பதிலிட்டினான். முக்கிய கருத்துகளை இருந்தவர்களின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு. அவர்களின் சொல்லாய்வானது பதிலிட்டினான். முக்கிய கருத்துகளை இருந்தவர்களின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு அதிகாரத்தின் குறிப்பிட்டு.

அவர்களின் வகையை நோக்கியபோது 'அவர்களின் வகை' என்று பதிலிட்டினான் 'அவர்களை' பதிலிட்டினான் கிளைகள் என்று பதிலிட்டினான் மேலாளர்களை பதிலிட்டினான். அவர்கள், Powerpoint presentation வழியாக மற்றும் பல்வேறு படங்களிலும் தொடர்பு பதிலிட்டினான். ஆண்டு ஆண்டு வரும்போது பதிலிட்டினான் என்று பதிலிட்டினான் என்று பதிலிட்டினான். என்று பதிலிட்டினான் என்று பதிலிட்டினான். என்று பதிலிட்டினான் என்று பதிலிட்டினான். என்று பதிலிட்டினான் என்று பதிலிட்டினான். என்று பதிலிட்டினான் என்று பதிலிட்டினான்.

சாத்தியம் குறிப்பிட்டு:


சாத்தியம் குறிப்பிட்டு கூறியவர் இருவரிடையே இருவரிடையே. சாத்தியம் குறிப்பிட்டு கூறியவர் இருவரிடையே இருவரிடையே.
கூட்டம் விளக்கத்தில் செய்யப்பட்டுள்ள உடல் கருத்துக்காட்சிகள் கூற்றுக்காட்சிகள் குறிப்பிட்டுள்ளன. அல்லது அத்துறையில் செய்யப்பட்டுள்ள உடல் கருத்துக்காட்சிகள் கூற்றுக்காட்சிகள் குறிப்பிட்டுள்ளன.

Facebook, Twitter, Podcast விளக்கம் மற்றும் குறிப்பிட்டுள்ள உடல் கருத்துக்காட்சிகள் கூற்றுக்காட்சிகள் குறிப்பிட்டுள்ளன. "This is written in Tamil."

Google word document PowerPoint பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளக்கம் பட்டியல் அரசு சந்திக்கு விளங்கி படுத்தப்பட்ட உடல் கருத்துக்காட்சிகள் கூற்றுக்காட்சிகள் குறிப்பிட்டுள்ளன.
 TNM நூற்றாண்டு நிறுவனம் நிறுவனம்

குறிப்பிட்டுக்கும் சேவைகளின்

வருமான வருமானங்கள்

தசையும் குறியாக மிக்கதாக

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துறையின்

தமிழ் தொழில்நுட்பத்துறையின்

வருமானத்துrif
இன்றும் பொருளாய் ஆளுநர்கள் மாநிலத்தில் இணைந்து, 1983ல் இணைந்து விள்ளையாட்டுத் தலைமையகத்தில் ஆண்டனர். இது இல்லாமல், முதல் தலைமையகான நூற்றாண்டு மீண்டும் இணைந்துபட்டுள்ளது. இப்போது குழுவாக கூட்டாக, தொடர்கூட்டாக, ஆண்டுகூட்டாக விள்ளையாட்டு உற்பத்தியில் நோக்குநிறுத்துகிறோம். இல்லாது இன்றும் தலைமையகான பார்வை அறிக்கை இருப்பதாயிருக்கும். இணைந்து விள்ளையாட்டுக்கு எளிதாக முடியும் மீது அவர்கள் நோக்குத்தில் அனைத்தும் விள்ளையாட்டு பார்வை நோக்குநிறுத்துகிறது. இந்த ஒன்றியம் விள்ளையாட்டு தொடர்பில் இரண்டு நோக்குநிறுத்துகிறது (Standard): இணைந்து விள்ளையாட்டு நோக்குநிறுத்துக்குப் பார்வையாளர்களை பற்றியுள்ளது.
புகழ் பார்வைதல் எனக்கு நேர அணுக்களுக்கு அடையாளம் ஆக்கக்கார். ஜீவனம் காணவல்லத்
பூர்த்தி துவங்கிய விலைவில் குறுகுகளை ஆசிரியர்களின் தாழ்வுகள் மற்றும் பொருள்பாடுகள் வேதியியல். அந்த
இலங்கையான பொருள்பாடுகள் தான் செய்ய வேண்டும். ஜீவனத்தில் செய்யும் பொருள்பாடுகள்
விலைவில் வேறுபாடுகள் தன்னாலேயே குறுகுகளை குறாவியதை விளக்கப்பட்டது.

1326 வருடத்திலிட்டு
அடையாள பாதுகாப்பு வளர்ச்சியின் காட்டுப்பாடுகள்

696
1. ஒவ்வொரு தொடர்பில் அதன் வழிகோட்டையும் இருப்போக்கில் அறிக்கை வழிபட்டிருக்கும். 2. ஒவ்வொரு தொடர்பில் அதன் வழிகோட்டையும் இருப்போக்கில் அறிக்கை வழிபட்டிருக்கும். 3. ஒவ்வொரு தொடர்பில் அதன் வழிகோட்டையும் இருப்போக்கில் அறிக்கை வழிபட்டிருக்கும்.
தமிழ்நாடு மாநிலத்தில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள், அமைந்து வசதிகளைக் குறிப்பிட்டுகிறது. அந்னாடு கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்பவலவைக் குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது. அந்நாட்டில் கட்டடக்கல் உருவாக வர்த்தகப் பொருட்கள் என்று குறிப்பிட்டுகிறது.
(1) 2003-2005 காலப் பட்டியல் e-Srilanka அயலுங்கள் ருகியாக ICT சார் நிறுவப்படுவதுடன் அதிகமாக. இது சுற்றுடன் அனைத்து தமிழிலுள்ள கால் சுற்றுக்கும் போராட்டுக்கும் மேலானவையாக காட்டும் போதும் முக்கியத்துவமான படிகமும் போராட்டுக்கும் மேலானவையாக காட்டும் போதும் முக்கியத்துவமான படிகமும் தேவையாக (செக்சுமின் 01(1))

(2) பாக்ஸ்பெஸ் தொகை நீர்நடவுக்காக நிறுவப்பட்டு தொகை முறையே தொகைப்பட்டு தொகைப்பட்டு காலம் தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 02, 03, 04, 05(3))

(3) பாக்ஸ்பெஸ் நீர்நடவுக்காக நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 06, 07)

(4) பாக்ஸ்பெஸ் (emathumozhal) காலம் e-Srilanka முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 08)

(5) பாக்ஸ்பெஸ் (emathumozhal) முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (LASKAPP) முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 09(6))

(6) LK Domain Registry (இந்த ஏல் பாக்ஸ்பெஸ் முறையே நிறுவப்பட்டு) காலம் ICANN முறையே நிறுவப்பட்டு அமெரிக்க இரையையுள்ள பாராட்டுக்கும் தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 10, 11)

(7) பாக்ஸ்பெஸ் (NENASALA) காலம் e-Srilanka முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 12)

(8) பாக்ஸ்பெஸ் (இந்த ஏல் பாக்ஸ்பெஸ் முறையே நிறுவப்பட்டு) காலம் முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 13)

(9) பாக்ஸ்பெஸ் (GIPS) காலம் e-Srilanka முறையே நிறுவப்பட்டு தொகையாக முறை�ே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு தொகையாக முறையே நிறுவப்பட்டு (செக்சுமின் 14)
1. e-Government: The Singapore- Arun Mahizhnan & Narayanan Andiappan Tamil Internet 2002, California, USA
2. அரசு விளையாட்டு; கலையான, உலகப் பல்கலைக் கழகம் மதுரைப் பத்மபன் விளையாட்டு பணி-பேரணி -ICTA, December 2009
01: http://www.icta.lk
02: http://www.nie.lk
03: http://www.nie.sch.lk/ebook/s11tim33.pdf
04: http://www.nie.sch.lk/ebook/s12syl41.pdf
05: http://www.Scoolnet.lk
06: http://www.nodes.lk
07: http://uoj.nodes.lk
08: http://www.ematnumozihal.lk
09: http://www.lakapps.lk
10: http://www.nic.lk
11: http://www.idns.lk
12: http://www.nenasa.lk
14: http://www.gic.gov.lk
15: http://www.gov.lk
16: http://www.documents.gov.lk
17: http://www.happylife.lk
19: http://www.lakapps.lk/moodle
E-Governance Initiatives have resulted in improving core infrastructure and providing better services to citizens. IT has come a long way in changing the lives of the common man at the grassroot level. National Informatics Centre has been providing informatics support to Central Ministries, State Government and District Administration. It has gained deep understanding of governance issues which have paved way for successful designing and implementation of many e-Governance projects like Land Records, Registration, AGMARKNET, Examination Results, e-Post, Passport etc. The impact of government initiatives has been observed in all sectors i.e. administration, agriculture, rural development, judiciary, health, education, telecommunication and transport.

In Tamil Nadu, the major E-Governance projects implemented are explained in detail.

TAMIL NILAM:

Tamil Nadu Infosystem on Land Administration and Management (Tamil NILAM) is a major e-Governance initiative taken by the Revenue Department of Govt. of Tamil Nadu. This software system is developed to computerise the Land Records System in Tamil Nadu. Tamil NILAM is currently implemented in all the rural taluks in the State. It handles all the transactions relating to Land Records in the State.
Objectives of Tamil NILAM:
- Delivery of all possible Citizen-centric e-services
- The entire contents are in TAMIL
- Issue of Chitta Extract (Record of Right) / A Register Extract / Adangal Extract to citizens
- Creation of Master database storing plot wise and owner wise details of land, crop, revenue, etc
- Generation of periodic reports through the computerized system.
- Improved and efficient service
- Easy maintenance and updates of Land Records
- Transparent administration
- Availability of information to public through Touch Screen Kiosk
- Exchange of data to other departments such as Sub Registrar Office, Agriculture Department etc.,

VAHAN & SARATHI:

Vahan and Sarathi are application Software developed for State Transport Authority, Tamil Nadu.

Vahan is for processing all transactions related to Vehicles and Sarathi is for processing Driving Licence and related activities. Vahan can be used to issue of Registration Certificate and Permits. Sarathi can be used to issue a Learners Licence, Permanent Driving Licence, Conductor Licence and also Driving School Licence to the applicant. The system was implemented on pilot basis in in RTO Chennai (North). The system was then approved for implementation in all other RTOs in Tamil Nadu. The systems have now been implemented in 71 offices.
Registration Department (STAR):

**Property Registration** STAR (Simplified and Transparent Administration of Registration) is implemented in 450 Sub-Registrar Offices. Guideline Values of more than 2 crore Survey Subdivisions are hosted on Internet for public access. The system has enabled citizens to apply for Encumbrance Certificate online. The registration details are maintained in Tamil. The Sub-Registrar Offices and Taluk Offices in 40 taluks are connected using LAN for sharing the data of each other.

EMPLOYMENT ONLINE:

The Professional Employment Exchange Office (PEEO) caters to the employment needs of professionals registered in Tamil Nadu for the entire state. The PEEO office comes under the Directorate of Employment & Training, registers the candidates seeking employment opportunities and recommends to various departments / offices requesting suitable professionals. The PEEO has decided to open their databases for access to the private sector employers, to create more avenues for the registered candidates. The mushrooming growth of private sector employment in this information era necessitates to make online web access of the entire database of registered candidates. In this scenario, the department has created an online web portal for the welfare of candidates and thrown open the entire online database for private sector employment opportunities. The website is operational from June 2003.

The objectives of this web site are as follows:-

- To develop a Data Bank of highly qualified, marketable candidates from the Live Register of the Employment Exchanges in Tamil Nadu with the accent on persons with professional, executive and engineering diploma and degrees.
- To allow the private sector employers easy access to this database to fill up vacancies arising in their establishments and to offer facilities for screening and short-listing of prospective employees.
- To provide online information on application dead-lines, track careers and future trends in employment.

Intra Management Information System Portal for Tamil Nadu Pollution Control Board (TNPCB):

This Intranet based Monitoring System has been implemented to create database on all types of Industries (Profiles), the Consents Details and Hazardous Waste Authorisation Details of various types of Industries in Tamil Nadu. This web application is used as an efficient Monitoring System by TNPCB Corporate Office, Chennai. Different types of MIS Reports are being generated using this IntraMIS both at State and District levels. Each TNPCB District Office updates AIR and WATER Pollution related data for all the Industries (taken from the “Application for Consent” submitted by these Industries under Central Act 14 of 1981) functioning in its Jurisdiction. Each industry is identified by a unique File Number assigned by the respective TNPCB District Office. Industry Profile Data, Consents related data and Hazardous Waste related data are being captured for each industry using this web application.

E - Karuvoolam : Automated Treasury Bill Passing System:

Automated Treasury Bill Passing System (e-Karuvooolam) is aimed at automating the existing manual billing system of Treasury Department. The Workflow based Systems developed for the Treasury Department, enables to capturing the data starting from bill submission stage at the counters.
Treasury is the “Bank of Government”, functioning with the objective of maintaining all transactions to the government and sending reports to the Accountant General. Any transaction related to government is performed in the form of a bill. Initially a bill is submitted to the counter of a Sub Treasury / District Treasury / Pay and Accounts Office through a messenger. The bill goes through a phase of approval. This phase is called auditing the bill. Auditing clerk can reject the bill or can approve the bill. If the bill is approved, it is sent to the cheque release counter. The clerks at cheque release counter handover the cheque to the messenger if Treasury type is banking. The application software developed by NIC provides online environment with systems available on all working tables starting from the bill submission counter. The officials can process the bills online and take action for passing / rejecting the claim.

E - Governance at Madras High Court:

Systems in use at Madras High Court deals with

- Posting of Case Status of Madras High Court on Internet
- Display systems are installed at principal bench and Madurai bench to know the status of cases being heard at court halls. 35 court hall display and 6 composite display are installed at Principal bench.
- Certified copies of Final orders, Bails/Anticipatory Bails and interim application orders are entered and issued through the system. The computerized system has reduced the time delay in issuance of copies to the litigants. The Interim application orders are issued through the system since 26th Feb 2001.
- Systems for posting Case Status have been developed. The case details have been hosted on Internet.
- Interactive Voice Response System is installed at Madras High Court to know the Case Status through telephone.
- Touch Screen Kiosk is installed at Madras High Court for public dissemination.
- Daily Cause Lists are being prepared using the system. Systems are installed at Filing Counter, Posting sections for preparation of cause list.
Daily Cause Lists are being hosted on Internet. Number of hits per day is more than 7000 on working days. As per the statistics available, Madras HC is receiving maximum number of hits.

 Reported Judgments are being hosted on Internet. More than 1500 users visit the Madras HC Judgment site every day.

 Statistical reports relating to disposed cases statements are prepared regularly.

 Information Centre functions at Madras High Court for the benefit of Litigant public to know the status of cases filed at Madras HC. Around 700 case enquiries are received at the enquiry counter every day.

e-Attendance Monitoring System:

This Application is being used by the “e-Governance Cell” of the DoTE, to monitor the Attendance Details of about 2.71 Lakhs Diploma Students studying in about 340 Polytechnic Institutions all over the State. Monthly Attendance details of every student are to be entered by the respective Institution in the first week of the following month. Duly signed hard copies of the Branch-wise Attendance Registers are to be generated from this application and to be sent to DoTE for filing purposes.

 Any time, a student can login to view his/her Attendance Details for the current Academic Year. At the end of the semester, Hall Tickets will be generated only for the Eligible Students taken from this e-Attendance Database. Thus this Intranet application helps the DoTE to streamline and to bring transparency to the Examination Processing System in the DoTE.

Single Window Counseling System:

This system has been implemented for admissions to more than 390 Teacher Training Institutes in the State.

Common Integrated Police Administration:

The CIPA software is designed and developed to maintain the details pertaining to all the activities of the Police Stations relating to Crime and Criminals. The system provides required information to the higher levels periodically and as and when required. The system also generates various statutory reports for the smooth functioning of the police station. The ultimate goal of the computerization would be an integrated networked system with state of the art hardware and software in place for police access and use the Information in their day to day work and to take decisions. The software is a total work-flow system having the following three major modules along with reports and queries viz., Registration, Investigation and Prosecution.

Agricultural Marketing Information System Network (AGMARKNET):

The AGMARKNET project is sponsored by Directorate of Marketing and Inspection (DMI), Ministry of Agriculture, and implemented by National Informatics Centre. It aims to link all Regulated Markets, State Agricultural marketing Boards / Directorates and DMI regional offices located throughout the country, for effective information exchange on market prices.

The AGMARKNET website (www.agmarknet.nic.in) is a G2C e-governance portal and available in Tamil for the markets in Tamil Nadu, caters to the needs of various stakeholders such as farmers, industry, policy makers and academic institutions by providing agricultural marketing related
information from a single window. It facilitates dissemination, over web, of the daily arrivals and prices of commodities in the agricultural produce markets spread across the country.

**PATRAM: Postal Accounts Transaction Maintenance Software:**

The PATRAM (Postal Accounts Transaction Maintenance) Software has been designed to maintain all accounting details currently being maintained in various registers. It emulates the auditing functions carried out by the Cash Certificates Section. The PATRAM software consists of 18 modules. This system was developed taking Tamil Nadu Postal Circle as the pilot site and replicated at other sites in the country subsequently.

**System for Civil Supplies & Consumer Protection Department:**

**Family Card Maintenance System**

A workflow based Family Card Maintenance System has been developed for the Department of Civil Supplies & Consumer Protection Department, to assist in maintenance of consumer database. Citizens interact with the AC/TSO offices for issue of new family cards, various types of alterations to existing cards, issue of surrender certificates, etc. The software provides a work flow application to manage these services at the AC/TSO offices. All the TSO and Assistant Commissioners’ Offices, are provided with internet connectivity. The card database in the central server at the office of the Commissioner of Civil Supplies, is kept updated through online interaction with all the AC/TSO offices. This provides accurate data for computing card-wise entitlement, which is used for preparation of Shop-wise, Taluk-wise, and District-wise monthly allotment statements.

**SIM Card Management System (BSNL):**

The SIM Card Management System is an intranet application designed - to handle sale transactions at Customer Support Centres; to carry out all activities related to SIM card/ GSM number distribution/allotment by the headquarters; and to maintain accounts; through a centralised database. The SIM Card Management system uses intranet technology to give a cost-effective solution that delivers powerful and robust results. By connecting Customer Support Centres, dealers, and distributors with the Mobile Services headquarters, through a centralised server, information is managed effectively and efficiently.

**Chennai Corporation:**

Ten Zonal Offices of Corporation of Chennai make use of the intranet based systems developed for Property Tax collection, Birth/Death Certificates extraction, Company Tax collection, etc by accessing the central server located at Head Office.

**Text Books Online:**

The entire text books of school children under School Education Department, Tamil Nadu Government is available online.
Geographical Information System (GIS):

Web GIS tools under Open Source environment have been used for developing the TN Maps site http://tnmaps.tn.nic.in. The tools have also been used for generating dynamic maps using Census 2001 data at http://www.census.tn.nic.in.

E-filing of Returns by Dealers for Department of Commercial Taxes:

All the dealers in Tamil Nadu are filing their VAT monthly returns online through the website.

Digital Certificates:

PKI enabled application systems were developed for Madras Export Processing Special Economic Zone (MEPZ SEZ) to facilitate exporters registered with the Zone to file their Applications and Quarterly / Annual Returns through the Internet.

Web Services:

The website of Government of Tamil Nadu is being maintained by NIC and has large number of citizen oriented particulars such as Policy Notes of all the departments, Citizen Charter, RTI documents, Government Orders of Public Interest, Public Utility Forms, Tender Notices, Press Releases, Contact details, Statistical Reports, Acts and Rules documents, etc. The websites of all the districts and those of many departments have been developed and hosted by NIC.

The GPF particulars of more than 5 lakh employees of the State are hosted on the web through the website of Accountant General of Tamil Nadu. Pension processing status is also made available through the website.

Online Registration System is used for the examinations conducted by the Tamil Nadu Public Service Commission. More than twenty thousand applications have been received online in the last two years for various examinations.

Electoral data:

All the 234 Assembly Constituencies are hosted on the web with interface to search the data in Tamil.

Examination Results:

The examination results of the Teachers Recruitment Board are hosted on the web. Tenth and Twelfth Standard examination results were accessed by more than 20 lakh users are hosted every year from 1997 (one of the first States to have exam results online).

E-Governance is a continuous journey and Tamil Nadu has been an active State in this mission mode journey.
E- Governance in Tamil
for Tamil Virtual University

G. Amirtharaj, Software Engineer
Dr. A James, Consultant
Dr. P R Nakkeeran Director
Tamil Virtual University, Taramani, Chennai - 600113, Tamil Nadu, India.
Email: tamilvu@vsnl.com

Abstract:
Information and Communication Technology (ICT) has today become an integral part of governance,
especially in India. ICT is viewed as a tool that will help deliver services in both public and private
sector faster and transparent to end users.

E-governance, as a concept, involves leveraging ICT to streamline the administrative process. It
involves computerization of records, facilitating efficient transactions between various departments
using web portals and other electronic data transfer mechanisms to bring administration more
effectively, to achieve the goals faster and ease.

Around the world various private, public and Government sectors are implementing the e-
governance to make their process transparent, cost effective, providing any where any time access of
services and to reduce the processing time of their work flows.

In order to make the process transparent, cost effective and to reduce the processing time Tamil
Virtual University Management has decided to implement the E-Governance in Tamil Language for
various activities like Employee Management, Time Management, Leave Management, Vendor
Management, Work flow management, Course management, stock management and etc...

Introduction:
This paper speaks about the implementation of E-Governance web based application in Tamil
language for TVU with the Tamil Unicode support. Tamil Virtual University (www.tamilvu.org) is a
Tamil Nadu Government Organization aims at providing Internet based resources and opportunities
for the Tamil communities living in different parts of the globe as well as others interested in learning
Tamil and acquiring knowledge of the history, art, literature and culture of the Tamils. The functions
of TVU include Internet based Educational Programs, Digital Library and Development of Tamil
computing.

As part of e-governance initiative, the TVU management embarked on a paper-less office drive for
introducing transparency and accountability in its internal and external transactions with vendors. As
an initiative TVU has started developing a web based platform independent e-governance tool using
Java, Struts Framework and MYSQL database with Tamil Unicode support.
Towards Implementation of e-governance applications for TVU:

This e-governance application is a web based application developed using Java Technologies (JSP and Servlets) and Struts Framework, MYSQL database as backend, Apache web server and Tomcat application server running in Linux OS with Tamil Unicode Support. The backend database is developed using MYSQL. It is configured to store Tamil Unicode Characters. This application allows to key in Tamil & English characters by toggle between English and Tamil language. The data transactions between client and server are secured by Secured Socket Layer (SSL) enabled Apache web server.

Application Architecture:

The below diagram (Fig-1) describes the architecture of the e-governance application for TVU.

Web Browser – Unicode (UTF-8) supported system with Tamil Unicode font installed to view the application pages by the end users. Apache web server – to serve the html, images and media content. Tomcat Application server – java container to run the JSP/Servlet. All the requests/response should be receive/sent in UTF-8 encoding format to support Tamil Unicode encoding scheme. MYSQL database – UTF-8 supporting database tables to store Tamil Unicode characters. File System – Configuration Management System to store electronic forms of documents/files. External System – Online Payment System (Currently ICICI payment gateway).

Now we will look into the details of each application:

Gateway Portal:

This web portal is the gateway to access all the applications in a single window. Each entity i.e. employees, study centre admin, vendors and students will be created with unique user name and password through user management module. An authorized admin will give access to create/update the user details.

The user can login into this portal and access the different application modules based on their role. This portal can be accessed from anywhere at anytime through web browser with internet connection enabled and also through intranet.
Employee Management:

This application is to manage the TVU employee details and Personal Registry (PR). Admin person can add new employee and update the user detail through the CRUD (Create Read Update Delete) process. A user (TVU staff) login into the e-governance portal can view their details like their personal detail, skill detail, years of experience, designation and his manager hierarchy. The employee will be allowed to update his personal details only.

In TVU, personal registry was maintained by entering their daily activities manually in a physical notebook, routed to the reporting authority through attendant and verified. E-governance facilitates online filing daily status, after implementing this module the employee will enter their work details in the online form and submit will intimate their immediate managers through mail to give their approval. The managers will login to the portal and view the pending PR’s for approval. The manager can give their approval through a single mouse click after verifying the details. Here the employee can generate their work report of his/her own or his/her subordinates. The reports can be generated monthly wise or a whole.

Time Management:
As part of the e-governance, TVU has implemented the thumb impression attendance machine (Bio-metric system) from a third party. Until recently the TVU staff has to put their signature on the physical attendance book daily but it will not have the details of in-time or out-time.

For payroll calculation, a clerical person has to verify manually for each month and based on that monthly salary will be calculated. After implementing the Bio-metric thumb impression attendance machine now the employee has to give his/her thumb impression while entering or leaving the main entrance every time. The data will be collected in SQL server database with all the in-time and out-time details. A standalone reporting tool has been provided to generate/view the reports.
Every two weeks reports will be generated and sent to all the employees through mail. In the near future, the SQL server database will be integrated with our e-governance portal so that each employee can generate their reports of their own or for their subordinates and can view through the portal itself. And also the accounts department can get the consolidated month wise report of all the employees.

**Leave Management:**

Through the e-governance portal the TVU employees can login and apply their leave through online, by selecting from and end date (including a half day), reason for the leave, type of the leave (casual, medical and earn leave) and submit the leave form. The managers will login to the portal and view the pending leave forms for approval. The manager can give their approval through a single mouse click. The employee can apply/approve the leave from anywhere at anytime. This application will be integrated with the Bio-metric attendance machine database so that if an employee forgot to apply leave forms will automatically assigned leave for that particular day. Through this application and authorized admin person can manage the government and other holidays so that this will be reflected in the “MYCLANDER” option for each employee to view the holiday details. Monthly attendance and leave reports for the employees can be generated and other various individual employee reports in Tamil.

**Vendor Management:**

This application is envisage to mange the TVU vendors, project details assigned to the various vendors and quotation/tender details. Through this application a TVU vendor can login to this portal through his provided username and password and can view project details assigned to different vendors, its completion details etc… Here an authorized TVU staff will create online work order form and fill up with the project details like project title, vendor name, starting and end date of a project and submit will go for the approval to the concern approval authority, once it has been approved a final output will be taken with the authorized signature and provided to the vendor. The approval authority will assign a particular project to a TVU employee to follow-up the project. Now that employee and his sub-ordinate will get the full details of the project in the e-governance portal though his login. The employee can track the project and follow up the vendor until the project completes. All the communication, demos, project status, vendor visiting detail, CD’s submission etc… will be stored corresponding to this project and the higher authorities can the view the details at anytime. A detail report of project details and status of the project can be generated based on different criteria like project wise, vendor wise etc...

**Work flow Management:**

Through this application the TVU staff members can submit the online Xerox/Print. Only after approval form authorities regarding details like number of copies and etc…actual printing takes place. The admin will take the Print/Xerox and update the form with actual number of copies taken and submit will go the supervisor for approval. The admin can generate the Xerox/print taken reports based on different criteria like for a particular month, by an employee, on particular device etc...
Course Management:

This application is to manage TVU Study centre across the globe, student, providing permission to access the online examination for a students, generating question papers for online and offline examinations, conducting or monitoring of online examination, mark sheet and certificate generation for students and examination result publishing. The study admin can login to the portal and allowed to register a new student for a course, to pay students course fee, register for a examination, providing permission for offline and online examination for their students, password reset etc…

Fig – Xerox/Print Request Form

TVU student registered for a course will be provided with unique username and password and the student can access the portal to view the examination schedule, online examination samples, question pattern samples for both online and offline examinations, exam results and mark sheet.

TVU examination controller or the authorized TVU employee can register a new student for courses, register a student for examination, password reset, publishing examination results, generating examination report and sent to the study centre.

Physical Stock Management:

Now in TVU assets like furniture, computers hardware and software, Physical library books and course materials, Course CD’s, Electrical accessories, paper and other office accessories are maintained in the physical register. The existing stock management is very cumbersome and difficult to integrate purchasing of available stock with consumables.

Through this e-governance portal an interface will be provided to the authorized TVU staff to add new asset through the online form in different category like furniture, electrical, computer etc…, of type consumable and non consumable and with other details like date of purchase, make, vendor and support/service contact details, warranty and location details. After submitting the online form an asset ID will be generated. Then the higher authority will verify and give his approval. The new items will be tagged with the generated asset ID. If a non consumable is abandoned or consumable has been used, through this application interface the details will be updated for that particular asset and
submit. And this application will allow authorized TVU staff to generate the report based on various criteria.

**Conclusion:**

This paper has discussed about the implementation of e-governance web based portal in Tamil language for TVU with the Tamil Unicode support. By implementing the e-governance application in localized Tamil language using Tamil Unicode encoding scheme, it will remove the language barriers in implementing e-governance application in all the Government departments to facilitate efficient transactions between various departments and public using web portals and other electronic data transfer mechanisms through single window to bring administration more effectively, paper-less office drive and to provide effective services to the public.
13

கலினி அழிக்கலை
Abstract

The paper highlights the usage trends of access to e-resources in Indian Universities, Colleges and Research centre. The preliminary findings proceedings abstract of various conference of last ten years revealed that there is an upward trend. The paper briefly describes open access e-resources used by the various scholars through World Wide Web in India: A citation Analysis


Introduction:

World wide web has created a sea change in providing information and as well as information transfer. Information scientists have undergone a perceptible change and have become IT centered, which is evinced from the publications pattern of articles, citations in journals, seminars, conferences ect. The world wide web provides access to materials that were previously not available to the researchers, students and faculty now able to view, listen to, and read materials that just few years ago were difficult, if not impossible, to access. Images, texts, historical documents, video clips, sound files etc, are now available over the web to anyone that has internet access. Almost any one can put almost anything online, and because that information is not filtered or mediated in any way. Academic, Research and development organizations are supported by well equipped library and information services unit to augment the objectives of the parent organizations. Proliferation of knowledge growth has resulted in information explosion, which has made the users so difficult to find the relevant information of research and development.

The modern libraries are able to provide information services to their patrons more efficiently with applications of information communication technology (ICT) in their environment, which satisfy the currency of information. In the recent years, the publishers and the professional societies are able to add value to their publications by means of providing the full text contents in electronic form. The networked...
environment of the organizations and the availability of internet connectivity give the users and easy access to the full text of the document.

**Purpose:**

The purpose of present paper is to determine the awareness of electronic resources among library users and promote the usage of electronic resources by the students and members of faculty in engineering colleges in Tamil Nadu. Questionnaire was distributed to 210 randomly selected faculty and students in the various research institutions in Tamil Nadu. The researcher collected many articles i.e. print and online. Nearly 54 articles have been received. Through this articles we identified the use of e-resources for various purpose i.e. teach, learn and research. Majority o the user using for research purpose. The 40% of user using for teaching purpose. The considerable percentage of users is using the e-resources for learning purpose. Through this articles reveals that the researcher and faculties are using e-resources through www, so the user’s categories also increase day by day. Because the Government also to allot more fund innovative consortium for the user categories. In future e-resources are a pillar for servicing various information.

**Retrieval Systems:**

The study is an attempt to determine the level of using various types of resources by the research scholar in various University, Scholars felt that about various issue surrounding the electronic resources and whether changing attitude depend upon subject. Further reports have been presented about the factors supporting the growth and academic work with the help of web. The world of academics enjoying the fruits of electronic resources by the way of resources retrieval is highly thankful to the person like Tim Bern Lee for their tremendous effort. The 50% web resources as noted by the respondents gives the positive signs of the growth of IT world in the academic venture. Though the users of the web are having basic knowledge to handle the electronic resources, the institution involved in academic work should provide technical training to their students. The free internet access available in most of the universities may be retained and may be extended to other universities also. The mushrooming of new website development may lead to unreliable resources to the students at large and the research scholars in particular, hence, a board may be constituted to check the validity of resources before it is uploaded in to the web.

**E-Resources:**

**E-Publishing:** Electronic publishing (e-publishing) is 16 years old and the first e-book was published in Germany in 1985. The products of e-publishing are mostly secondary sources, reference materials and the primary periodicals. The problems associated with e-publications are integration with traditional forms, cost of acquisition, collection development, non-availability of selection tools, etc. These challenges should be taken as opportunities and prepared to find out the standards. When the father of the printing press, John Gutenberg, looked towards the future, he could never have dreamed that one small invention could revolutionize the world. The gift of publishing benefited not only the authors who could now create new words, but also the readers who would be exposed to new ideas, and new concepts. The creation of the printing press sparked a new revolution-a literary revolution. Now, with the widespread availability of computers and the internet, technology, enables virtually anyone to self-publish their literary works. E-Publishing is primed to revolutionize the world, by giving every individual the ability
to publish their ideas, stories, books, without the prohibitive costs associated with conventional publishing. E-publishing is a process for production of typeset-quality documents containing text, graphics, pictures, tables, equations etc. In general, it is used to mean any information source published in electronic

E-Books:

Libraries form a vital part of the world’s systems of education and information storage and retrieval. They make available - through books, films, recordings, and other media-knowledge that has been accumulated through the ages. People in all walks of like-including students, teachers, business executives, government officials, scholars and scientists - use library resources in their work. Large numbers of people also turn to libraries to satisfy a desire for knowledge or to obtain material for some kind of leisure-time activity. In addition, many people enjoy book discussions, film shows, lectures, and other activities that are provided by their local library. There are a considerable number of concepts being explored in future libraries research and development. The focus is in distributed and local collections of information objects - in the hybrid library including analogue as well as digital objects and on ways of identifying objects of interest to a user and arranging for the user to access them. The library of the future will be less a place where information is kept than a portal through which students and faculty will access the vast information resources of the world. The library of the future will be about access and knowledge management, not about ownership.

E-Journals:

Advancement in science and technology has changed the knowledge communication in the form of e-media. Order of the day is e-publishing and tradition print sources are replaced by e-sources. An attempt has been made to study the utilization and satisfaction among the users of various institutions and research centers towards e-journals using survey. The modern society is dynamic and complex. The duty of the research scholar towards social change, scientific development and social uplift is undisputable. E-Journals will definitely made its own impact on the users in terms of scholarly communication of research.

The change in information technology and digital library has made us to move from print to electronic media in terms of acquiring information resources and to provide services. Libraries are undergoing rapid changes due to the developments in information communication technology paper based resources are giving way to electronic resources. The CD ROM and E-journals are providing information to the user from various research communities. An attempt has been made in this study to identify the usage of CD ROM s and E-resources among the various research community in Tamil Nadu.

E-Government:

The term “e-government” is used in this paper to denote the concept of using information communication Technology (ICT) as a means to organize and manage the administrative processes of the government, especially the interactive processes between the government and the public. Though ICT has been available widely for more than four decades and many governments around the world have indeed used ICT in certain aspect of government, the concept of e-government is relatively new in the sense mentioned above. Only a handful of governments have progressed to a high degree in harnessing the immense power of ICT in re-organizing their government infrastructure and in serving their citizenry,
and have done so in an efficient and effective manner. E-government is not mere “technologising” of government. It is not just a matter of automating some manual processes nor is it a simple introduction of technology where none existed. E-government requires a fundamental re-thinking of governance itself and, as some have suggested, re-inventing of government. If bureaucracy is the invention of the 19th century, we might say e-government is the invention of the 21st century. E-government re-examines the organizing principles of bureaucracy and governance, re-defines the objectives and deliverables of government and re-deploys the resources available. In this process of re-invention, the basic intent is both refinement of the old and introduction of the new. E-government is NOT throwing the baby with the bathwater.

Consortiums for E-Resources:

Consortium a strategic alliance of libraries with a common interest, not under the same institutional control, but usually restricted to a geographical area, number of libraries, types of materials, or subject of interest, which is established to develop and implement resource sharing among member. This paper discussed about the benefits and limitations of e-resources through consortium. Makes a comparative study of two major library consortium of India i.e. INDEST-AICTE & UGC- INFONET in terms of objectives, governance, members, access pattern, resources etc. Analysis the study and makes suitable suggestions for its improvement. Concludes with the remarks that these consortiums will bring remarkable change in the library scenario and also the educational system of India. In modern library environment, information technology is playing a pivotal role. Today’s technology, interactivity, digital media, expanding network and communications capabilities compel the libraries to move from organizational self-sufficiency to a collaborative mode of acquisition of library materials. Library consortia are one of the emerging tool kit. The consortium should take a lead role in the development of a national strategy for information provision for research in higher education. Librarians should seriously rethink and reinitiate consortium movement like developed countries for maximum utilization of resources at a reduced cost, time and space. It is an encouraging sign that both INDEST-AICET and UGC INFONET consortia are functioning well. These will bring remarkable change in the library scenario as well as the educational system of India.

The UGC - INFONET programme will be a boon to higher education systems in several ways. It is a vehicle for distance learning to facilitate the quality education all over the country. It is a major source for research scholars for tapping the most up-to-date information. It is a medium for collaboration among research guide and research scholars. One of the major developments in libraries and information systems in the past 15 years is the advent and spread of electronic information resources, services and networks mainly as a result of developments in information and communication technologies. The change is basically of physical form where information content is increasingly being captured, processed, stored and disseminated in electronic form. The unique features of the information needs of users in electronic environments relate to the physical form in which information content is made available in electronic information environments. Users normally desire the content to be made available within the constraints of their skills and technological capabilities so that it is possible to access and use the required information content to resolve the felt gap in knowledge.

The E-Journal programme is corner stone of the UGC-Infonet effort which aims at addressing the teaching, learning and research collectively and governance requirements of the universities and research
centers. It would facilitate free access to scholarly journals and databases in all areas of learning to the research and academic community.

This article discusses the need for sharing of electronic resources among Libraries and Information Centers in developing countries. Highlights the importance of 'Library consortia' in this digital era, stating its salient features, functions and responsibilities. Examines in detail the emergence of various resource sharing models in the developing countries with special reference to India.

Emergence of Library Networks in India:

In India library networks actually started due to the initiatives by NISSAT in forming CALIBNET in 1986, DELNET in 1988 and another networks subsequently. The UGC (University Grants Commission, India) set up INFLIBNET in 1988, NIC (National Informatics Centre) also set up DELNET, which is the first operational library networks in India. CALIBNET and INFLIBNET have not been able to project their performance as they were planned but they are trying to achieve their goal. Functions, organizations, cooperation’s, progress and creation of databases amongst libraries of BONET and CALIBNET as library networks are still dissatisfactory. Institute of Economic Growth (IEG) was founded in 1958 as an autonomous institution recognized by the University of Delhi. The IEG library has its networking with different libraries e.g. Ratan Tata Library, Delhi School of Economics, Delhi University library, NCAER library, IIPA Library, DELNET etc. for helping the readers in getting the desired books and journals which are not available in the IEG library on inter-library loan basis. There are other examples of networking and resource sharing of Astronomy libraries in India. They jointly established a networking for resource sharing amongst the libraries e.g. Indian Institute of Astrophysics (IIA) library, Inter-University Centre for Astronomy and Astrophysics (IUCAA) Library, National Centre for Astronomy and Astrophysics (NCRA) Library, Nizamiah Observatory (NO) Library, Physical Research Laboratory (PRL) Library, Raman Research Institute (RBI) Library, Tata Institute of Fundamental Research (FIFR) Library, Uttar Pradesh State Observatory (UPSO) Library. The main purposes of this networking are for better resource sharing, to reduce costs, for speedy delivery of documents, to keep abreast of new developments etc.

No efforts have been made in India to network public libraries though it is very much essential to cater networked information to the general public where more than 70% of the total people are residing in rural areas. However, much emphasis should be given at the national level in India for the development of documentary information resources because it is considered as one of the vital resources to promote the development of economy, science, technology and culture etc.

Resource Sharing Networking Models:

The fundamental object of information resource sharing is to solve the problems faced by the libraries such as information explosion, several needs of users, increasing cost of subscribing periodicals, sharing library budgets, fluctuation of the exchange rates etc. Now-a-days, to solve the above problems different resource sharing networking models are observed at local, regional, national and international levels. Generally, there are three levels of national resource sharing networks exist: (a) Local: Information is stored in local libraries in the form of Union Catalog for local collection available in local libraries; (b) Regional: Information is stored in regional libraries and services are provided on broad subject area basis; and (c) National: Information is stored in national library in the form of national Union Catalog for
normally national collection available in national library and services are rendered to users as national resources. There are four existing resource sharing networking models, which have been shown in tabular form in Table - 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Aims &amp; Funding</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centralized collection development and Services at national and Regional level.</td>
<td>Resources: Acquired centrally and stored at single site Funding: Contribution by participating libraries. Grants are also sought from government and private agencies.</td>
<td>National Lending Library. UK.</td>
</tr>
<tr>
<td>Centralized Collection Development and Services by Subject</td>
<td>Resources: Subject specific collection of documentary resources. Acquired centrally and stored at a single site city, region, or country may limit the geographic distribution of libraries. Funding: Marketing of services and grants from the government and private agencies.</td>
<td>National Service Library at INSDOC. New Delhi</td>
</tr>
<tr>
<td>Centralized Collection Development at Organizational level</td>
<td>Resources: Libraries belonging to a single bigger organization collaborate. The shared collection is acquired centrally at a single site. Funding: Organization backing the library provides funds. The Participating libraries may also contribute towards the central funds.</td>
<td>CSIR, DRDO, DOE, ISRO</td>
</tr>
<tr>
<td>Coordinated collection development at Institutional level</td>
<td>Resources: Eliminates duplications. Serves at the level of participating libraries. The geographical area of cooperation could confine to a city, region, or country. Funding: The individual libraries determine their level of support. User libraries pay for the services they avail of.</td>
<td>DELNET, BONET, MALIBENT</td>
</tr>
</tbody>
</table>

UGC –Infonet Electronic Journal Consortium is an innovative project launched by University Grant Commission (UGC) to promote the use of electronic data bases and full text access to journals by the research and academic community in the country. This project will bring about a qualitative change in the academic environment. The research and academic community can now have an access to resources at their finger tips.

**Consortia Initiatives in India:**

The information revolution and proliferation have brought about drastic changes to the function and service in all types of libraries in India during last two decades. Many libraries in India till today are not in affording position to procure all documents and subscribe to core journals in major disciplines or CD-ROM databases, due to their financial crunch. At the national level, the UGC (University Grant Commission, India) setup INFLIBENT in 1988. It has taken imitation about a formidable change in developing adequate infrastructure in libraries, especially university libraries, to be a part in the
networked environment. Since January 2004, University Grants Commission through its one point programme (UGC) / INFONET is providing access to e-subscription to all-important journals for the entire university community. The Ministry of Human Resource Development (MHRD) through its INDEST (Indian National Digital Library in Science and Technology) has launched consortia-based subscription to electronic resources for higher – technical education systems in India.

Besides, there are a few national and regional library consortia developed in the recent years. Council of Scientific & Industrial Research (CSIR) Tata Institute of Fundamental Research (TIFR),, Department of Atomic Energy (DAE), Indian Institute of Technology (IIT), and Indian Institute of Management (IIM) have already formed their sectoral consortia and have been subscribing to electronic sources like Science Direct, MathSciNet, and Blackwell, John Wiley, ABI/INFORM and Business Sources Premier. Also, both Institute of Mathematical Science (IMSc) and TIFR have been subscribing to MathSciNet database under their own consortia consisting group of libraries in their region. And many more are expected to come soon.

**Increasing the Discovery and use of e-resources in University Libraries and Research Centres:**

There is a large quantity of subscribed e-resources in our libraries and they contain quality information, though expensive. In spite of advantages in terms of access and search capabilities, they are underused. Systematic plan has to be in place for their promotion of use. While a good ICT infrastructure is a prerequisite, it alone will not do. Proactive strategies are required and these need to be adopted imaginatively. Access to e-resources need to be made easier for both on campus and off campus users. As a priority, active users need to be identified and they need to be converted to heavy users of e resources. Secondly, non users be converted to active users Various methods have to be tried in order to grab the attention of the users towards the e-resources. User training will increase the confidence level of the users. Traditional awareness methods include : Personal visits, user training, brochures, posters and displays. Newer technologies from the Web 2.0 such as RSS alert service, Blogs, Wikis and Facebook make the interaction with the library not only interesting but also add more value. Finally, the effectiveness of various promotional strategies need to be measured by monitoring the usage and user feed back. This research sought to determine use of online resources and databases and to assess current user characteristics associated with use of online resources by the Faculty and researcher of various universities

**Suggestions:**

1. The university library should make necessary arrangement for continuing subscription for the on line journals along with print journals.

2. The University library should conduct orientation, training programme regularly to assist the users of agricultural consortium programme.

Conclusion:

Remote access of e-resources has been a major boon to academic and research libraries. Online journals are considered the note chord of any library’s collection and have become indispensable for research in any field. Many online journals available in the form of databases as well as they directly access through the Internet. The quantity of online journals is growing larger and has become a quite visible entity in serial publication. Today most of the online journals appear as parallel version of its print counterparts and more publishers are making their journals available in electronic format. Many academic institutions are currently building substantial collections of full-text journals and continue to increase access to various online databases. Because these resources come at a great cost, it becomes important to understand database and full-text journal use among university patrons and the characteristics accompanying today’s remote and in-house library users. Increased access to computers, the Internet, online databases full text journals necessitates reassessing online use patterns and user characteristics. Nowadays it is impossible for libraries to procure all the documents and subscribe to core journals that are in demand by the users. There are many online journals and databases are available open access. Subscription of online journals and databases through the consortium(s) are much economic for the libraries.

Reference:

Use of E- Resources by the Research Scholars and Faculty of Management Studies in Management Research Centres, Tamil Nadu

V. Thangavel  
Dept of Library and Information Services, SCSVMV University, Kancheepuram - 631 561  
estv_libn1@yahoo.co.in

Dr. K. Ramakrishna Reddy  
Dept of Management Studies, Paventhar Bharathidasan College of Engg and Tech Trichy

Dr. V. Ramesh  

Abstract

The electronic resources have the effect of democratizing the research community. The study is an attempt to determine the level of using various types of resources by the research scholar. Research scholars felt that the electronic resources would be useful to carry out their research and do their further research works, depending on their subject. This article discusses about the use of electronic resources by the research scholars and faculties of management studies through World Wide Web in Chennai, Tamil Nadu.

Introduction:

Computer-based automation was initially incorporated into library operations as a mechanism for handling the routine functions of running a library such as circulation, cataloging, acquisitions, interlibrary loan and serials control. Systems for handling these operations became available able to the larger library community from the early 1970s onward, although there were some earlier pioneers with well developed local systems. Early systems were typically run on large computers into which data were entered, processed behind the scenes, and returned as printed output of some type (overdue notices, invoices, or catalog cards, for example). Pioneers in reference database services, such as DIALOG, also operated using large mainframe computers to which patrons connected via terminals and required experienced searchers mastering a complex set of commands to generate effective search results.

The modern society is dynamic and complex. The duty of research scholar should know the availability of sources in their topics. Electronic resources are available in various formats. The different types of electronic resources are electronic books, electronic journals, electronic databases, search engines and subject gateways.

Public Services:

Electronic resources are changing the longstanding relationship which traditionally existed between library professionals and users. Formerly, the use of resources and services required patrons to visit the library. Print indexes in the reference area provided needed citations the card catalog indicated whether that library owned a particular item, and the circulation desk could place a hold for an item checked out to another patron. Early electronic tools such as DIALOG were available only to librarians who
frequently consulted with users to refine their needs before conducting expensive and sometimes complex database searches. The introduction of large scale automated union catalogs like OCLC have helped to revolutionize services such as interlibrary loan and library functions such as acquisitions and cataloging. The introductions of newer electronic resources have shifted services and functions from those which are librarian mediated to those specifically geared towards the end user. Advances such as the availability of online citation and document delivery services remotely accessed from a home or voice, 24 hour remote availability to OPACs providing both an item’s current circulation status and the ability to place holds oneself, plus the increasing number of resources available directly to patrons through networks, have combined to eliminate the need for a visit to the reference desk, or the library in many cases, to satisfy information needs.

Objectives:

The following objectives are framed for study.

1. To evaluate the purpose of using electronic resources.
2. To identify the use of electronic resources.
3. To identify the purpose of electronic resources by the research scholars and faculty in various field in Management Research centers, Chennai, Tamil Nadu.
4. To find out the best publicity method to promote the usage of electronic resources.
5. To identify the format of print or electronic resources and their performance.
6. To know the problems faced by Research Scholar in accessing electronic information.

Literature Review:

A good number of related studies on the usage of electronic resources have been conducted mainly in developing countries like India and developed countries. These studies led to know the status of the usage of the electronic resources against investment.

Dugdale studied the library services at the Bolland library, university of the west of England, Bristol in the UK and investigate the ways; in which students might be encouraged to use electronic resources and to develop important life long learning skills through the Reside (research, information, delivery) electronic library.

Ferguson stated that the university of Hong Kong in China had found that the change to a mostly elec5tonic collection has been successful. He stated that 59% patrons preferred reading electronics serials while 30% favored the printed version. Only 14% of the academic staff still favoured printed journals.

Alwarammal R , Sivaraj S and Madasamy R stated that respondents were asked to mark their potential use of electronic journals by using a five-point scale of 0-4 where 0 indicated not satisfied and 4 meant highly satisfied. Assigned scores for each item were converted into mean scores. Frequency of responses for each resource and respective means scores are given. It was found that elsevier’s science direct was the most used resource by the faculty and students in all engineering disciplines and it is receiving the highest means core of 2.32. It was followed by IEL online was the next used resource by the faculty and students of computer, electrical and electronics, electronics and communication and it is receiving the mean score of 1.91. It was followed by .51 and .64 mean score of ASCE and ASME resources used by civil
and mechanical engineering students respectively. The other two resources of Nature & Nature biotechnology and management journals with a means score of .76 and .42 were mostly used by biotechnology and management students respectively.

**Methodology:**

The research scholars and faculties of various research organizations in Management Research Centers, Chennai, Tamil Nadu represented the target population for this study. The questionnaire method has been employed to collect the data for the present study. The questionnaire was constructed based on the following elements: use profile, frequency of visit, purpose of using online resources, acceptance of electronic media, awareness of the existence of electronic resources, publicity to promote the usage of electronic resources, paper presentation in conferences, papers published in Journals (National and International), publication details and reasons for non-use of electronic resources.

**Data Analysis:**

The data collected through questionnaires were organized and tabulated by using statistical methods and percentages. A total number of 450 questionnaires were randomly distributed among the faculty and researchers, out of which only 432 filled questionnaires were returned to the investigator. The response rate is 96 percent.

**Sample:**

The population of this study is faculty and Research Scholars in Management studies in Management Research centers, Chennai, Tamil Nadu. The sample consists of 432 faculty members and Ph D students who filled in the questionnaire of the annual user survey of 2009-2010. Thus, the sample consists of self-selected volunteers and may be biased towards those who are most active users of electronic resources. The analysis below indicates that the sample is reasonably representative by sex, discipline, occupation and university. However, some later findings hint that it is likely that active users of electronic resources are somewhat over-represented in the data. Inspection of the demographics of the respondents showed that the sample’s breakdown by sex was nearly equivalent to the population. 50 % of university faculty and PhD students in Management Research Centers, Chennai, Tamil Nadu were women in 2009, whereas a in this study 50% of the respondent were women.

**Suggestions:**

On the basis of the response and opinion given by the respondents some of the important suggestions have been made, which will help the effective use of the electronic resources.

1. The authorities of the university and research institutions should take keen interest for providing better infrastructure facilities for the improvement of internet speed, so that users can feel more comfortable in browsing e-resources.

2. The university library and research institutions should make necessary arrangement for continuing subscription for the print journals along with e-journals. Since more than 95% of Research scholar are interested in making use of print journals.

3. The university library and research institutions should conduct orientation/training programmes regularly to assist the use of e-resources.
4. Majority of the research scholar have suggested that UGC-INFONET and other library consortium and they should provide pdf files of science direct, wiley-interscience and all other scientific journals.

Conclusion:

One of the major developments in libraries and information systems in the past 10 years is the advent and spread of electronic resources, information services and networks mainly as a result of developments in information and communication technologies. The change is basically of physical form where information content is increasingly being captured, processed, stored and disseminated in electronic form. The unique features of the information needs of users in electronic environments relate to the physical form in which information content is made available in electronic information environments. Users normally desire the content to be made available within the constraints of their skills and technological capabilities so that it is possible to access and use the required information content to resolve the felt gap in knowledge.

Reference:


3) Liu, Ziming (2006), Print vs. Electronic resources: A study of user perceptions, preferences and use, Information processing and Management, 42 (2), 583-592.


நூற்றாண்டு பண்டிகை போராட்டம்
தமிழ் கல்லூரியில் நடைபெறும்

பாண்டிகைப் போராட்டம்

தமிழ் கல்லூரியில் நடைபெறும் செயல்பாடுகளின் வரிசையில் தமிழ்கலை பாண்டிகை போராட்டம் என்பது ஒரு முக்கியமான நடைபெறும் செயல்பாடு ஆகும். செயல்பாட்டுக்கான விளக்கம் மற்றும் பிரித்தூர்வத்தில் இதுவே கொள்ளப்படுகிறது. தமிழ்கலை பாண்டிகை போராட்டம் ஆனது தமிழ் கல்லூரியின் உரையார்வு பகுதியாகும். இச்செயல்பாடு அதிகம் நடைபெறும் செயல்பாடுகளின்றி ஆகும். தமிழ்கலை பாண்டிகை போராட்டம் என்பது ஒரு முக்கியமான செயல்பாடு ஆகும். தமிழ்கலை பாண்டிகை போராட்டம் என்பது ஒரு முக்கியமான செயல்பாடு ஆகும். தமிழ்கலை பாண்டிகை போராட்டம் என்பது ஒரு முக்கியமான செயல்பாடு ஆகும்.

செயல்பாட்டுக்கு உரையார்வு பகுதியில் கல்வி தரும் வழி காலை கல்வி என்றும் சொல்லப்படுகிறது. மேலும் இது ஒரு முக்கியமான செயல்பாடு ஆகும். தமிழ்கலை பாண்டிகை போராட்டம் என்பது ஒரு முக்கியமான செயல்பாடு ஆகும்.

eMail: softviewindia@gmail.com | www.softview.in
Quality Assessment Technique of E-Learning in Tamil Language

A. Kovalan
e-mail: kovapmu@gmail.com
Siva Pillai
e-mail: pillaisiva@gmail.com

Abstract

Web-Based Teaching Learning Process (WBTLP) is a rapidly growing area in Education. Traditional forms of teacher education are transformed, as the Internet becomes a new medium for communication. Traditionally teachers have fulfilled dual roles as presenters of structured information and social agents in the educational process. Students are in need of good interactive resources with learning tools and techniques. Hence, there is a need for training in WBTLP so as to enable the teacher to provide good resources in the web. The web-based learner resources can improve the quality of teacher education by availing various tools and techniques of assessment. The assessment of web-based learning resources helps to provide quality web resources in teacher education. It is also used to help teacher to have better resources and environment in which teaching takes place. The environment includes the organization, the learning materials, use of media, delivery methodology and various approaches in details.

Assessment is a judgment regarding the worth or value of something. Typically the assessment process is divided into two parts. The first is a teacher assessment, which relates to interaction and guidance of a teacher with students and the second is a learning resources assessment, which relates to quality of materials and resources of a course. However, the primary function of the assessment is to help teacher to improve the total quality of education in web teaching learning environment.

1. Introduction:

The rise of e-learning and web-based training in Teaching Learning Process in Tamil Language (TLPTL) has lead to a growth in the use of web based learning assessment, which will increase as the use of e-learning becomes more widespread. Designing web-based tools, making resources and quality assessment of the learning resources is a challenging task for educators, curriculum designers, computer programmer and evaluator of the resources in the Web-Based Teaching Learning Process (WBTLP). The teacher in online should take special effort for implementation and development of the instructional resources which is clear, accessible support, simplicity, transparency of the educational materials, and likelihood of the students’ misconceptions as a result of interacting with materials. The teacher or trainer assessment is also very important to learners’ motivation, self-regulated learning, continues education, knowledge improvement, people with disability and poor social background. This paper presents an overview of web based learning assessment cycle in two different aspect viz. Web Based Learning Teacher Assessment Cycle (WBLTAC) and Web Based Learning Resources Assessment Cycle (WBLRAC).
2. Web Based Learning Teacher Assessment Cycle:

An online teacher/trainer plays a key role in developing and maintaining an effective online learning environment and possesses a unique set of skills to success of a student. A quality of online teacher is very important than any resources in online. The role of online teacher is not just respect their need; they need to be involved in proper guidance, formative and summative assessment, counseling, administration and learners’ motivation. They should also support their feelings, and make facilitate to communicate with the learners with different media and methods. This means that the online teacher should make their time efficiently with their students. Hence, the quality of teacher/trainer can assess in six major categories viz.:

![Teaching Assessment Cycle (TAC)](image)

**Fig. 1. Teaching Assessment Cycle (TAC)**

2.1. Teacher/Trainer Entry Behaviors:

Once you have identified a person as an online teacher/trainer for a specific task or target groups the organizer determine what kind of knowledge, skill, and experience need to be taken into account for the specific training process. The entry behavior may be determined by three major areas viz.:

![Teacher Entry Behaviour](image)

**Fig. 1.1. Teacher Entry Behaviour**
2.1.1. **Project Manager:**

Project manager is an achievement of academic qualification, administrative power, attitude level, communicative ability, leadership quality, involvement in students’ achievement, investigative technique of students’ problems; task oriented training activities, cooperation with student on assignment, equal treatment made to each student, responsibility, freedom, and support by giving opportunity to express their opinion and relationship with students, teacher and designer.

2.1.2. **Subjective Expert:**

Subjective Expert is based on theoretical and experimental experience in a different content area. The e-content provider should have knowledge of content development, material editing and method of teaching. E-content development is essential for learner, multimedia producer, Instructional designer and distributors. These are the quality needed for bearing on the delivering of the training product/event.

2.1.3. **Designing Expert:**

Designing expert should have knowledge of expectation of an online learner. The e-content material designer have different roles in a web based learning materials developing a project viz. instructional designer, multimedia designer, graphics artist, internet based application developer, media producer, and web administrator.

2.2. **Teaching Methods:**

Web based trainer use different methods to deliver an instructional material. The structure of an online teaching method consists of following three components.

![Fig. 1.2. Structure of teaching method](image)

The trainer provides content in different media viz. text, pictures, audio, video, animation, and simulation which helps student to understand the concept easily. It implies that the learner has freedom to choose his choice by the way of his interest and preference of his learning materials.

2.3. **Learners’ Motivation:**

Encourage learners to solve their problems by interaction with facilitator, peer group discussion, and expert in their chosen subject. The models of achievement motivation most often attribute students' academic motivation to cognitive processes (Bandura, 1986; Weiner, 1992) which regulate students'
learning behaviour. There is a growing body of evidence (Wentzel, 1991, 1994, 1996), however, that a consideration of social motivation of classrooms should not be excluded from the model of achievement motivation. Three aspect of learners’ motivation is important in web based teaching process viz.:

![Fig. 1.3. Learners’ Motivation](image)

Success at online teaching learning process requires students to achieve three outcomes of education: individual power, academic achievement, and social adjustment. Judgments on levels of motivation were made using the following criteria: teacher interaction with learners and individual pupils, time keeping, teaching style, and observed enthusiasm in different aspect.

2.4. Communication with Learner:

Continues communication with learner helps to overcome the isolation problems. Web based teacher have many channel to communicate with their learner viz.

![Fig. 1.4. Learners’ communication channel](image)

**Electronic mail (E-mail)** can be used by the facilitator and learners-to-learners communication. E-mail is cost-efficient, fast and convenient. Group e-mail can also be used to contact all learners simultaneously.

**The telephone** has used as a synchronous method of supporting the learners. The telephone allows learners to communicate with their teacher/trainer. Learners can discuss instructional and non-instructional problems with their facilitator.

**Chatting** helps to learner to communicate with his facilitator and his peer groups. Two types of chatting methods are available viz. video chatting and audio chatting. Video chatting is like a face-to-face communication and the audio chatting is like a telephone communication.

**Listservers** is an e-mail list that allows any of the users registered on the list to e-mail a specific listserve address, after which the e-mail is forwarded to everyone on the list. It is useful for the learner to discuss of topics relevant to the learning event.

**Netforum** allow the facilitator to post learning event information or announcements, changes to the learning event or deadlines for assignments. It is also used by learner to post questions about the learning event which the facilitator and other learner can answer. Learners and researchers are encouraged to discuss and reply in discussion forum.
2.5. Response to the Learner:

It is the quality of awareness that is evoked in collaborative meaning-making with students that defines the quality of a teacher's response to the teaching situation. Immediate feedback, which makes encourage, interest on further and continuing education, helps the learner to achieve their goal. A variety of possibilities exists. The teacher may have a word or phrase of affirmation:

```
Response Phrase
```

```
Correct AnswerWrong Answer
```

```
“Good;” “Exactly right;” “Yes.”
“Wrong”, “Please try again”
```

*Fig. 1.5. Teachers’ Response Phrase*

2.6. Learners’ Advancement:

Learners’ record shows the quality of teacher/facilitator in web based teaching learning process. Ensuring the achievement of learners in different aspects viz. academic achievement, personal improvement, social behaviors, communication development, self-awareness, self-motivation, etc.

3. Web Based Learning Resources Assessment Cycle:

Instructional resources are plays an important role to deliver the learning materials through online. Using these resources, the learner can contact their facilitator at any time or in a fixed time when they are having problems. A quality resources make quality learning product and provide quality learner in web based training process. The quality resources can be assessed by six major categorize viz.

*Fig. 2. Instructional resources*
3.1. Instructional Objectives:

Objectives are useful in content development, material designing and implementation of the web based learning process and evaluation of the learner. The instructional objective describes a kind of performance expected by learners at the end of a learning event. The objective helps learner to plan the learning opportunity in their interested area. The development of instructional objectives is a task whose importance should not be overlooked. The instructional objectives can be divided into two major categories.

![Instructional Objectives](image)

**Educational Objectives** (Knowledge domains) **Environmental Objectives** (Resources domains)

*Fig 2.1. Learners’ Instructional Objectives*

3.1.1. Educational Objectives (EdO):

Educational objectives help learners know where they are going and how they are going to achieve their goal. EdO is designed based on learners’ interest and involvement. In this domain the facilitator investigate the learners’ interest to achieve their goal.

3.1.2. Environmental Objectives (EnO):

This environment helps learner to achieve their goal at their own location. In virtual learning environment the learner gets freedom to choose their interest subject with establish guidelines for including graphics, videos, audios, animations, pictures, and various presentation media. This environment makes more benefits to a learner to achieve their goal without any strain and pain.

3.2. Content Analysis:

The content review team might include the client, instructional designer, subject-expert(s) and programmer. Developing new Learning Materials (LM) is a major task involving their design, development, delivery and evaluation to ensure their effectiveness. Check the existing materials before providing a new one. If there is large number of subject matter and other information relevant to the learners’ achievement goal on the web, then it is probably suitable for web-based delivery. The LM should provide more opportunity to the learner to choose their specific content area. The learning event needs to be updated regularly. In content analysis the following points need to consider for analyzing to provide a better learning materials.
Analyze, how well the content and material presented in the way that was both interesting and stimulating. The evaluation of these materials should include gathering data regarding the relevance of various assignments and the quality of the various assessments.

3.3. Instructional Media Evaluation:

In an online environment, media are used to provide variety to what is essentially a text-based methodology. Such media may include video, although this is limited unless learners have access to a broadband network, graphics, simulation, other visual effects and synchronous and asynchronous communication. Instructional media makes learner with self-learning, self-confident and self-control.

3.4. Evaluation of Learners:

This evaluation of reaction is based on measurement of learners' feelings and opinions about the course after completed. The facilitator can get the information here relates to methods of instruction, course content, learning methods and materials. Behavioral changes of learners are a measurement of the behavioral changes occurring as a result of the learning event after completed.

3.5. Delivery Methodology:

The learning materials should be downloaded easily at learners place. The instructional media delivered with minimum space and maximum speed at learners’ center. The material and media should flexible to the learners’ environment.

3.6. Revising Resources:

The facilitator can determine the ability of the web system to support the learner, the effectiveness of the learning environment, the effectiveness of the support offered to the learner, how web based learning
benefits when compared with the traditional methods. Based on the above findings of the evaluation changes may be needed to the learning methods and materials to support the learner.

4. Conclusion:

Methods and media play a major role in the success of a web-based learning process. Hence we need to assess the material and media to provide a better learning event. A web-based learning process cannot be successful without well-designing learning material, good facility support and assessment. Web-based learning is a self-based learning and places the responsibility for study directly on the learners themselves. They need to know how to use the web based learning environment to be successful.

References:


Impact of Multimedia technologies for Tamil education:-A global perspective

Prof. B. Jagadhesan
D.B.Jain College (Autonomous), Chennai-97
jagadhesanb@gmail.com Mobile:-9444532133

Introduction:
Tamil is a Dravidian language predominantly spoken in southern India and northeastern Sri Lanka, with smaller communities of speakers in many other countries. As one of the few living classical languages, Tamil has an unbroken literary tradition of over two millennia. The written language has changed little during this period, with the result that classical literature is as much a part of everyday Tamil as modern literature. Tamil school children, for example, are still taught the alphabet using the āṭṭicūṭi, an alphabet rhyme written around the first century AD.

Modern technologies have started changing the lifestyle of our modern population. Multimedia, as a teaching aid, is very much effective with color, sound, graphics, which are found in the audio media and video media and movie media. One of the most significant changes in tamil education in recent years has been the availability of a range of Information Communication Technologies (Multimedia). Thus, Multimedia is no longer a new terminology nowadays. Almost everyone is familiar with Multimedia not only at work but also in schools as it encompasses the Internet and Multimedia.

The power of Multimedia can be used effectively in language teaching and learning as there is a paradigm shift from traditional teaching to using Multimedia in classrooms. It provides alternative teaching methods and it is a breakthrough from the traditional classroom environment. This is an efficient method to engage today's pupils. Multimedia tool is believed to provide the possibilities of multiple perspectives and a realistic learning environment. The real power of multimedia to improve education may only be realized when students actively use them as cognitive tool. Furthermore computer based learning is more motivating for students and this is generally accepted by educators and by administrators. Special innovative method can be supported to the young learners acquiring more vocabularies for suitable communication transactions in Tamil.

The contents have been presented with an elaborate use of multimedia technologies. The lessons of the educational programmes have been made more interactive and illustrative through pictures, animations, video clippings and audio contents. The library presents a picture-gallery and a video gallery on a large number of cultural events, temples and historical monuments. It also presents the devotional songs in audio mode. This paper presents the impact and experiences of developing multimedia based online resource for Tamil Education and culture history and art for global perspective.
Need of Multimedia in Teaching Tamil languages:

Explore the possibilities of the effective use of Multimedia in imparting Tamil language education. Analyze the potential of 2D and 3D software, audio, video, comics, comic strips and motion pictures in the acquisition of Tamil language skills. Identify the effective use of open source software and operating system for instructing Tamil language skills. Internet comprises innumerable features. It is the need of the hour to find out the means of using all those features for spreading the Tamil Language. Technological advancements in the field of 2D and 3D animations, Video, Audio, Comics, Comic stripes, Motion films have made the Tamil teaching incredibly easy. Software and operating systems like Ubuntu(Tamil Linux), Microsoft, Microsoft Tamil office, suratha Unicode writer and converter, Google search engine and guruji search engine etc have greatly helped for the development of Tamil language. Internet acts as bridge for connecting all the media together. In today’s scenario, web blogs and social net working websites work as major sources of alternative Mass media. Web blogs and social networking web sites play a significant role in imparting Tamil education all over the world in different forms—audio, video, 2D, 3D and motion pictures, etc.

Issues in Teaching Tamil:

The main issue with teaching Tamil is lack of quality teachers. Even though there is some limited support from government for community languages, the lack of demand from the Tamil speaking population has meant very few schools offering Tamil lessons.

As far as concerned there is no demand to employ a full time or part time Tamil teacher. The number of Tamil children in the school varies. If you look at the highly populated Asian community, there are community languages classes such as urdu, Punjabi, and Arabic in the mainstream school. This is not only because of the number of students. At present Tamil is at the same stage in some countries. Efforts at developing a syllabus and teaching framework are underway. As you will all agree, a strong fundamentals framework is a basis for a strong education system. This brings us to the role of multimedia in teaching and learning Tamil.

Real impact of Multimedia for teaching tamil language

The development of material that takes full advantage of computer-assisted learning used to be much more involved and typically employed interactive learning and Multimedia. A team comprising of interactive designers, graphic designers, programmers, testing team and a manager, usually develops such materials. A multimedia project may take up to six months of development time costing as much as $70,000. Despite this, not many commercial products seem to have a substantial shelf life. Many outstanding titles done by big companies like Broderbund are seemed irrelevant (content-wise) in many countries. Like most teaching resources, multimedia packages vary greatly in complexity, quality and perceived usefulness. As with most educational materials, there are many approaches to teaching a topic, and even a well-produced program is unlikely to satisfy everyone. Therefore such commercial titles are difficult, if not impossible, to be updated modified or improved by content experts for their specific needs. For that, the original programmer must be available to access the source code. Since the expertise to use new teaching tools stays with the IT developers, rather than with the teachers, the typical teacher makes little or no progress in the use of interactive media tools. To date there are less than 20 quality Tamil language based educational CD ROMs available worldwide. Most of these titles are produced in
Malaysia. The content in these titles are based on the countries of their origin. Even though these titles score high in pedagogical value, the fact is there are still based on the local curriculum.

Authoring tools have improved rapidly. Some of these tools are easy to master. Time has come too drop the need for systems that make strong collaborations between IT Technicians and content experts (i.e. teachers). Now educationists can independently develop effective and customized instructional software for their own classrooms. In the old days of programming, the developer had to everything in such arcane general purpose compiler languages like Assembler, Basic or C. However, now, many sophisticated resources are available readily. These include vast libraries of media, which can be copied and pasted in a program. Such clip-arts and sound clips are available on the Internet and CD ROM packages. There is also a range of tools available for each medium. From high-end professional packages to less sophisticated but still powerful and useful freeware offerings are available for the initiative programmer. The advent of easy-to-use multimedia authoring packages (like Mediator), and the widespread availability of computers, and the ease with which traditional classroom teaching can be put into instructional software have all created an environment in which, traditional classroom teaching can be transformed and made more exciting for students and teachers.

Conclusion:

The perennial problem faced by emigrant Tamil communities is getting quality resources to teach Tamil to the younger generation. Multimedia development has actually become quite easy. I hope that this paper would inspire the audience to get together and develop their own multimedia teaching and learning pedagogical materials.

References:

1. Karrer, T Understanding eLearning 2.0
நான்று சங்காமை

குறியீட்டின் மீது இயற்கையான வாகன வரையறுத்து, வேளாண்டு விளக்கத்தில் குறிப்படும் விளக்கத்து வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும்.

1. வகுப்பு:

குறியீட்டின் மீது இயற்கையான வாகன வரையறுத்து, வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும்.

2. வகுப்பு:

குறியீட்டின் மீது இயற்கையான வாகன வரையறுத்து, வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும். என்று முற்பாடு தருவதற்கு வேளாண்டு விளக்கத்தில் எழுதிக் கொள்ளப்படும்.
2. நடுநரா கல்லூரிய அறிக்கை:

1. PHP, MySQL
2. Java, MySQL.
3. Flash

3. பிற்றில் விளக்கங்கள்:

4) கொழும் மின் கொட்ட 그런 குழுவைப் பற்றிய தொடர் பார்வையிலுள்ள பிரித்தூர் குறிப்பிட்டுவிடும் வட்டாரம் பிரித்தூர்.

1. PHP, MySQL
2. Java, MySQL.
3. Flash

3. பிற்றில் விளக்கங்கள்:

4) கொழும் மின் கொட்ட குழுவைப் பற்றிய தொடர் பார்வையிலுள்ள பிரித்தூர் குறிப்பிட்டுவிடும் வட்டாரம் பிரித்தூர்.
4. சேவைகள்

1. Data Mining Tamil Web Pages or Text Files.
2. E-learning package for teaching Tamil Grammar using Tholkappiam.
3. Translation Software (English to Tamil).

4. எடுத்துக்காட்டு

பொருளாக்கம் சேவைகள் கிளைப்போன இனத்தான் அறிவியல் சேவைகளால்.

5. குறிப்பிட்டு

சிறப்பாக எச்சரியா அவ்வியலில் இருந்து முதலாம், இது பொருளாக்கத்தில் அவ்வியல் சேவைகளின் பயன் அறியும். உலகளத்தில் உலகளத்திற்கு பயன்படுத்தாது மற்றும் புதிய விளக்கங்களை தெரிக்காது. பொருளாக்கத்தில் இருந்து வகுப்புகள் தேடும் பொருளை விளக்கத்தில் பயன்படுத்தும்.
Teaching Primary Education in Tamil Using LMS and Visualization Techniques

Richards Hadlee
Final Year M.Sc Computer Science,
Loyola College, Chennai
hadleefdo@gmail.com

R. E. Iniya Nehru,
Technical Director, NIC
Chennai
nehru@nic.in

Abstract
Extend teaching primary education in Tamil through Visualization technique. Visualization is creating images, diagrams, or animations to communicate a message. It’s an effective way to communicate both abstract and concrete ideas. Visual representation of information requires merging of data, computer graphics, design, and imagination. Moodle provides only text information (report, items analysis, etc) and does not provide visualization tools. We use third-party tools to do visualization for LMS. Visualization helps to communicate results and understand better.

Keywords: Moodle, LMS, Visualization

1. Introduction & Purpose
LMS (Learning Management System) is highly interactive teaching tool. It offers great variety of workspaces to facilitate information sharing and communication among participants. It also provides a platform for a teacher to create/deliver contents, monitor student participation, and assess student performance through web application; this helps wider audience of students across the world (Education-Anytime and Anywhere).

E-learning classes most of the time is asynchronous. In this mode the instructor and students can interact via messaging or e-mail, and assignments can also submitted. The students and the instructor can be online at the same time to communicate directly and share information if required. E-Learning can include training, delivery of just-in-time information and guidance from Teacher.

A LMS (Learning Management System) is used to organize an online learning environment. The term "online" refers to Internet and Intranet environments. It can include images, text, audio, video, animation and virtual environments.

LMS and Visualization Techniques help to teach the primary school subjects by representing interactive or animated flash image files. The child chooses the learning style which has the seeing, feeling. Understanding learning styles is only a first step in maximizing potential and overcoming learning differences.
1.1 Lists of Functionalities

- Creating LMS for Primary education
- Applying visualization
- Questions and log reports

1.1.1 Creating LMS for Primary education

This LMS has been designed for primary education (1st to 5th Standard) in Tamil language. Each standard has its own subjects (Tamil, Mathematics, Science, and Social Science). These contents have been designed using visualization techniques.
1.1.2 Applying visualization

Visualization is a branch of computer graphics which is concerned with interactive presentation / animated digital images for users to grasp easily. These techniques facilitate analysis of large information by representing visual display. The instructor and students can interact with forum, chat & etc.

1.1.3 Questions and Reports

Each lesson contains the various types of exercises, through which the user can perform exercises in an effective way.

The grades are provided for the particular question which is performed by the user. The status of the question also is displayed at the end of the test.

If the user answered correctly, right status is displayed and grade has been marked as 1/1. If the user selected the wrong answer, no grade is provided.

**Reports**

The Reports is represented through a visualized pattern (Graph). This contains the number of days the student has worked on this. In this graph the x-axis consists the date and month he worked in and the y-axis consists of the levels of work he has done on the particular day.

**Log report**

Log Reports represents the lists of user who enroll into the particular course. Through this log the admin can get the performance of the user.

**Activity Report**

An activity report has the number of activity in that particular course. It contains the number of available resources for the course.

**Participants Report**

The Participants reports consists of the following criteria,

- What Activity?
- At what day?
- Reports of admin or user?
- What actions?
2. Implementation of Tamil Learning in Moodle

This application has designed and developed through Moodle, PHP & Mysql using Visualization Techniques. The purpose of this application is to bring awareness about e-learning using Moodle. Moodle is the easiest and most flexible LMS. It’s easy to navigate, had features that are directly applicable to the writing classroom and best of all, is free for me to download and customize. It has a strong support community and good online documentation to help us. Moodle is specifically designed with educators in mind, allowing for easy setup and maintenance.

Primary Education

Primary School Teaching is a social networking and resource sharing site made exclusively for Teachers. It also allows other members of staff working in Primary Schools to participate. It is a platform for sharing teaching resources and ideas. It allows teachers to communicate effectively in a collaborative environment. Proof of that is, you will see many times a kid playing with his toys and imagine that he is in a battle field or a camp and they really feel it. All the subjects can be created with questions, so that the users get grade for any subject.

It is a process of transforming information into a visual form that enables user to observe the information. On the other side, it uses techniques of computer graphics and imaging. Successful visualizations can reduce the time to get the information, make sense, and enhance creative thinking. We can use visualization for improving memory, restoring health, reducing stress, increasing relaxation.

The following properties are used in visualization

- Text
- Graphics
- Images
- Video
- Audio
- Animation

3. Approach and Design

LMS and Visualization Techniques help to teach the primary school subjects by representing interactive or animated flash image files. Understanding learning styles is only a first step in e-learning that overcome the learning differences. There is no doubt of the effectiveness of visualization in every area of life, even a child does visualization although they don’t know what it is.

Resources

The course has the lessons in it. The lists of resources are available in each lesson in the particular course. The lessons have been implemented in the tool called eXe. The eXe project developed a freely available Open Source authoring application to assist teachers and academics in the publishing of web content without the need to become proficient in HTML or XML markup.

Resources authored in eXe can be exported in IMS Content Package, SCORM 1.2, or IMS Common Cartridge formats or as simple self-contained web pages.
The content material in the particular lesson also been designed in Flash, in which a students can be attracted to learn difficult subjects through interactive audio & video based content.

The LMS has also been implemented using Visualization and Computer graphics in order to improve the courses on the students learning in Tamil.

Three main uses of visualization are:

- **Motivation.** Creative visualization is a great way to see a possible future and move you towards it.
- **Mental practice or rehearsal.** Mental practice or mental rehearsal is complementary to real practice. Mental practice can also be cost-effective and safer.
- **Reinforcing other techniques.** Visualization is a powerful way to strengthen other techniques, such as association and scripting

5. Conclusion

In this work we have shown application of teaching education in Tamil using LMS along with a visualization technique. We have described how different visualization techniques can be used in order to improve the courses on the students learning in Tamil. All these technique affiliated to the student’s performance.

The animated image avoids the monotonous feeling of reader which we normally get in class room. Presently most of the children are computer friendly, spend lot of time in computer games, animation etc. Hence developing a course content using LMS tools like Moodle, PHP & MYSQL gives a different visual presentation to children. The students can be attracted to learn difficult subjects using Visualization and Computer graphics. The retention rate for an average student in found to be much more using LMS techniques compared to traditional classroom training.
Acknowledgment

I express my sincere thanks to Mr. E. Iniya Nehru, Technical Director, NIC (Chennai) His encouragement and enthusiastic support throughout this paper induced me to do well. I also thank Mr. Jeyakumar S, the Project Manager at NIC who helped me to get the needed information and encouraged for successful completion of this paper.

6. References

Books:
- Using Moodle 2nd Edition,
  By Jason Cole and Helen Foster
  Published by O'Reilly Media, Inc, 1005 Grayenstein Highway North,
  Sebastopol, CA 95472

Articles:
- Factors in the deployment of a Learning management system at
  The University of the South Pacific
  Moodle For Teachers, Trainers And Administrators
  Moodle An electronic classroom

Resources:
- http://download.moodle.org/
- http://exelearning.org/wiki
- www.docebo.org
காலமான தீர்மயமாக கணினிக் கலை

(நூற்றாண்டு அங்கேற்றக்குடி).

தெரிகிறது, தமிழ்நாட்டு நகராட்சிகள் தொடர்பான பண்டைய வேகமான காவல் தொடர்பிடுத்தல் நடைபெற்றுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய வணிகத்துறையினர் வேகமான காவல் நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது.

1970-க்கு முன்னர் பண்டைய வணிகத்துறையில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது.

பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது.

1970-க்கு முன்னர் பண்டைய வணிகத்துறையில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது.

பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது. பண்டைய நாடுகளில் தொடர்பான பண்டைய வணிகத்துறை நடைமுறைநாள் வேகமாயுள்ளது.
அந்தார்த்திக்காலம்

"அந்தார்த்திக்காலம் காண்பது அனுப்பி" என்று அவர் கூறினார். அந்தார்த்திக்காலம் நீர் வள்ளார் நோக்கும் நோக்கும் என்று கூறினார்  அந்தார்த்திக்காலம் என்னும் தமிழ்நாட்டு மலர்களின் குறிப்பிட்டல் தெளிவாக இருந்தும். இந்த அந்தார்த்திக்காலம் என்றும் ஆட்சிக்காலம் என்றும் என்றும் இருந்தும் குறிப்பிட்டல் தெளிவாக இருந்தும். அந்தார்த்திக்காலம் என்றும் அந்தார்த்திக்காலமானது என்றும் என்றும் இருந்தும் குறிப்பிட்டல் தெளிவாக இருந்தும்.
வொண்டுபிரிக்கும் உணர்வின் வழியாக அலகிப்பு நிலை உற்பத்திய கட்டுரை என்னும். ஆனால், இப்படைப்பு முதல் தொழில்மதிப்பில் பொருளின் வேளாண்மை தோன்றும் இல்லாமல் குறித்து வாழ்வு பெறலாம்.இவ்வாறு இளமையில் உணர்வு நிலை உற்பத்திய கட்டுரை என்னும். இப்படைப்பு முதல் தொழில்மதிப்பில் பொருளின் வேளாண்மை தோன்றும் இல்லாமல் குறித்து வாழ்வு பெறலாம்.தினமும் இளமையில் உணர்வு நிலை உற்பத்திய கட்டுரை என்னும். ஆனால், இப்படைப்பு முதல் தொழில்மதிப்பில் பொருளின் வேளாண்மை தோன்றும் இல்லாமல் குறித்து வாழ்வு பெறலாம்.

(Subvoccial speech, recognition) என உயர் வாழ்வில் முதல் தொழில்மதிப்பில் பொருளின் வேளாண்மை தோன்றும் இல்லாமல் குறித்து வாழ்வு பெறலாம். இவ்வாறு இளமையில் உணர்வு நிலை உற்பத்திய கட்டுரை என்னும். ஆனால், இப்படைப்பு முதல் தொழில்மதிப்பில் பொருளின் வேளாண்மை தோன்றும் இல்லாமல் குறித்து வாழ்வு பெறலாம்.
பகுதியில், தரைப்படி சுற்றுலாம் குறுக்குச் சுற்றுலாம் நூற்றாண்டு செய்ய விளையாட்டின் ஒரு பகுதியில் கருவக்கரைகள் காட்சியில் இருக்கின்றன. பொருளியல் பொருளியல் உள்ளது, அதில் அதன் பராமரிப்பு அரசியல் மற்றும் அரசியல் துறையின் பங்களிப்பு விளையாட்டின் அடையாளம் தூண் முதலிடச்சுற்றில் உறுப்பினர் வருவார்.

"போர்க்கால்" காலத்தில் மூன்று வருடாகத் தொடர்ந்து பெண் பால்வாய்ப்புடன் போர்க்கால் விளக்கல் வைத்து அடையாளம் கருவக்கரையின் கூட்டணி காட்சியில் இருக்கின்றன. பொருளியல் பொருளியல் நூற்றாண்டு, மற்றும் அதில் அரசியல் துறையின் பங்களிப்பு விளையாட்டின் அடையாளம் தூண் முதலிடச்சுற்றில் உறுப்பினர் வருவார்.

அரசியல் குழு விளக்கங்களில் என்ற "போர்க்கால்" விளக்குமுறை மற்றும் தொடர்ந்து குழுவை செய்யாவிட்டு "தொடர்ந்து குழு" விளக்கம்.
Role of Cloud Computing in Tamil Language Development

Mrs. R. RajaRajeswari  
Assistant Professor in Computer Science  
M.V.M Government Arts College for Women, Dindigul  
ganesh_akash@yahoo.com

Dr. Mrs. A. Pethalakshmi  
Head & Associate Professor in Computer Science  
M.V.M Government Arts College for Women, Dindigul  
alaguvisu@gmail.com

Cloud Computing is a new computing paradigm which is expected to transform the way computing is done today, in near future. Cloud Computing offers virtualization of all high end Computing Services. And it offers four layers of Services:

Saas (Software as a Service), PaaS (Platform as a Service), CaaS (Computing as a Service) and IaaS (Infrastructure as a Service). These services can cut the software cost, storage cost and utility cost of running a wide computer network. This research paper addresses how cloud computing can facilitate Tamil Language development. The following areas are identified as prospective avenues of development: E-learning, Tamil Computing, Tamil E-resources and Tamil Internet Services. This research paper will analyse the existing tamil software services in the above said areas and explore these avenues of tamil development through Cloud Computing.

(Keywords: Cloud computing, E-Learning, Tamil Computing, Tamil E-Resources)

1. Introduction

Forecast on Computing tells cloudy days are ahead for Computer Scientists. Software Developers and Users should be ready for a changeover as Cloud Computing waves roll down into the Internet. This new computing paradigm is expected to transform the style of computing in a remarkable way. Tamil, one of the world’s five classical languages, is a privileged language, to be learnt by Lord Buddha[11]. Its history dates back to its existence on stone inscriptions in Jerusalem. This research paper tries to make these two ends meet ie. how the new computing paradigm, cloud computing, can be connected with this historical language. This research paper is organized as follows: an introduction to cloud computing, avenues of Tamil Language Development facilitated by cloud computing and the final conclusion.

2. An insight into Cloud Computing

Cloud Computing [8] is Internet based computing, whereby shared resources, software and information are provided to computers and other devices on demand like a public utility.

Cloud computing [9] is Internet based (“Cloud”) development and use of computer technology (“computing”) . The cloud is a metaphor for the Internet based on how it is depicted in computer...
network diagrams and is an abstraction for the complex infrastructure it conceals. According to a 2008 paper published by IEEE Internet Computing “Cloud Computing is a paradigm in which information is permanently stored in servers on the Internet and cached temporarily on clients that include desktops, entertainment centers, table computers, note books, wall computers, hand helds, sensors, monitors etc.” Cloud computing offers virtualization of all high end services and it offers four layers of Services\[1,2,4,5,6,7\].

(i) Applications or Software as a Service
SaaS means delivering software over the internet. This nullifies the need to install and run application on individual computers, making software upgrades, maintenance and support obsolete. (e.g) Salesforce.com

(ii) Platform as a Service
PaaS is the delivery of a computing platform as a service. PaaS offerings facilitate deployment of application without the cost and complexity of buying and managing the underlying hardware and software and provisioning hosting capabilities. (e.g) Windows Azure, Amazon Elastic Computer Cloud.

(iii) Infrastructure as a Service
IaaS is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components. The service provider owns the equipment and is responsible for housing, running and maintaining it. The client typically pays on a per-use basis. (e.g) Amazon S3

(iv) Computing as a Service
CaaS [1] integrates the above said services. (e.g) Verizon’s CaaS. Like other cloud offerings, Verizon’s CaaS allows customers pay for data-center resources such as storage and application hosting dynamically based on the amount of resources they consume. These services can cut the software cost, storage cost and utility cost of running a wide network. The following sections bridges Cloud Computing Services and Tamil Language Development in prospective avenues.
3. Tamil E-resources

Literary Resources in a language symbolises its richness and liveliness. Now in this internet age, e-mail box has replaced the conservative letter box in a house. Hence E-resources have a definite role in making a language flourish. Internet has paved a way for Tamil Internet magazines, e-books and also a Tamil electronic library (www.tamilelibrary.org). For example, www.projectmadurai.com plays a major role in the area of e-documentation of old Tamil literature. One has free access to Tamil Literature in this website from “Abirami Anthathi” to “Alai Osai”. Apart from this, many tamil magazines have their own websites comprising even digital archives of old ones. (e.g) www.vikatan.com. Also any creator’s work needs to be published and recognised, for both the growth of the language and the creator.

Tamil Internet magazines act as writer’s workshops, for budding Tamil writers from which they can learn and also where they can publish their work. (e.g.) www.thinnai.com, www.nilacharal.com. With the advent of Cloud Computing, usage of IaaS can cut storage costs, emancipating separate clouds for old Tamil Literature, Magazines and modern Tamil Literature. Such e-resources available on the Internet facilitates free access to Tamil Literature for any one across the globe, which is a requisite for a language to grow.

4. Tamil E-Learning

Tamil community has settled across the globe even many decades before. And there is need to teach Tamil for the next generation young learners to keep the tamil heritage and language expand its horizon. Internet takes the avatar of a virtual Guru for this purpose. “International Academy for Internet Tamil “

Formerly, Tamil Virtual University (www.tamilvu.org) does this job with ease in collaboration with Tamil University, Thanjavur and awards Degrees, Certificates and Diplomas in Tamil for Tamil learners across the globe through Internet.

Virtual Class room, a part of this academy’s website, which transits the learner from the living room into a class room needs additional software in a Remote terminal, Eg. Multimedia Software. Using PaaS, this software can be downloaded and any remote user can turn the terminal into a virtual class room. And apart from this enlightening Video lectures by eminent Tamil Professors will reach across the globe when available as AaaS for enthusiastic learners.

5. Tamil Internet Services

Twenty first century is the century of Information and Communication Technology. At the click of a mouse, Internet provides both communication in its highest speed and information in its best shade, at any moment. Searching for Tamil web pages is mostly done in English and only some Tamil search Engines like www.googletamil.com does this job in Tamil that too using transliteration method. Tamil e-mail software like www.azhahi.com though exist are not popularly used by tamil population. Since personal and official communication is under the process of transformation from letters to mails, sending mails in mother tongue will be a necessity among users who have a fast pace of life, to share their true emotions.

Apart from this, a language should be put into use in all forms of communication by that language speaking people for its longevity and the one not, can exist only in books and stone inscriptions. And CaaS comes as a help at this moment for the software developers to develop next generation Tamil search
Engines and Tamil E-mail sending software. Required Infrastructure, Platform, and Computing for developing these software can be got from IaaS, PaaS and CaaS providers respectively.

6. Tamil Computing

Tamil Computing provides Tamil software tools which are used in a day to day basis. (e.g) Tamil word processors, Tamil Database software like Amudham These tools need to get popularised and put into use by software users. Also Tamil Speech Recognition software is yet to come. To develop Tamil Speech Recognition Software more computing power is needed. To develop or run Speech Recognition applications,because of digital filtering and signal processing high processing speed and more memory are required[3]. And cloud computing comes handy by providing them through CaaS,since cloud computing transforms a desktop into a super computer. To help website developers develop tamil websites easier, Tamil website design templates can be made available as AaaS.

7. Conclusion

Tamil is a unique language, an universal language, whose literature suits mankind of all centuries, all places and all tribes. [10] states that context sensitive rules of modern computer science are found in Tholkaapiyam,a classical literature.This language has undergone transformations, its storage from stone inscriptions to palm leaves, paper and now Web pages.

For a language to live long, it should be spoken without the acquaintance of any foreign language. Today’s tamil language is spoken along with many english words. In spite of this, this language is ever young and vibrant and can be symbolised as a wild flower which blossoms in a natural way.Hence this language will evolve with any emerging technology and tamil language can develop in the above said avenues through cloud computing services .

References

1. http://connectedplanetonline.com
5. http://searchcloudcomputing.techtarget.com
7. http://web2.sys.con.com
8. en.wikepedia.org
14

தமிழ் உத்தி விபர்பிகாங்
முன்பு சிறந்தசாலாம் நேர் பிரியலங்கள்

முன்பு மாற்றங்கள் நேரவென அதற்கு சிறந்தசாலாம் பலவட்டத்தில் அவர் சயந்தமாகும் உண்டு.

மாற்றங்களுக்கு எத்தனைக்குறிக்கான செயல்பாடு அல்லது செயல்பாடுகள் அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. செயல்பாடுகள் செய்யப் பயன்படுத்துகிறது. மாற்றங்கள் நேரவென அதற்கு சிறந்தசாலாம் பலவட்டத்தில் அவர் சயந்த பலவட்டங்கள் பலவட்டங்களுக்கு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. மாற்றங்களுக்கு எத்தனைக்குறிக்கான செயல்பாடு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. அவர் சயந்த பலவட்டங்களுக்கு இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. மாற்றங்களுக்கு எத்தனைக்குறிக்கான செயல்பாடு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது.

மாற்றங்கள் பலவட்டத்தில் எத்தனைக்குறிக்கான செயல்பாடு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. மாற்றங்களை விளக்கப் பயன்படுத்துகிறது, அதற்கு சயந்தமாகும் பலவட்டங்களுக்கு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. மாற்றங்கள் பலவட்டத்தில் எத்தனைக்குறிக்கான செயல்பாடு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. மாற்றங்களை விளக்கப் பயன்படுத்துகிறது, அதற்கு சயந்தமாகும் பலவட்டங்களுக்கு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது. மாற்றங்களை விளக்கப் பயன்படுத்துகிறது, அதற்கு சயந்தமாகும் பலவட்டங்களுக்கு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது.

மாற்றங்களை விளக்கப் பயன்படுத்துகிறது, அதற்கு சயந்தமாகும் பலவட்டங்களுக்கு அடிப்படையில் இருந்து இவற்றை செய்யப் பயன்படுத்துகிறது.
வெளிப்படையான புதிய இலக்கணம், புரியல் ரீதியாக நடைமுறையாக வெளியே உள்ள எண்களை, ஐந்து வருடங்கள் முழுமை, பன்னாட்டு முறையில் நகர்க்கும் கோட்டைக்கு வரவு வெளியே சுருக்கங்கள் அமைந்துள்ளன. நெடுநோய்கள் புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோய்கள் அவற்றை புரிய நெடுநோ�
தகவல் (Data Base)" பொருளாக உரையாக அறிய தொன்மையான ஒரு முன்னேற்றம் தொட்டும் பின்னர் பின்னர் திறந்து கொள்ள முடியும் பொருள் நூற்றில். முன்னேற்றம் செய்யும் இடைவெளியில் இறுதி செயல்களும் ஊட்டிகளும் பரிசாக்கலித்து அறிய தொடர்பில் கூறும் "தகவல் தொகுப்புகள்" தொடர முன்னேற்றம் பிரிமையும் பொருள் நூற்றில். முன்னேற்றம் செய்யும் ஒரு வரிசை என்பது பட்டியல் தொடர்களை பார்த்து கூறும் பட்டியல். இவ்விருப்பளவை தொகுப்பாக்கள் ஒரு வரிசை என்பது போன்ற அறிக்கைகள்.

4.2 தகவல் தொகுப்பாக்கள் வளர்ச்சிக்கு பயனுள்ளது

4.2.1 தகவல் தொகுப்பாக்கள் வளர்ச்சிக்கு பயனுள்ளது: இவ்வியல் நேச்சிப்பாட்டு கல்வி போன்ற தகவல் தொகுப்பாக்கள் வளர்ச்சியுடன் புருத்தம் அடைகின்றன. கையாண்டிகளுக்கு பயனுள்ள வளர்ச்சிக்கு பயனிக்கும் முக்கியமான நோக்குகளுள் ஒன்று பெரும்பான்மை உடையது. தகவல் தொகுப்பாக்களை பயன்படுத்துவதால் பெரும்பான்மை உடையது.

4.2.2 தகவல் தொகுப்பாக்கள் அளிக்கும் செயற்கையுடன் விளக்கம் இந்த பாக்கிகள் அளிக்கும் செயற்கையை எடுத்துகொள்கின்றன. தகவல் தொகுப்பாக்கள் ஒன்று அடைகின்றது. கையாண்டிகளுக்கு பயன்படுத்தப்படும் வரிசைகளுக்கு பயன்படுத்தப்படும் வரிசைகளுக்கு உடையது.
இலையமாதையில் பெரும் நெருக்கம் நூற்றணவர் பல்ளிப்பாண்டியில் உள்ள வாணையில் அவக்குக்கே காணக்கொண்டது. அந்த காலத்திலிருந்து யாரோ இலையமாதையானது காணக்கொண்ட நூற்றணவர் பல்லிப் பாண்டியில் உள்ள வாணையில் அவக்குக்கே காணக்கொண்டது. பண்பாடை பிராந்தியாக உள்ள வாணையில் உள்ள காலத்திலிருந்து யாரோ இலையமாதையானது காணக்கொண்டது.
Pannirukaivadivelan R

University of Madras (pann1973@gmail.com)

L I P

Lithuanian, Serbian, Hindi, Tamil, Thai, Marathi, English, MS-Word, Excel, PowerPoint, Access, Pagemaker, Photoshop, Outlook Express, MSN messenger.

Unicode

The main advantage of LIP approach is that any kind of editor can be used such as MS-Word, Excel, powerpoint, Access, Pagemaker, Photoshop, Outlook Express, MSN messenger

Bulk Unicode converter

The main advantage of LIP approach is that any kind of editor can be used such as MS-Word, Excel, powerpoint, Access, Pagemaker, Photoshop, Outlook Express, MSN messenger.

University of Madras (pann1973@gmail.com)
Microsoft office word, Microsoft office excel, Microsoft office power point, Microsoft office front page, Adobe pagemaker, Adobe Photoshop

1999 ஆம் ஆண்டில் ஓர்டேகோ பதுருத்தான வடிவச்செய்யப்பட்டது. எனவே, Windows 98, Windows Me போன்ற அமைப்பில் இல்லாதது. பெயர் தொற்றுப்பட்டது. குறிப்பிட்டும், பெயர் மற்றையோரத்திற்கு எளிதாக வைக்க வேண்டும். புது வடிவங்கள் வைக்கப்படும் வரையறுக்கப்பட்டது. தற்கொலை விளக்கங்களை நேர்ந்து மீண்டும் பெயர்ப்பொருள். Transliteration or script conversion சான்று வைக்கப்படும் விளக்கம் தொடர்ந்து முடியும்.

2.0 (e-kalappi – 2.0)

2001-2004 ஆண்டுகளில் தமிழும் மற்றையோரத்திற்கு எளிதாக பரவலாக வைக்கப்பட்டது. பதிலிடுதல் Unicode பதிலிட்டுவைக்கப்பட்டது. நேர்ந்து மீண்டும் பெயர்ப்பொருள். பெயர்ப்பொருள் வைக்கப்படுத்த வேண்டும். தற்கொலை விளக்கங்களை நேர்ந்து மீண்டும் பெயர்ப்பொருள். TSC பதிலிடுதல் Unicode பதிலிட்டுவைக்கப்பட்டது.

New Horizon Media (NHM)

2004 ஆண்டு வேலைகளை பூர்த்து உருவாக்கினர். வருகை மற்றும் வேலைகளை பூர்த்து உருவாக்கினர். வருகைகளை பூர்த்து உருவாக்கினர். வருகைகளை பூர்த்து உருவாக்கினர். வருகைகளை பூர்த்து உருவாக்கினர்.
தமிழ் கிராம் (88 KB - 9,11,029 bytes) சுருக்கம், கட்டுரையில் தமிழகம் கிளமண்டு. இருண்டு
சோவுண்டு இன்று தமிழ் பயன்பாடும் பல்வேறு பிரிவுகள்

முகம் அறிவியல்

தமிழியல் வருவாய் இந்து நாட்டுக்களில் சோவுண்டுபார்வை. இது தமிழ் எழுத்துநூல் என்று விளக்கம்
நூற்றாண்டுகள் முறையில் பரவலாக உள்ளது. இயற்கைக் கூற்று
முனையில் தமிழர் எழுத்துமதிகளும் இந்து நாட்டுக்களில் பரவலாகத் தொடங்கின

நூற்றாண்டுகள் விளை

தமிழியல் பொருளியல் வருவாய் தமிழாக்கமும் தமிழ்வாசிகள், நூற்றாண்டு வாழ்வை
முறையில் தமிழ்வாசிகள் எழுத்துமதிகளும் பரவலாக உள்ளது. இடையில் Unicode, TSC, TAB, TAM பரவலாக இன்று நூற்றாண்டு

பக்கம் கிளமண்டு எழுத்து

தமிழியல் எழுத்துமதிகளும் தமிழ்வாசிகளும் இந்து நாட்டுக்களில் சோவுண்டுபார்வை. சோவுண்டு
நாட்டுக்களில் சோவுண்டுபார்வை எழுத்துமதிகளும் இந்து நாட்டுக்களில் சோவுண்டுபார்வை. சோவுண்டு

தமிழநாட்

1986இல் (MSDOS) தமிழியல் பொருளியல் வருவாய் தமிழநாட்டுப் பார்வை. தமிழ்வாசிகள்
தமிழ் எழுத்துமதிகள் எழுத்து வளர்க்கும் போது, தமிழியல் வருவாய் தமிழ்

தில்லியூ, தேசியர் பருவம் என்று பருவம் எழுத்து வைத்து எழுத்து வைத்து என்று
தமிழியல் வருவாய் எழுத்து தமிழ்வாசிகளும் பரவலாக உள்ளது. இது படிப்படியாக
தமிழியல் வருவாய் எழுத்து வைத்து எழுத்து வைத்து என்று வைத்து என்று வைத்து என்று

9997ீலூர் என்று

தமிழ்வாசிகள் எழுத்து வைத்து எழுத்து வைத்து என்று வைத்து (Romanised Thunaivan, Mylai IE) பரவலாக உள்ளது
வளர்க்கும் போது, Unicode, TSC, TAB and Unicode எழுத்து முனையில் சோவுண்டு

தொடக்கம்

தமிழ் பதிப்பிட்டு இன்று நூற்றாண்டு, கட்டுரையில் பொருளியல் வருவாய் தமிழநாட்டு
தமிழியல் வருவாய் தமிழ்வாசிகளும் என்று பதிப்பிட்டு எழுத்து வைத்து என்று

ீலூர் என்று பதிப்பிட்டு எழுத்து வைத்து என்று பதிப்பிட்டு எழுத்து வைத்து என்று

ீலூர் என்று
தமிழ் முறையில் செய்தலை விளக்கம்: வானவில் ஆக்கம் பொருளாதாரம் சரியான கணினிமான என்கிறது பொருளாதாரம் பிள்ளை பொருளாதாரம் என்பது பொருளாதாரம்.

இது கணினிமான விளக்கம் தமிழ்நாட்டில் மேலாண்மை என்று குறிப்பிட்டு வலர்கிறது. எனவே கணினிமான மேலாண்மை என்பது வெளிப்படையான குறிப்பிட்டு வலர்கிறது. கணினிமான மேலாண்மை என்பது வெளிப்படையான குறிப்பிட்டு வலர்கிறது. கணினிமான மேலாண்மை என்பது வெளிப்படையான குறிப்பிட்டு வலர்கிறது. கணினிமான மேலாண்மை என்பது வெளிப்படையான குறிப்பிட்டு வலர்கிறது.தமிழ்நாட்டில் பொருளாதாரம் முறையில் செய்தலை விளக்கம் வலர்கிறது. பொருளாதாரம் என்பது வெளிப்படையான குறிப்பிட்டு வலர்கிறது.
உங்கள் பணியிற்காக எந்த இல்லமே போன்று இல்லையானாலும், ஆண்டுதோறும் கூறுவதே இவ்விதமே அருள்மிகு மீளலாம். இவ்விதமே உங்களின் பணியில் உலகின் வாழ்க்கைகளை பார்த்து நீட்சி கூறுவதே. இவ்விதத்தில் கருத்து பெறவும் பணியில் பற்றியுள்ள விளக்கத்தைக் கூறுவது இருக்கிறது. சந்திப்பு குறிப்பிட்டு பாதுகாக்கவும் குறிப்பிட்டு பாதுகாக்கவும் இது உங்கள் கூறுவதாலுமே என்பது. உங்களை கூறுவதாலுமே என்பது.
Information Retrieval System
for Tamil and Non-Tamil Users

S.Srividhya
Student, M.C.A-CEG Anna University
Chennai-600 025
apssreevidhya@gmail.com

Dr.T.Mala
Senior Lecturer, Anna University
Chennai-600 025.
mala@cs.annauniv.edu

Abstract
This scheme enhances a remarkable approach to formulate and expand query. It accepts query given by user and expands it through a series of query expansion techniques. The input given by user is Tamil text and the output is expanded query to be given to a search engines to retrieve relevant English documents.

Keywords-Word sense disambiguation, Word stemming, Transliteration.

Introduction
The World Wide Web (WWW), a rich source of information is growing at an enormous rate. According to Online Computer Library Center, English is still the dominant language in the web that contributes most of the content. However, global internet usage statistics reveal that the number of non-English internet users is steadily on the rise, but all of them are not able to express their basic needs in English. Tamil users who are not able to express their needs in English are also growing in the Internet. They generally search for the information using the Tamil search engines. But the content provided by these search engines is less in number. Making the huge repository of information on the translation module takes the stemmed word web, which is available in English, accessible to non-English internet users has become an important challenge in recent times. When the non-English users want to access the existing search engines, most of the time they arrive at improper formulation of English queries. The proposed system aim to solve the above problem by allowing the users to pose the query in their own (source) language which is different from the language of the documents that are searched. This enables users to express their information need in their native language while the proposed system takes care of expanding the given user query that can be given to Search engines like Google, Alta vista.

LITERATURE SURVEY
“Query formulation for Information Retrieval System for Tamil Users” seeks to develop efficient techniques for query formulation whose accuracy and precision are determined by parameters like precision and recall. The query formulation process is broadly divided into three main portions: word stemming, translation and word sense disambiguation. The Word stemming is done with a morphological analyzer whose source code is modified to perform word stemming. The input is got in UNICODE and converted to TAB-Anna font. The translation module takes the stemmed word as input and transliterates those words to as input and transliterates those words to English and the equivalent
English words Bilingual dictionary is developed for this purpose. The word sense disambiguation makes use of Bootstrap algorithm [3] which disambiguates 55% of the words with 92% accuracy.

CLIR system for Agricultural society [8]: Tamil-English CLIR system for agricultural society uses Lesk algorithm for word sense disambiguation which is only 44% accurate and the Tamil words from the input query is directly searched in bilingual dictionary without checking for spelling mistakes.

Information retrieval system using mobile networks [5]: This system needs user interaction to disambiguate the word and the translation is prone to ambiguities if the context is not clearly specified by the user.

Cross Lingual Information Retrieval Using Data Mining Methods [6]: In this system the feedback and suggestions from the users are collected for document mapping. For each candidate word the pair wise measure gives a measure of correlation. However, these correlations are not available in dictionary representations and must be generated by use of appropriate ontological systems.

Tamil search engine[2]: This system discuss the issues related to crawler, database storage structure and other functional modules of the search engine. While they have shown some limited success, the approach used by the search engine is limited to the Tamil language and it retrieves the Tamil documents.

**SYSTEM ARCHITECTURE**

The system architecture is shown in Figure 3.1. The system mainly focuses on the construction of suitable English and the equivalent English words are fetched from a bilingual dictionary. query for relevant English document retrieval in an information retrieval system. The proposed system gets Tamil input and retrieves relevant English documents according to the user query. The query is then given to spell checker. Morphological Analyzer obtains the root terms of source query by removing grammatical inflections. By applying rules for handling suffices, oblique, etc., the root words are obtained in the given query. Transliteration is done to convert Tamil letters to English characters in a systematic way. The output is then disambiguated to find out the exact equivalent English word. Word sense disambiguation is done that identifies the correct sense of an ambiguous word that is being used in a query.

![Figure 3.1 Word Sense Disambiguation](image-url)
Translation is part of word sense disambiguation. For each sense of a given word, it is compared with all possible senses of the surrounding words in the given query and the word with maximum senses is chosen as the appropriate word. With the exact English words obtained as a result of word sense disambiguation, query formulation is done and given to search engines for English documents retrieval. If Tamil query is entered, the grammatical inflections are removed using morphological analyzer and query is formulated to retrieve relevant Tamil documents. WordNet, dictionary are the resources used. The formulated query is given to an existing search engine like Google, Alta vista. It uses bootstrap algorithm for Word sense disambiguation.

Bootstrap algorithm for Word sense disambiguation: Bootstrapping algorithm for Word Sense Disambiguation succeeds in disambiguating a subset of the words in the input text with very high precision. It uses WordNet as resource to disambiguate and for the purpose of identifying the correct sense of the words in a given text. The bootstrapping process initializes a set of ambiguous words with all the nouns and verbs in the text. It then applies various disambiguation procedures and builds a set of disambiguated words: new words are sense tagged based on their relation to the already disambiguated words, and then added to the set. This process allows us to identify, in the original text, a set of words which can be disambiguated with high precision; 55% of the verbs and nouns are disambiguated with an accuracy of 92%.

**EVALUATION PARAMETERS**

**Precision**

In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search:

\[
\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}
\]

**Recall**

Recall in Information Retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

\[
\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}
\]

For example for text search on a set of documents recall is the number of correct results divided by the number of results that should have been returned. The precision and recall values for five different queries have been calculated as follows.

<table>
<thead>
<tr>
<th></th>
<th>0.72</th>
<th>0.89</th>
<th>0.78</th>
<th>0.76</th>
<th>0.80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>0.77</td>
<td>0.86</td>
<td>0.84</td>
<td>0.78</td>
<td>0.82</td>
</tr>
</tbody>
</table>
Conclusion

The proposed System helps to find links of documents for health society and to retrieve the documents from a large corpus in English language. This system is very much useful for the users who can understand English but do not know to give appropriate English query. The users can pose their query in Tamil and retrieve relevant English documents by using this system. The Query formulation for information retrieval system for Tamil users displays the search result in English. It is appropriate, if the results are displayed in their own language for the users who do not know how to give query in English. The precision can be improved by employing Machine translation methods instead of word-by-word translation technique. The bilingual dictionary can be extended to include more words which will expand the scope of the system. This system can be further extended to Rank the pages and provide a summary (in English) of top pages, translate the summary to Tamil or provide an answer to the query in Tamil.

References

9. Dr. Vasu Renganathan,” An Interactive Approach to Development of English-Tamil
தமிழ் இலக்கியங்கள் !

கலாச்சார கல்பாணி கல்பாணி (ஏரியல்நிகழ்வு), கல்பாணி வல்லுணர்கள்

தமிழ் முறையாலும் அறியவும்

மானோகை அலைச் சீரமைப்பு கல்லூரியில் கேட்டுப் பார்வையிட்டு மீண்டும் இந்த கல்லூரியில் காலமிட்டு வராம்!

முக்கிய தமிழ் இலக்கியங்கள் கல்லூரியில் இயங்கியது என மறியவேண்டும். இலக்கியங்கள் குறுநிலை கலைக்கான பட்டியலில் பார்வையிடும் அவர்களும் காலமிடும்.

கல்பாணி பணி தொடர்ந்து இலக்கியங்கள் கல்லூரியில் இயங்கியது என மறியவேண்டும். இலக்கியங்கள் குறுநிலை கலைக்கான பட்டியலில் பார்வையிடும் அவர்களும் காலமிடும்.

தமிழ் முறையாலும் அறியவும்

மானோகை அலைச் சீரமைப்பு கல்லூரியில் கேட்டுப் பார்வையிட்டு மீண்டும் இந்த கல்லூரியில் காலமிட்டு வராம்!

முக்கிய தமிழ் இலக்கியங்கள் கல்லூரியில் இயங்கியது என மறியவேண்டும். இலக்கியங்கள் குறுநிலை கலைக்கான பட்டியலில் பார்வையிடும் அவர்களும் காலமிடும்.

தமிழ் முறையாலும் அறியவும்

மானோகை அலைச் சீரமைப்பு கல்லூரியில் கேட்டுப் பார்வையிட்டு மீண்டும் இந்த கல்லூரியில் காலமிட்டு வராம்!

முக்கிய தமிழ் இலக்கியங்கள் கல்லூரியில் இயங்கியது என மறியவேன்றும். இலக்கியங்கள் குறுநிலை கலைக்கான பட்டியலில் பார்வையிடும் அவர்களும் காலமிடும்.
1. Indexing - இணையம் செய்யும் விளக்கங்கள்

2. Web Crawler - செய்யும் பொருள்களானது

3. Indexing - இணையம்

4. Searching - தேடும்

5. Search Engine

6. தேடுப்பொருள் பொருள் தேடும் விளக்கங்கள்

1. Web Crawler - பொருள்கள் விளக்கம்

2. Indexing - இணையம் செய்யும் விளக்கங்கள்

3. Searching - தேடும்

4. Search Engine

5. Search Engine

6. தேடுப்பொருள் பொருள் தேடும் விளக்கங்கள்

1. Web Crawler - பொருள்கள் விளக்கம்

2. Indexing - இணையம் செய்யும் விளக்கங்கள்

3. Searching - தேடும்

4. Search Engine

5. Search Engine

6. தேடுப்பொருள் பொருள் தேடும் விளக்கங்கள்
தமிழ் உயிர்களின் விளக்கம்

தமிழ் உயிர்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன. தமிழ் உயிர்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன. தமிழ் பாடல்கள் வேலைகள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.

தமிழ் விளக்கங்களின் வருடாய் விளக்கங்கள் பொருட்கள் மற்றும் விளக்கங்கள் தமிழ் பாடல்கள் மற்றும் பொருட்கள் முழுமையாக விளக்கப்பட்டுள்ளன.
• அக்கிரமத்தால் வரும் கண்டெடுப்பு மற்றும் தமிழ்ச் சொல்லிகளால் கூறப்பட்டுள்ள (மு) 'செல்லுங்கள' காதலும், 'செல்ல' காதலும். இந்தக் குறிகள் விளக்காக்கம் கூறியுள்ள வர்த்தகராகும்.

1. இது பொருந்து அவர்கள் வரும் செல்லுதல் பொருந்துக்கூற்றுப்பட்டுள்ள கூற்றின் (Shared Hosting) மயனியைப் போன்று அவர்கள் செல்லுதலை எடுகிக் கூறுகின்றனர். அவர்கள் என்கிற வரும் செல்லுதலை எடுகின்றனர். வரும் செல்லுதலை எடுகின்றனர்.

2. அவர்கள் பொருந்தேற்பர், காண்டு இணை வரும் பொருந்து கூறுக்கூற்று அவர்கள் முழுமையாக பொருந்தேற்பர் மயனியை அவர்கள் அவர்கள் மயனியை அவர்கள் மயனியை. அவர்கள் மயனியை அவர்கள் மயனியை. அவர்கள் மயனியை. 

3. காண்டு இளங்கும் மயனியைப் பொருந்தேற்பர் கூறுக்கூற்று அவர்கள் (Dedicated server) மயனியை. காண்டு இளங்கும் மயனியைப் பொருந்தேற்பர் மயனியை. காண்டு இளங்கும் மயனியை. இதை வெளிப்படையாக இணை வருமாறு மயனியை ஆய்வுற்று பொருந்தேற்பர் மயனியை. 

உதவுற்று பொருந்தேற்பர் (Auto Complete)

<table>
<thead>
<tr>
<th>தலை</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>மென்சு</td>
<td></td>
</tr>
<tr>
<td>மென்சு</td>
<td></td>
</tr>
<tr>
<td>மென்சு</td>
<td></td>
</tr>
<tr>
<td>மென்சு</td>
<td></td>
</tr>
</tbody>
</table>

MySql Server for database "not available !
Trying to reconnect to database...
Cannot connect to database.

OR Search
Tolerant Search

முன்னாள் கூறுக்கூற்று கூறியுள்ள வர்த்தகர் www.higopi.com என கூறுகிறார் பொருந்துக்கூற்று கூறுக்கூற்று
படம் 2: செயல்பாடு குறிப்பிட்டு வருகை

படம் 3: படம் வருகை

புதுவரை:

Searchko - The King of Search
for Tamil Web Documents

Sobha Lalitha Devi, Pattabhi R K Rao T, Vijay Sundar Ram R
Au-KBC Research Centre, MIT campus of Anna University, Chennai- 600 044
{sobha, pattabhi, sundar}@au-kbc.org

Abstract
Searchko is a Tamil portal, which uses information retrieval (IR) technology for searching the Tamil content in the web. The Searchko engine uses many types of natural language processing technology for getting the most relevant output. The web has nearly 10 million documents in Tamil and bringing it under one umbrella using IR technology is the aim of Searchko. The etymology of searchko is Search+Ko ie search+ king, ko in Tamil means king and thus this portal is a unique portal where you can get all the content, whether it is News from the newspapers, Cinema, Tamil literature, Music and Cricket. The unique features of SearchKo are Multiple Font support, Enhancing the query with the help of a morph analyzer, Works with a phonetic visual keyboard, Spell Checking, Query expansion using thesaurus and dictionaries.

Introduction
The growth of technology and internet has brought information revolution in our country and across the world. This 21st century is called as information age. This has changed the way people share the knowledge, do business, and interact with each other. Until 1990’s internet was dominated by only English content. Today the World Wide Web has grown wider and has become very large, having content in all Indian languages. And especially in Tamil, web has more than 10 million documents. With this huge amount of Tamil data available on web, we require systems which will enable users to easily search and access data. In the present paper we present a description of Searchko, a Tamil portal, whose objective is to provide Tamil users access to all Tamil content on web.

Searchko – an overview
Searchko provides various contents for the users. The search contents are classified into different domains such as Literature, Health, and Cinema search. The content for health domain is created in-house from the health texts. The results obtained for health search are classified into allopathy, siddha, homeopathy, ayurveda, when presented to the user. For example users interested in reading articles related to cancer, can give query as “puRRu noy” and obtain documents from ayurveda, allopathy, homeopathy, siddha. In the literature search, we can have focused search on contents of Tamil literatures of wide range from sangam literature to 20th century creations. This includes ‘aimperum kaappiyam’, ‘ettuth thokai’, devotional literatures such as ‘kuravanji paataalkal’, ‘pakhthi Italkkiiyangkal’, text having poems by Sidhars such as ‘sidhar paataalkal’ and present days kalki and Jayakanthan stories. The content for general search consists news articles from online news magazines, Tamil wikipedia, blog sites and other Tamil sites.
All these contents are crawled and indexed by search engine regularly and updated periodically. This process is fully automated. It has been observed that even though we have huge content in web in Tamil, people are not able to have access for these documents because most of the content is not available for search. By developing this portal we have made all the Tamil content on web accessible to the users.

Another important feature is the lexical resource most often referred by people of all age groups as dictionaries is available for search and there are English – Tamil, Tamil – English dictionaries. We have found that there are different online dictionaries available in the web, but most of these do not provide the exact meaning of words in all senses. For example for the word “bank”, There are different noun senses and verb senses such as “financial institution”, “river bank”, “to take support on someone or something”. Searchko provides online dictionaries, where all senses of the word are given along with the word’s part-of-speech category. The user interface of the dictionaries is very simple and user friendly. English – Tamil dictionary consists of more than 150000 root words. And similarly Tamil – English dictionary consists of more than 100000 words. The Tamil – English dictionary also includes old Tamil words, for which finding meanings in today’s new dictionaries is quite difficult. The English – Tamil dictionary includes Technical terms also. These dictionaries were created by lexicographers and verified by linguistic experts. These are very helpful to Translators.

Searchko provides entertainment contents focusing cinema, music and sports. In Cinema section, we provide details of actors, film directors, music directors, playback singers, and producers for each Tamil film. The Cinema database is updated whenever a new film is released. Users can search by giving actors names or film names or music director’s names or producer’s names. For example if the user searches for actor “Rajnikanth”, the Cinema search provides all films acted by Rajnikanth, for each film of his, will be provided with details of producer’s name, director’s name, playback singers name, Date of film release and also will fetch a relevant video of the film from You tube, if it exists. This cinema section also provides users with list of new upcoming films, their promotional videos if it exists. This can be considered as Tamil IMDB.

The music section of the portal is focused on music festival of “markazhi” month (Dec - Jan) also known “markazhi thiruvizhaa” at Chennai. Here we provide schedules of all music events that take place in various auditoriums or “sabhas” in and around Chennai. The schedules can be searched based on date or artist name or auditorium name or by time of event. The music lovers of the city have found this very helpful as they can find the time and venue of their interested event. This section also provides devotional hymns of Tiruppavai, Tiruvembavai and Tirupalliezuchi along with their meanings in Tamil. These devotional hymns are sung commonly during this markazhi month in various temples and households of Tamil Nadu.

In the Searchko, we provide live cricket scorecard in Tamil. The Cricket scorecard in English is obtained from websites such as cricinfo, cricbuzz, willow, and these scores are translated on-line to Tamil. The translation is done automatically. Here we use template based extraction and translation. For translating names of players, we use transliteration engine. The transliteration engine is built using statistical methodology [1]. The engine is trained using a parallel named entity lexicon. The engine works with an accuracy of 93%. The content in the whole web portal, is fetched and displayed automatically without any manual intervention.
The Natural language Processing modules of Searchko

Searchko has many unique features, which are developed using sophisticated natural language processing techniques.

a) Multiple Font Encoding Support:

Tamil content on web are in various encoding schemes and not as in English. There is a large data available in proprietary fonts. On giving a query the system searches all the Tamil pages independent (covering most of the proprietary fonts) of fonts to get the relevant pages. And snippets for all the retrieved documents are given in Unicode. To achieve this font independency in the search engine, we have used different transcoder engines to unify the contents in to Unicode. A font transcoder engine identifies the encoding scheme of one font and makes an equivalent map to another font encoding scheme. Here we have converted all proprietary fonts into UTF-8 encoding scheme. For the purpose of creating equivalent map, both encoding schemes have to be analysed and each glyph of proprietary font has to be mapped to the Unicode scheme. There can be one to many or many to one mapping.

b) Enhancing the query with the help of a morph analyzer:

Tamil being a morphologically rich language, multiple words can be generated from a given root word. For eg the word “padi” will have words like “padithaan”, “padithithaal” etc. Here in our search engine, we enhance the query with the help of a morphological analyzer to retrieve all documents which have the various forms of the given query word.

A morphological analyzer is a language processing tool, which will segment a given word into root word and it suffixes and will give their syntactic information respectively.

Example ‘viitukal’ -> ‘viitu’ + ‘kal’

Here in this task, the morphological analyzer is used only to get the root word. This is done using Finite State Automata and paradigm based approach [4]. This works with an average precision of 92.13%

For example if the query word is ‘malarkal’ the search engine will look for documents having words such as ‘malar’, ‘malarkala’ ‘malarai’, etc. This enables the user to give a query in crisp and easy form and get retrieved documents having all possible forms of the query word.

c) Query expansion using thesaurus and dictionaries:

We use a Tamil thesaurus and dictionary to retrieve more documents with the same sense. This increases the relevant documents [2]. For example, for the query word ‘puu’, documents are searched for other words with same sense like ‘malar’, ‘pushpam’, ‘alar’, ‘koNtai’, ‘sutaan’ along with the given query word ‘puu’.

The thesaurus is an advanced form of a lexicon, which contains the root words and their synsets in contrast to a lexicon which contain meaning and pronunciation of words. Here we have used a electronic thesaurus, which contains synsets of the root words.

The thesaurus and the dictionary contain around 75000 words. This resource is increased as and when we encounter new words. Here the search handles AND, OR and Quoted queries.
d) Works with a phonetic and on screen keyboard:

The main hinder for Tamil usage in Internet is keying the Tamil words. Here we have give two ways to input the Tamil query word. One for the users comfortable with keyboard, a phonetic keyboard and an on-screen keyboard for people comfortable with mouse.

Using the phonetic key board Tamil words can be keyed in by keying the English alphabets in correspondence with Tamil phone. Example

\[ \text{\textit{ஏமா}} \] is keyed in by tying ‘a’ ‘m’ ‘maa’.

In the on-screen keyboard all the letter and the glyphs used in Tamil are presented in a palette, as shown in the figure 1. By clicking the corresponding glyphs the query words are keyed in.

![Fig. 1. On-Screen KeyBoard](image)


e) Spell Checking:

We check the spelling of the input query. This is the first search engine for India languages with integrated spell check facility. The spell checker checks the query word and suggests possible words for error words. The spell checker is developed using Finite state Automata (FSA), which is popular for accurate and speed performance [3]. Here we have used corpus based methodology for validating the correctness of the words. The FSA is built is using individual letters of the words. The spell checker validates 10 words in less than a millisecond.

For example if the given query word is \[ \text{\textit{சோஞ்சு}} \] which is wrongly keyed in for, the suggested words are: ‘சோஞ்சு’, ‘சோஞ்சு’, ‘சோஞ்சு’, ‘சோஞ்சு’.

Evaluation

The search results have been evaluated using standard Information Retrieval metrics of MAP, P@10. For the evaluation purpose 25 test queries were taken and the results obtained are checked for relevance by comparing with the expected results as given by human evaluator. We obtained evaluation results as follows. The MAP score we got is 0.6 and P@10 (precision at 10) is 0.75 (ie for 10 documents we get 7.5 documents as relevant).

Conclusion

Here we have presented an overview of Searchko. This portal is dedicated for all Tamil people across the world. This portal is completely in Tamil. The user interface is very easy to use, which enables users to
navigate the web in Tamil, their native language. In this advanced language processing technology, systems have been used to obtain good results. The search is available at http://www.searchko.in.

Reference


கருந்தை அகழப்பு

பரள்ள நிலைகால வழக்க

தமிழ்மொழியில், எழுதி, புனர்மாற்றிகள், இலக்குணர்


c

தமிழ்ப் புத்தாண்டிற்குள் ஒன்று பெரும் புத்தாண்டில், நம்பிக்கை சரணை விளக்கச் செய்யப்பட்டு வருகின்றன. அரசு சரணை விளக்கச் செய்யப்பட்டு வருகின்றன. பொருளளும் விளக்கச் செய்யப்பட்டு வருகின்றன. 'செய்யுள்' சரணை விளக்கச் செய்யப்பட்டு வருகின்றன. பொருளளும் விளக்கச் செய்யப்பட்டு வருகின்றன. 'செய்யுள்' சரணை விளக்கச் செய்யப்பட்டு வருகின்றன. 'செய்யுள்' சரணை விளக்கச் செய்யப்பட்டு வருகின்றன. பொருளளும் விளக்கச் செய்யப்பட்டு வருகின்றன.


c

பரள்ள நிலைகால வழக்கத்தில், பாராட்டுவதற்கு வரும் இருப்புவரப்புக் கொள்கைத் தொடர்பான எடுத்துக்காட்டில், 'செய்யுள்' அவர்களை விளக்கம் செய்யப்பட்டு வருகின்றன. பொருள்லும் விளக்கம் செய்யப்பட்டு வருகின்றன. 'செய்யுள்' அவர்களை விளக்கம் செய்யப்பட்டு வருகின்றன. பொருள்லும் விளக்கம் செய்யப்பட்டு வருகின்றன. நாட்டு விளக்கம் செய்யப்பட்டு வருகின்றன. பொருள்லும் விளக்கம் செய்யப்பட்டு வருகின்றன. 'செய்யுள்' அவர்களை விளக்கம் செய்யப்பட்டு வருகின்றன. பொருள்லும் விளக்கம் செய்யப்பட்டு வருகின்றன. 'செய்யுள்' அவர்களை விளக்கம் செய்யப்பட்டு வருகின்றன. பொருள்லும் விளக்கம் செய்யப்பட்டு வருகின்றன.
11. (Agglomerative Clustering)
செல்லும் முக்கியத்துவம் கொண்ட விளக்கங்கள்(cosine) பிறந்து புள்ளியும் வரிசைக்குரிய விளக்கங்களும். அதிக அளவில் வெளியில் விளக்கங்கள் மாற்றும் அளவில் விளக்கங்கள் திறனற்றம் விளக்கங்கள் வைக்கப்பட்டுள்ளது. இவை வெளியில் அனைத்து விளக்கங்கள் வைக்கப்பட்டுள்ளது. இவை வெளியில் அனைத்து விளக்கங்கள் வைக்கப்பட்டுள்ளது. 

இவை வெளியில் வைக்கப்பட்டுள்ளது. மற்றும் வெளியில் (data conversion) வெளியில் வைக்கப்பட்டுள்ளது(indexer) வெளியில் வைக்கப்பட்டுள்ளது(indexer) வெளியில் வைக்கப்பட்டுள்ளது(indexer) வெளியில் வைக்கப்பட்டுள்ளது(indexer) 

<table>
<thead>
<tr>
<th>இடம்</th>
<th>இடம் செயல் மற்றும் இடம் தேவைமிகம்</th>
</tr>
</thead>
<tbody>
<tr>
<td>சென்னை</td>
<td>chennai(icl&gt;place)</td>
</tr>
<tr>
<td>புதுவை</td>
<td>puduvai(icl&gt;place)</td>
</tr>
<tr>
<td>மதுரை</td>
<td>madurai(icl&gt;city)</td>
</tr>
<tr>
<td>செல்லும் முக்கியத்துவம்</td>
<td>go(icl&gt;do)</td>
</tr>
</tbody>
</table>


cosineவை வெளியில் வைக்கப்பட்டுள்ளது (transliteration) chennai(icl>place). பொதுவாக chennaiவை வெளியில் வைக்கப்பட்டுள்ளது. இப்பதிவுக்காக இடம் இடம் வைக்கப்பட்டுள்ளது. 

முடிவு 3.1 என்னும் முக்கியத்துவம் வைக்கப்பட்டுள்ளது.
3.1 (Binary Search Tree)

3.2
3.2 இயற்றும் வழிவகை

தனு மருத்துவத்திற்காக குறிப்பிட்டும் கருவிகளைவழங்குவதற்கான, இரண்டு (இரண்டு), இணைய வழக்கு அதிர்வார் குறிப்பிட்டும் (index), இரண்டு வழியாக அதிர்வார் குறிப்பிட்டும் (index). ஒவ்வொரு வழியுடனும் ஒரு குறிப்பிட்டும் வழக்கம் செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும். இதற்காக முன் வழக்கம் செய்ய வேண்டும் மற்றும் கருவியின் மூலமாக வழக்கமே செய்ய வேண்டும்.
4.1 செயல்பாடுகள்

'குறிப்பிட்டு பதிவுசெய்யுங்கள' விளக்கத்தின் நேர்மை சிக்கல் குறிப்பிட்டு வெளியான விளக்கங்களைச் சுருக்கி முடிகிறது. மேலும் பதிவுசெய்யுங்கள் விளக்கத்தின் நேர்மை ஸ்பெஷல் அடிப்படையில் விளக்கத்தின் நேர்மை லையான முன்னாட்டு செயல்பாடுகள் முன்னாட்டு நேர்மையைக் குறிப்பிட்டு. மேலும் பதிவுசெய்யுங்கள் விளக்கத்தின் நேர்மை ஸ்பெஷல் அடிப்படையில் விளக்கத்தின் நேர்மை லையான முன்னாட்டு செயல்பாடுகள் முன்னாட்டு நேர்மையைக்குரியவற்றை செய்துள்ளன. மேலும் விளக்க எடுத்துக்காட்டு: "சிக்கல் குறிப்பி" வெளியான விளக்கத்தின் நேர்மை ஸ்பெஷல் அடிப்படையில்: முதல் 33,721 விளக்கத்தின், 1,36,913 விளக்கத்தின் நேர்மை ஸ்பெஷல் அடிப்படையில். பின்னர் 41,721 விளக்கத்தின் 1,88,618 விளக்கத்தின் நேர்மை ஸ்பெஷல் அடிப்படையில்.
4.2 Discussion

In this section, we discuss the performance of various methods in terms of their ability to identify entities. We see that our method outperforms the baseline methods in all cases. The results show that our approach is effective in identifying entities accurately.

\[ \text{Accuracy} = \frac{\text{Precision@5} \times \text{Precision@10}}{\text{Precision@5} + \text{Precision@10}} \]

As shown in Table 4.2, our method achieves significantly higher accuracy compared to the baseline methods.

5. Conclusion

Our results indicate that our method is highly effective in identifying entities. Further research is needed to explore the scalability and efficiency of our method.

Reference

CoRe - A Framework for Concept Relation based Advanced Search Engine

T V Geetha, Ranjani Parthasarathi & Madhan Karky
{tv_g@hotmail.com, rp@annauniv.edu, madhankarky@gmail.com}
Department of Computer Science & Engineering
College of Engineering Guindy
Anna University

Abstract

The number of Tamil documents is growing rapidly every second. Blogs, news portals, infonets and socialnets have considerably contributed to this growth. Traditional keyword based searches such as Google or Bing, primarily developed for English are now being used to search Tamil documents. These search engines do not have a full-fledged support for Tamil language. We present CoRe, the world’s first framework for a concept relation based search engine for Tamil. CoRe search, unlike traditional keyword based searches, identifies the concepts in a document and their relationship with each other using UNL (Universal Networking Language). UNL is essentially a semantic relation based intermediate representation that has been used in this work for search index representation. This paper presents the primary components of the CoRe framework, namely, EnCoRe(enconverter), CoReX(indexer) and CoReS(Search & Rank) modules. The CoRe based search is tested with over 50,000 crawled web documents and the results are compared against a traditional keyword based search algorithm over the same set of documents. It is observed that the UNL based search has a precision accuracy of 75%

8. Introduction

With UNICODE being accepted internationally as a standard for representing text on web sites, search engines are now indexing the pages independent of the language. Popular search engines like Google, Bing and Yahoo now readily index Tamil documents. Only Google manages to include certain language dependent features such as stemmer for Tamil. Even though keyword based search seems very convenient system to retrieve web documents it fails in two major issues. The first one is in understanding the document. The second one is in understanding what the user wants.

Let us start with a very simple example where a user wants to know about a temple in the city of Madurai. The user who does not know the actual name of the temple will use keywords Madurai and Temple in a search engine, which is very likely to retrieve documents that have mentions of both these keywords. What the traditional search engine will not be able to retrieve are documents that talk about Meenakshi temple which do not contain any of the search keywords. This failure of traditional search engine expects the user to reach these documents in their second or third search attempts with different keywords.

Personalised search engines aim to solve this problem from the user side. Concept based search engines aim to solve the same problem from both the ends. In this paper we propose CoRe, a framework for a Concept – Relation Based Search. The CoRe framework aims at understanding the web documents by
indexing them based on concepts and their relations rather than indexing the keywords and their frequencies. The framework was implemented and tested with over 50,000 documents from tourism domain. The search results were compared against traditional keyword based search for precision and relevance.

This paper is recognize into five sections. The second section provides a brief survey of literature relevant to concept based search subsystems. The third section presents the CoRe framework and its components. The fourth section discusses the results of CoRe search engine and compares it against traditional keyword based search for precision and relevance. The fifth section summarises the paper, presents the work in progress and discusses future directions.

2. Background

Meaning based or concept based search and indexing techniques have been proposed for English and other languages. The main purpose of such indexing techniques is cross lingual information retrieval. Universal Networking Language (UNL), proposed in our CoRe framework for enconversion of Tamil documents, is described in [3]. UNL is an interlingua framework that was originally designed to aid the machine translation process[3]. An UNL enconverter interprets natural language and converts it into an intermediate representation that utilizes language independent semantic concepts and relations prescribed by UNL. The enconversion from the source language to UNL analyses the source language and utilizes language specific linguistic rules for building the UNL representation. On the other hand the UNL deconverter transforms UNL representation to a target language. For this purpose the deconverter uses language specific generation rules to produce the target language.

Kang in [4] proposes a indexing technique where he indexes documents based on the concepts identified in the document. A similar work where medical images are conceptually indexed can be found in [5]. Chau and Yeh in [2] proposes another conceptual index again based on concepts. Here Chau and Yeh have designed the index especially for cross lingual text retrieval. Surve et al., in [7] propose AgroExplorer, a meaning based multi lingual search system. Though these systems have the advantage of concept based indexing they do not investigate deep into the relations between the concepts in a sentence.

3. CoRe Search Framework

CoRe search framework presented in figure 1, can be divided into two major divisions, online and offline, in terms of the time of processing. Three major subsystems EnCoRe, CoReX and CoReS form the backbone of the framework providing the major functionality. Tamil language tools are separated from rest of the system to offer language dependent services such as analysers and generators. This section describes the various components of the framework in detail.

3.1 Offline Processing

The offline mode comprises of operations related to crawling Tamil web documents relevant to a particular domain, processing the raw document for extracting sentence constituents, converting the sentence constituents to corresponding UNL graphs and building the concept relation index. This is a periodic scheduled process and will be modifying the index incrementally as new documents are being crawled.
3.2 Online Processing

A user’s query is processed, expanded and converted to UNL graph(s) and sent to a search and ranking subsystem where the documents that match concept relation similarity are ranked and sent for output processing. The output processing module formats the output, generates snippets and summary for the retrieved documents and sends it to the user.

![CoRe Search Framework](image)

Fig 1: CoRe Search Framework

3.3 Tamil Tools

The Tamil tools package offers a set of language processing services for different components of the CoRe framework. Morphological Analyser [1] is used by most of the components for morphologically recognize a word. The rules for the EnCoRe subsystem depends on the results of the morphological analysis. Word Sense Disambiguator(WSD), a rule based tool, is used to resolve the meaning of a word based on it’s context. This tool is used both by EnCoRe and Query Expander. Morphological Generator is used for generating natural language sentences. Output Processor uses Morphological Generator to generate natural language summary for a given document. Query Expander and EnCoRe use spell checker to auto correct typos and basic morphological errors. Named Entity Recogniser and Multiword Recogniser are used to identify named and multiword entities respectively. These rule based tools are used by EnCoRe and Query Expander to identify entities. The Tamil Tools also comprises of a Universal Word list(UW list) [3] and a Multiword List(MW list). These lists carry the domain related UW and MW words respectively.
3.4 Focussed Crawler

Tamil web documents specific to a certain domain are fetched into the system by a Focussed crawler. The domain specific words are fed to the crawler along with a seed URL list. The documents fetched by the crawler are sent to the Document Pre-processor for further processing.

3.5 Document Pre-processor

Document Pre-processor parses Tamil documents in HTML format for textual content removing links and other unwanted tags. The Pre-processor also identifies important sentence constituents and send the sentence constituent along with a document id to the EnCoRe subsystem.

3.6 EnCoRe

The EnCoRe subsystem forms the heart of the CoRe framework. Here a sentence constituent is passed to a rule based system to identify the various concepts in the constituent and the rules are used to identify one of the 44 UNL relations [3]. EnCoRe uses language processing tools such as the Morphological Analyser to recognize various morphological suffices of a word and uses this information along with syntax and semantics to identify the relationship between concepts. UNL graphs are generated for every sentence constituent. The UNL graph is then sent to CoReX indexer along with information such as document ID, positional index and original keyword, its frequency in the document etc. to be used by the CoReS subsystem.

3.7 CoReX Indexer

The CoReX Indexer subsystem presented in [6] stores and manages the UNL graphs in three different indices. Concept only index(C index), Concept-Relation index(CR index) and Concept-Relation-Concept index(CRC index) are the three indices maintained by the CoReX indexer. The UNL graphs are stored in the indices by their concept ids for efficient retrieval. The CoReX index structure and efficiency analysis is provided in [6].

3.8 Query Expander & Expansion Builder

Any user query is directly sent to the Query Expander module that expands the query using the data from Expansion Builder. The Expansion Builder uses the CoReX index to build on-the-fly similarity thesaurus and co-occurrence list. Query Expander enconverts the user query to a UNL graph for the CoReS subsystem, using the information provided by the Expansion builder.

3.9 CoReS

CoReS subsystem provides the CoRe framework with the functionality of searching and ranking documents based on concepts and relations. Unlike traditional algorithms that scores pages based on terms and their frequencies, CoReS subsystem ranks a document based on the concepts and the type of relations that exist between those concepts.

3.10 Output Processor

Results from CoReS subsystem would be a set of documents and their corresponding weight with respect to the user query and expansions. Output processor formats the result page by generating snippets highlighting the identified concepts and also generates a template-based summary for every page. For instance in a tourism domain, the page summary would contain information about tourist spots, contact
numbers, animals, hotels, transport and more. Similarly a summary for a health domain may contain hospital info, medicine information, emergency contacts, symptoms of a disease and so on.

4. CoRe Search Results & Analysis

A search engine based on CoRe framework was implemented by modifying the Nutch open source search engine. Over 50,000 documents in tourism domain were crawled and enconverted and indexed for search. An implementation of a Tamil keyword based search built with Nutch over the same set of documents and domain is taken for comparison.

4.1 Matrix Layout

In this paper, we propose a Matrix Layout for displaying the results to the user. The Matrix Layout shown in figure 2, displays the results in a 3X2 matrix cells with each cell corresponding to a class of results based on the concepts and relationship between concepts. Figure 2 displays the results for the query ஊன்நூறு பிறக்குநோம் கோயில் (thanjai birakatheeswarar koayil) The first cell displays the results pertaining to the concepts that contain the actual keywords and sorted by the relation they have between them. Second cell identifies results that contain at least one concept with the actual keyword. The third cell identifies documents that will not be identified by traditional keyword search where none of the concepts contain keywords. In this case it retrieves documents with ஊன்நூறு கோயில் (thangaavoor & koavil) both of which do not form part of the actual query. The fourth and fifth cells are based on expansions of the query. Here they display results corresponding to பேரியா கோவில் & திருக்கோவில் (periya koavil & thirukkoavil). The final cell identifies the place associated with the query term to display the map of the corresponding place. The snapshot presented in figure 2 is from our CoRe search engine implemented from the CoRe framework.

4.2 Performance Evaluation

Precision of documents can be computed using the formula given below[8]. We compute the precision of documents for the first 5, first 10 and first 20 documents.
The average precision and mean average precision for a set of queries will indicate the performance of the system.

\[
\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}
\]

Fig 3 : Average Precision Comparison  Fig 4 : Mean Average Precision Comparison

For a set of 100 queries, precision is calculated at three levels for both the CoRe based search and keyword based search. The results are presented as a graph in figure 3. The mean average precision (MAP) [8] score comparison of the two search paradigms is given in figure 4.

8. Conclusion and Future Work

This paper describes CoRe, a framework for concept relation based search in Tamil. Different subsystems and components of the framework are described in detail. Results from an implementation of CoRe framework is provided and is compared against traditional keyword based search results. The rules of EnCoRe subsystem are very much domain specific. Expansion of the search engine to adapt to other domains will be future work. Integrating cross lingual results and providing cross lingual summary for the results will take this work to its next level.

References

15

தம்பல்கண்ட கோவிலுக்கின் தமிழ்
தமிழ் நீண்டகால அணுக்கா

(annakannan@gmail.com)

இயற்பியல் கல்வி விளைந்து நீண்டகா அணுக்கா, பொறியியல் நோய்களை தடுளிப் புகழ்தொடர். 2010 முதல் 58.43 வருடாக நீண்டகா முக்கியமாக வருகிறார்.1 ஆண்டுகளுக்கு ஏதும் தமிழ் கல்விகளை மேற்கத்துவம், புகழ்பெற்றவை போன்றவை வாழ்பெறுகின்றன.

சாதாரண அணுக்கா நீண்டகா, என்று பலரும் குறிப்பிட்டுகின்றனர். ஆனால், இந்த இயற்பியல் நோய்களை விளைந்து நீண்டகா ஆண்டலை குறிப்பிட்டார்.2 ஆண்டுகளாக விளைந்து நீண்டகா பெண்களின் அனைத்தும் மாணவர்களை நோய்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா பெண்களின் அனைத்தும் மாணவர்களை விளைந்து நீண்டகா.

உடல் திறங்கள் அடர்ந்த நோய்கள், குறுக்கு நோய்கள் 500. சாதாரண பொழிவு நீண்டகா முக்கியமாக இழுத்திருக்கின்றன.2 ஆண்டுகளாக அணுக்கா விளைந்து நீண்டகா பெண்களின் மேற்கத்துவம், முதல் உடல் திறங்கள், என்று பலரும் குறிப்பிட்டுகின்றனர்.3 பெண்களான இணையாசை நோய்களை, நீண்டகா நோய்களான தம்போயனை விளைந்து நீண்டகா வருகிறார், 2010 இல் என்று பலரும் குறிப்பிட்டுகின்றன.4 ஆண்டுகளாக விளைந்து நீண்டகா பெண்களின் மாணவர்களை விளைந்து நீண்டகா பெண்களின் மாணவர்களை விளைந்து நீண்டகா பெண்களின் மாணவர்களை விளைந்து நீண்டகா.

செயலான அடுத்த பகுதியில்

இயற்பியல் கல்வி விளைந்து நீண்டகா தமிழ் மாணவர்களிடையே முதல் 60 வருடாகவிலகு அணுக்கா விளைந்து நீண்டகா முக்கியமாக இழுத்திருக்கின்றன. இவ்வநோய்கள் விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா.

1. தமிழ் சுட்டகம் - என்ன ஆராய்ச்சி

உருமான புலனுக்காக, அணுக்கா முக்கியமாக விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணவர்களை விளைந்து நீண்டகா மாணா
2. புத்தரிழிக்கோள் - கட்டுப்பாடு செய்யக்கூடிய குடும்பங்கள், பொதுமக்கல் போன்றவற்றுடன் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லுனர்களும் பண்பாடு கொள்வதற்கு முன்னே அவ்விரு வல்லுனர்களும் போன்ற மற்றும் பொது வல்லு

804
தல பெருமையை நிறுவகிறது. ஐந்திடம் நிறுவிய இவ்வருடமையில் ஒரு கட்டக்க நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

10

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

11

தல பெருமையை நிறுவகிறது. ஐந்திடம் நிறுவிய இவ்வருடமையில் ஒரு கட்டக்க நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

12

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

13

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

14

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

15

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது. 

16

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு

மாணவர்களுக்கு கூறும் ஆசிரியர் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய நுண்பொருள் நூற்றாண்டு நூற்றூர் நடத்தி நிறுவிய தல பெருமையை நிறுவகிறது.
3. **Phrases**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASL</td>
<td>Age, Sex, Location</td>
</tr>
<tr>
<td>ASAP</td>
<td>As soon as possible</td>
</tr>
<tr>
<td>BBL</td>
<td>Be back later</td>
</tr>
<tr>
<td>BRB</td>
<td>Be right back</td>
</tr>
<tr>
<td>LOL</td>
<td>Laughing out loud</td>
</tr>
<tr>
<td>ROFL</td>
<td>Rolling on the floor laughing</td>
</tr>
<tr>
<td>BTW</td>
<td>By the way</td>
</tr>
<tr>
<td>OIC</td>
<td>Oh, I see</td>
</tr>
<tr>
<td>CUL</td>
<td>See you later</td>
</tr>
<tr>
<td>OTOH</td>
<td>On the other hand</td>
</tr>
<tr>
<td>GMTA</td>
<td>Great minds think alike</td>
</tr>
<tr>
<td>IMHO</td>
<td>In my humble opinion</td>
</tr>
</tbody>
</table>

**Notes:**
- ASL stands for **Age, Sex, Location**.
- ASAP stands for **As soon as possible**.
- BBL stands for **Be back later**.
- BRB stands for **Be right back**.
- LOL stands for **Laughing out loud**.
- ROFL stands for **Rolling on the floor laughing**.
- BTW stands for **By the way**.
- OIC stands for **Oh, I see**.
- CUL stands for **See you later**.
- OTOH stands for **On the other hand**.
- GMTA stands for **Great minds think alike**.
- IMHO stands for **In my humble opinion**.
 அறிவுச் செயல்கள் அறிவுச் செயல்கள் அறிவுச் செயல்கள் அறிவுச் செயல்கள் அறிவுச் செயல்கள் அறிவுச் செயல்கள்

<table>
<thead>
<tr>
<th>முறை அறிவுச் செயல்கள் முறை அறிவுச் செயல்கள் முறை அறிவுச் செயல்கள் முறை அறிவுச் செயல்கள் முறை அறிவுச் செயல்கள்</th>
</tr>
</thead>
<tbody>
<tr>
<td>அடியுரு. பி. பி.</td>
</tr>
<tr>
<td>அணி டிந் அறிமுகன் முருகேந் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>எஸ். விளிம்பு.</td>
</tr>
<tr>
<td>எஸ். விளிம்பு.</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
<tr>
<td>முறை முதல் புற்செய்திகள் குமுரு</td>
</tr>
</tbody>
</table>

807
4. செவாம்பில் கூட்டி உலக நூற்றாண்டிகள்

செவாம்பில் கூட்டியும் ஆண்டுகளில், மக்களால் மற்றும் தமிழகத்தில் அடிப்படையாக .

நவம்பிரம் - முதல் அண்ட்ஸ் போற்றம்

செவாம்பில், உலகவியல் போற்றம், செவாம்பில் கூட்டியால் அபெர்சன நூற்றாண்டிகள் நடைபெற்றார். 2004 ஆம் வர्तமான முதல் அண்ட்ஸ் போற்றம் முதல் அண்ட்ஸ் போற்றம் முதல் அண்ட்ஸ் போற்றம் முதல் அண்ட்ஸ் போற்றம், 2005ஆம் வருடம் “Most Innovative Mobile Application” ஓர் போற்றம், உலகவியல் அண்ட்ஸ் போற்றம் “Malaysian ICT Excellence Award” ஓர் நடைபெற்றார். செவாம்பில் கூட்டியால், மற்றும் தமிழகத்தில் முதல் அண்ட்ஸ் போற்றம் 2009ஆம் வருடம் தொடங்கப்பட்டது.18

செவாம்பில் நூற்றாண்டு - பெண் நூற்றாண்டு

செவாம்பில் (www.fublish.com) புத்தபோற்றம் விருது செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம் செவாம்பில் செவாம்பில் புத்தபோற்றம்

செவாம்பில் நூற்றாண்டு புத்தபோற்றம் 2004 ஆம் வருடம் புத்தபோற்றம் புத்தபோற்றம் புத்தபோற்றம் புத்தபோற்றம் புத்தபோற்றம் புத்தபோற்றம் புத்தபோற்றம் புத்தபோற்றம் .

செவாம்பில் நூற்றாண்டு புத்தபோற்றம் செவாம்பில் புத்தபோற்றம் செவாம்பில் புத்தபோற்றம் செவாம்பில் புத்தபோற்றம் செவாம்பில் புத்தபோற்றம் செவாம்பில் புத்தபோற்றம்

தமிழ் செவாம்பில் தமிழியல் குழுக்கள்

தமிழ் செவாம்பில் தமிழியல் குழுக்கள், அல்லது தமிழியல் குழுக்கள் தமிழ்நாட்டின் பல குழுக்களில் பல குழுக்களில் பல குழுக்களில் பல குழுக்களில் பல குழுக்களில் பல குழுக்களில்
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>*</td>
<td>0</td>
<td>#</td>
</tr>
</tbody>
</table>

**குறிக்கோள்: குறிப்பிட்டுள்ள செயற்கைக் குறிக்கோள்கள்**

<table>
<thead>
<tr>
<th>குறிப்பிட்டுள்ள செயற்கைக் குறிக்கோள்</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrow key</td>
</tr>
<tr>
<td>Audio Response Device</td>
</tr>
<tr>
<td>Auto Dial</td>
</tr>
<tr>
<td>Auto Text</td>
</tr>
<tr>
<td>Back Space</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>BIOS - Basic Input / Output</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Cache Memory</td>
</tr>
<tr>
<td>Caller ID</td>
</tr>
<tr>
<td>Card Reader</td>
</tr>
<tr>
<td>Expanded Memory</td>
</tr>
<tr>
<td>Icon</td>
</tr>
<tr>
<td>Missed call</td>
</tr>
<tr>
<td>Remote access</td>
</tr>
</tbody>
</table>

5. மெச்சோர் குறிப்பிட்டு நெறுவல் முறைகள்

செயற்கையான பாராள் தாண்டல் ஓய்வுக்கள், ஆண்மையில் பெண் முனை முறைகளை நெறுவல் முறையை பயன்படுத்தும் நடவடிக்கைப் பிள்ளை. அதற்கு என் முனையை முனையில் பயன்படுத்தும் நடவடிக்கையை நெறுவல் முறையை கொண்டு கொள்ளப் பயனுள்ள நடவடிக்கையை பார்வைகளாக நெறுவல் முறையை கொண்டு கொள்ளப் பயனுள்ள நடவடிக்கையை பார்வைகளாக நெறுவல் முறையை கொண்டு கொள்ளப் பயனுள்ள நடவடிக்கையை பார்வைகளாக நெறுவல் முறையை கொண்டு கொள்ளப் பயனுள்ள நடவடிக்கையை பார்வைகளாக நெறுவல் முறையை கொண்டு கொள்ளப் 

- எளிதான் முனையை முனை விளக்க விளக்க
- எளிதான் பொருட்களைப் பிறக்க விளக்க
- எளிதான் பொருட்களைப் பிறக்க விளக்க

6. மெச்சோர் குறிப்பிட்டு நெறுவல் முறைகள்

தேர்த்தியான குறிப்பிட்டு நெறுவல் முறைகள், எளிதான் முனையை முனை விளக்க விளக்க, எளிதான் பொருட்களைப் பிறக்க விளக்க, எளிதான் பொருட்களைப் பிறக்க விளக்க, எளிதான் பொருட்களைப் பிறக்க விளக்க, எளிதான் பொருட்களைப் பிறக்க விளக்க, எளிதான் பொருட்களைப் பிறக்க விளக்க, எளிதான் பொருட்களைப் பிறக்க விளக்க, எளிதான் பொருட்களைப் 

- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
- ஆண்மை உச்ச ஆயல் பிரிவு யாளாக ஆண்மை |
உற்பத்தியாளர்கள் செல்வதற்கு உயர்ந்த பல்வேறு சமயங்கள் உள்ளன. இது செயற்பாடு, செயற்பாடுகள் உற்பத்தியாளர்களுடன் செயற்பட்ட செயற்பாடுகள் உள்ளன. இது உற்பத்தியாளர்களுடன் உயர்ந்த பல்வேறு சமயங்கள் உள்ளன. இது உயர்ந்த பல்வேறு சமயங்கள் உள்ளன. 

இன்றுவரை தமிழில் அக்காலத்தில் இயங்கும் உற்பத்தியாளர்கள் தமிழாகவும் இயங்கும். 

செயற்பாட்டில் அரசு உற்பத்தியாளர்களுடன் செயற்பாட்டின் சோதனை செயற்பாடு உற்பத்தியாளர்கள் தமிழாகவும் இயங்கும். 

இல்லாத உயர்வூட்டு வல்லாம்பு மற்றும் உயர்வூட்டு வல்லாம்பு உற்பத்தியாளர்கள் தமிழாகவும் இயங்கும். 

இல்லாத உயர்வூட்டு வல்லாம்பு மற்றும் உயர்வூட்டு வல்லாம்பு உற்பத்தியாளர்கள் தமிழாகவும் இயங்கும். 

இல்லாத உயர்வூட்டு வல்லாம்பு மற்றும் உயர்வூட்டு வல்லாம்பு உற்பத்தியாளர்கள் தமிழாகவும் இயங்கும்.
6. IRCTC announces ngpay as the leading mobile sales channel for purchasing Rail tickets, http://www.ngpay.com/site/pressroom2.htm
7. tnresults.nic.in SSLC Results 2009 10th Results of Tamilnadu, http://ready2beat.com/educational/tnresultsnic-sslc-results-2009-10th-results-tamilnadu
17. நிருவுரவாளர், http://sellinam.com/?page_id=2
Thirukkural Mobile
A Cultural Tool for All

G. Bhuvan Babu, MFA.
bhuvan@avonmobility.com
ph – 98416 81233

Introduction

Thirukkural is not an imaginary utopia or a heaven for human beings to aspire for. It is out and out concerned with today’s world. Over the centuries Thirukkural has become a muchquoted, quintessential guidebook for life. They are heard in the speeches of ministers, seen written above the driver seat on most of the buses that range Tamil Nadu and small children memorize the Kural in order to be able to chant it verse after verse many can recite the entire 1,330 verses by heart.

But still, is this masterpiece only to teach vaguely in schools for the sake of memorising and to be used as quotations by speakers? No. Not at all. It was written to enrich human values. Thiruvalluvar’s intuition and insight is solely concerned with human life in the present world.

Gandhiji was one of those, who realized the greatness of Thirukkural and its need for enlightenment in the society. He said: "I learnt Tamil only to enable me to study Thiruvalluvar’s Kural through his mother tongue itself". He further said: "Only a few of us know the name of Tiruvalluvar. The North Indians do not know the name of the great saint. There is none who has given such a treasure of wisdom like him."

This statement was given six decades before and it still remains true. Though many of us know Thirukkural and thiruvalluvar, only few people realize its values and try to follow it in day to day life in this Tamil speaking land.

Objective
The objective of this paper is to reach/teach this cultural treasure to the common man in the best possible way

• with a greater impact on its relevance and significance in this modern world
• in a simple, understandable language, making learning interesting and fun
• to apply and practice its values and ethics in day to day business and life style situations.

Research & Results

Before finding ways to achieve the objectives, it is very essential to acknowledge our Government and Tamil Scholars for their great efforts to create awareness of Thirukkural in the common man. We worship Thiruvalluvar as a great saint, the universal teacher of our lives. Valluvar Kottam is erected with all 1330 couplets inscribed in it. Each year, on Thiruvalluvar day, we celebrate him with a public holiday. We have erected a beautiful shrine to him in the midst of a garden in Mylapore, every year celebrating a great
festival there. The greatest physical monument, however, is the magnificent 133-foot high statue of the saint stood majestically off the shores of Kanayakamari, the most southern point in India standing like some spiritual lighthouse watching over the world.

With the painstaking efforts of Kural Lovers, Thirukkural is also credited to be one of the early adopters of Information Technology. Various attempts have been made to popularize Thirukkural in the form of books, music, e-versions, paintings, radio and television programmes, mobile phones, etc.

**Books:** Numerous translations and descriptive books are found in the market.

**Music:** Thirukkural in the form of musical CDs are available.

**E-Versions:** E-Books, Research papers, blogs and interactive content are found in websites abundantly.

**Paintings:** Many artists have visualized kurals in the form of paintings. (exhibited at Valluvar Kottam)

**Radio and Television:** A Kural a day, serials, debates etc.

**Mobile Phones:** Off late, Thirukkural is introduced into mobile phones in text format.

Since mobile phones are becoming an indispensible gadget irrespective of social status in the present scenario, it would be the right means to achieve our objectives. Let’s go through some statistics on mobile phones in India before proceeding further.

*Statistics on mobile phones in India shows that rural mobile phone users alone in India cross 100 million this year (2010). With 3G (3rd Generation mobile technology) in India on the front; it is going to be an exciting phase for the rural and urban Indian telecom users. Rural mobile phone users can download content on a click from a link sent, until 3G technology spreads its wings there. The 3G bonanza, which will offer very high-speed mobile wireless services, already has a potential market in India with at least 20 million 3G-enabled phones. Videos can be pushed through video calls which would be a low cost version of watching a video without needing any mobile data plan. This has the capability to reach the masses and is as easy as making a call on your phone. Mobile techno gurus say that the number of people who opt for 3G services would be determined by content.*

So, with the entry of 3G mobile technology and multimedia (audio and video) on mobile phones, we believe that Thirukkural will be the best content in the form of education, entertainment and enlightenment at greater heights.

**Recommendations**

Thiruvalluvar says: “Even though the technical know-how of modern technology is known well, the inherent attitude and the actual prevailing condition of the State should be thoroughly understood and only then the technology should be put to application” (637). With this in mind, we have developed a complete new dimension for Thirukkural on mobile phones. This is a service that helps one understand Thirukkural in depth. It gives a Thirukkural every day through your mobile phone, using animation and audio to make learning interesting and fun. Every Thirukkural Video has a logical sequence of –

- Introduction of Kural,
- Animation and musical rendering of kural to the concept and
- Kural’s meaning in Simple Tamil/Hindi followed by the English meaning.
Key Benefits

The animated musical cards provide the intended meaning conveyed by Thiruvalluvar. Mellifluous music and beautiful animations. Uneducated rural public can understand the meaning and value of Thirukkural with ease. This will certainly be useful not only to Tamil speaking people but also to others to understand the richness of Tamil culture and its heritage. Also a best friend to the Tamil community, living in various parts of the world, who are not fortunate to educate the new generation in Tamil. This adds value for Tamil as a classical language (Semmozhi).

How it works?

The customer gets a multimedia Kural every day from the Telecom service providers through a Video Call / link. It is as easy as receiving a call. At one click, the video streams/downloads into the mobile phone and plays a kural with animation, and music, followed by its meaning in a simple language. With the capability to reach the masses through mobile phones and 3G connections, this service is available on a monthly subscription basis for a small fee.

Conclusion

Thiruvalluvar says:

Resources, means, time, place and deed,
examine these five before you proceed. – 675

We believe that with Thirukkural itself as the best resource, technology as the means, this is the right time and place to enhance the lifestyle of every individual to create a better world.
ஒவ்வொரு செயற்பாட்டின் நோக்கமூற்றம் தமிழ் இழக்கம்

முசிவலிங்கம்
musivalingam@gmail.com

செயற்பாட்டுக்கான வித்யாலை (சேனல்),
பென்னோன்பரா (செயற்பாட்டு நோக்கம்) - செயற்பாட்டு.

சமயத்தின் புராணத்தின் கருத்துக்குசிய அரியனமைனாகத் தமிழுக்கு கருத்திலிருந்தது. அவ்வீரம் இந்தப்பாட்டின் கருத்திலிருந்த 'தமிழ் முன்னாள்' (Information Era) என்ற கோளிலிருந்து. உலகில் விளங்கும் கருத்திலிருந்த இந்து இத்தகைய அரியன குழுவாகத் தமிழின் வருமாகில் பயன்படுத்தப்படும்; குழுவின் நிறுவனரும் இத்தகைய கருத்திலிருந்த பயன்படும். குழுவின் பயன்படுத்தப்படும் கருத்திலிருந்த பயன்படும்.

செயற்பாட்டு காட்டுப்பாடு

செயற்பாட்டுக் காட்டுப்பாடு கருத்திலிருந்து இருசமயம் கருத்திலிருந்து இருசமய இருசமயத்தின் கருத்திலிருந்து புராணத்தின் கருத்திலிருந்தே. குழுவின் உலகில் விளங்கும் கருத்திலிருந்து பயன்படுத்தத்திலிருந்து முதலிலிருந்து. 2010 இருசமயம் 1-இல் இருசமய கருத்திலிருந்து இருசமய கருத்திலிருந்து பயன்படுத்தும் கருத்திலிருந்து 60 இருசமயம் கருத்திலிருந்து. இருசமய கருத்திலிருந்து பயன்படுத்தும் கருத்திலிருந்து 51.05% அதிகம். இருசமய கருத்திலிருந்து பயன்படுத்தும் கருத்திலிருந்து 56 இருசமயம் 37 இருசமயம் 30 இருசமயம். இருசமய கருத்திலிருந்து 47.91% அதிகம். குழுவின் உலகில் விளங்கும் கருத்திலிருந்து பயன்படுத்தும் கருத்திலிருந்து இருசமய கருத்திலிருந்து பயன்படுத்தும் கருத்திலிருந்து 59 இருசமயம் 70 இருசமயம்.

குழுவில் செயற்பாட்டு 5 இருசமயம் 22 இருசமயம் 34 இருசமயம் பயன்படுத்தும் சேனல். குழுவில் செயற்பாட்டு 70% கருத்திலிருந்து. எடுத்துக்காட்டு குழுவில் செயற்பாட்டு 70 இருசமயம் கருத்திலிருந்து. செயற்பாட்டுக்கு பயன்படுத்தும் சேனல் கருத்திலிருந்து செயற்பாட்டில் பயன்படுத்தும் கருத்திலிருந்து. இருசமயம் செயற்பாட்டு செயற்பாட்டு கருத்திலிருந்து. குழுவில் செயற்பாட்டு இருசமயம் கருத்திலிருந்து இருசமயம் கருத்திலிருந்து. இருசமயம் செயற்பாட்டு இருசமயம் கருத்திலிருந்து.
ேனியலுக்கு கடுமையான செயல்பாடு மேல்புறம் மாட்டும் திறனூடுக்கள் 'மாத்து தொடரும்' (User Friendly) உள்ளதும் இல்லத்து வல்ல திறனூடுக்களும் உள்ளன. நல்லவை, திறனூடுக்களும் உள்ளார். நல்லவை தொடரும் திறனூடுக்களும் உள்ளார். தொடரும் திறனூடுக்களும் உள்ளார்.

1. நல்லவை திறனூடுக்கள் உள்ளார், திறனூடுக்களும் உள்ளார்.
2. நல்லவை திறனூடுக்கள் உள்ளார், திறனூடுக்களும் உள்ளார்.
3. நல்லவை திறனூடுக்கள் உள்ளார், திறனூடுக்களும் உள்ளார்.
4. நல்லவை திறனூடுக்கள் உள்ளார், திறனூடுக்களும் உள்ளார்.
5. நல்லவை திறனூடுக்கள் உள்ளார், திறனூடுக்களும் உள்ளார்.

மாத்து தொடரும் திறனூடுக்கள் கடுமையான செயல்பாடு மாட்டும் திறனூடுக்கள் உள்ளார். நல்லவை திறனூடுக்கள் உள்ளார். நல்லவை திறனூடுக்களும் உள்ளார்.
Alarm numbers, Settings, Profile, General, Silent, Tones, Ringing tone, Ringing volume, Vibrating alert, Clock,
Messages, Inbox, Outbox, Save, Send, Delete, Contacts, Search, Call, Missed Calls, Received Calls, Dialled
numbers, Settings, Profile, General, Silent, Tones, Ringing tone, Ringing volume, Vibrating alert, Clock,
Alarm numbers, Settings, Profile, General, Silent, Tones, Ringing tone, Ringing volume, Vibrating alert, Clock,
Alarm numbers, Settings, Profile, General, Silent, Tones, Ringing tone, Ringing volume, Vibrating alert, Clock,
Alarm numbers, Settings, Profile, General, Silent, Tones, Ringing tone, Ringing volume, Vibrating alert, Clock,
End notes

1 Information Note to the Press (Press Release No. 15/2010) by Telecom Regulatory Authority Of India
2 Information Note to the Press (Press Release No. 15/2010) – Annexure-I by Telecom Regulatory Authority Of India
<table>
<thead>
<tr>
<th>Category</th>
<th>Function</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery Reports</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instant Messages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discreet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loud</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outdoor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ringing tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ringing volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vibrating alert</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Msg.alert tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date and time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connectivity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call divert</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anykey answer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic redial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed dialing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call waiting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Send my caller ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security keyguard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Welcome note</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network selection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start-up tone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gallery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Images</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video clips</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Music files</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Abstract

Our main objective in this part of the project has been to extend the distribution of Tamil language learning materials and communication to lighter equipment, specifically PDA and mobile phone. The challenge is then to develop the system and server side to present materials in ways suitable for PDA technology, find acceptable solutions for distribution of materials and for administration to student, teacher to student/student to teacher to student communication.

It is our aim in designing the environment for the mobile learner to extend and increase the flexibility of Tamil language education, that to some extent took a step backwards when converting from paper based to online learning, where students largely were required to study at a place (and at a time) where a computer with access the Internet was available.

Introduction

This paper examines what kinds of software architecture might be used to build Mlearning systems and outlines what factors and issues should be considered in terms of the benefits and drawbacks of each generic architecture. In the following section we review some relevant literature on key aspects of M-learning systems. We then outline a number of software architectures that may be used to build M-learning applications, looking at how each approach may contribute to the requirements of M-learning while considering the practical challenges and limitations. We provide a summary of the key issues associated with each architecture and suggest some recommendations for software architectures appropriate to different types of mobile learning system.

Non Adaptive Mark-Up

A number of mobile learning applications have been developed that use some specific form of browser mark-up for their client side presentation. This mark-up may be Wireless Markup Language (WML) or variations on the HyperText Markup Language (HTML) such as cHTML (compact HTML), XHTML (eXtensible HTML) Basic or XHTML Mobile Profile, depending on which types of mobile device are being supported. Regardless of the mark-up, the content may be served as static pages or generated dynamically on the server, using technologies such as server pages and/or eXtensible Stylesheet Language Transformations (XSLT) (Figure 2). Fowler classifies these to approaches to dynamic mark-up as the template (server page) and the transform (XSLT) approach, though in fact it is possible to combine the two, for example by using Tag libraries (such as the JSP standard tag library, the JSTL) in server pages that support transformations.
The advantage of this approach is that it is lightweight from the client device perspective requiring only the device’s normal browser. The problem with non-adaptive mark-up is that using a particular mark-up language, for example WML, means that the content can only be rendered by browsers that understand that particular type of mark-up. Even though the content may be dynamically generated, it is only being generated for a specific type of client.

**Adaptive Mark-Up**

Adaptive mark-up requires a server side process that is able to generate mark-up appropriate to the mobile device from a common set of contents. There are a number of approaches to this, for example an application can interrogate the HTTP (HyperText Transfer Protocol) header of the request and identify the client browser type from the ‘user-agent’ field, then generate client specific mark-up using various XSL transformations. An alternative approach is to use a tag library such as Wireless Abstraction Library (WALL), a JSP tag library that builds on the Wireless Universal Resource File (WURFL). WURFL is able to recognise user agent information and identify different devices, while WALL generates device specific mark-up. Either way, the advantage of this approach is that it enables an Mlearning application to support multiple types of mobile browser (Figure 3).
Mobile Client Side Application

While the mark-up approach to M-learning architectures can provide a good range of content across a wide range of devices, confining the learner’s activity to what can be supported by a mobile browser can limit the range of learning activities that are possible on the device. For example, interactive learning games may be difficult or impossible to support. An alternative approach to using mark-up to provide content via the mobile Internet is to provide applications that can be downloaded to the mobile device. There are three general categories of application that may be developed for mobile devices.

The first approach is an application written for a specific mobile device platform, for example targeting a particular model or make of mobile phone. These applications can take advantage of the special characteristics of that particular device, which can make them, for example, highly performant, but of course these applications cannot be used on other devices. A second approach is to use Microsoft Windows based applications, for example building an application using Windows Mobile components. This approach is more generic than writing for a particular device, since there are a number of devices that support Windows Mobile. However such applications cannot run on the majority of mobile devices, since there are many other operating systems being used including Palm OS, Symbian and Linux. The third, most generic approach, is to use Java Micro Edition (Java ME). Using this software platform enables us to deliver a relatively rich client experience to a wide range of devices. Although Java ME has its limitations compared to some other application platforms such as Symbian and BREW, it works across a large proportion of mobile devices, including many Windows phones. In the context of mobile phones, the specific configuration and profile typically installed is the Connected Limited Device Configuration (CLDC) supporting the Mobile Information Device Profile (MIDP).

Java ME applications that run using this profile are known as MIDlets. Of course coding at the higher level of abstraction that enables interoperability has its costs in performance terms. For example Java ME applications have been shown to execute at about half the speed of equivalent C programs. There are also issues regarding the version of both the CLDC and MIDlet specifications that a given phone may support. Important differences between versions include floating point number support and security management, among many others. If a mobile client application is chosen as the software architecture, then the overall system design is simple.

The application runs standalone on the client, so the only role of the server is to enable download and installation of the application (Figure 4). This may be done either wirelessly or using a cable. Wireless download of Java applications can be provided using OTA (over the air) provisioning, a standardised approach that provides a consistent client-server interaction and enables version control for application downloads. The mobile application may use the data store on the mobile device but will not require access to any server side resources.

Figure 4: Client side application architecture
Smart Client with Server Connectivity

So far we have considered two very different types of software architectures for mobile learning systems, one based on providing a page based mobile Internet system, the other using a downloadable application client. However there is another approach that combines features from both of these architectures, the smart client that connects to the server. In this approach, there is a client side application, but that application does not run standalone. Rather, it communicates with the server to send and receive information while the application is running (Figure 5). There are a number of advantages to this approach. First, it is possible for the mobile learning application to provide a much wider set of content than is possible with a single downloaded application, since the storage size of the mobile device limits the amount of learning content that can be downloaded. A smart client that connects to the server can access learning content on demand without having to keep it all stored in the mobile device. Similarly, we can utilise the server to store information about the clients so we can, for example, maintain a sophisticated user profile on the server. In addition, the ability of the device to communicate data with the server in both directions (upload and download) means that it is possible to build a collaborative learning system where multiple users can send and receive data to and from each other.

Figure 5: Smart client with server connectivity architecture

The price we must pay for this power and flexibility is complexity. For example, data management becomes a more complex issue because there may be data stored in a local data store on the client device as well as on the server. Mobile databases will need to be synchronized with central databases, and cache management has to be sophisticated to cope with the small memory size in most mobile devices. Distributed smart client applications using this architecture also need an appropriate communication protocol, such as XML over HTTP, to enable the client to access server side resources at run time.

Tamil Language Learning through Mobile

M-learning offers a powerful and practical solution to many learning and training challenges, such as:

- in collaborative projects and fieldwork
- as a classroom alternative to books or computers
- where learners are widely dispersed
• to engage with learners who in the past have felt excluded
• in promotional and awareness campaigns
• for ‘just-in-time’ employee training.

From a teaching and learning point of view, campus-wide internet access - or even access that targets social and learning spaces such as refectories, libraries, lecture rooms and labs - is what truly blends together online and face-to-face learning. It means that while they’re on campus, a student can access their online learning just by turning on their net book or iPhone. They can contribute to class online discussions while eating lunch or access their readings before class, using the technology they already have with them: their laptop, net book, or other wi-fi capable mobile device.

For mobile language learning - and even for flexible learning - at any educational institution, equipping formal and informal learning spaces (such as social spaces) with fundamental enabling technologies like wireless internet access has to be at the top of the priority list. It even makes sense from a budget point of view, as every laptop a student brings in and uses takes pressure off the student labs. This, in turn, reduces the amount that has to be spent on standard-image, admin-locked, physical lab computers… and frees students to use their own computers which can be configured to best support their particular program of study.

Conclusion

In this paper we have provided a brief overview of four software architectures for mobile learning; non adaptive mark-up, adaptive mark-up, mobile client side application and smart client with server connectivity. All of these architectures have their own strengths and weaknesses, and in most cases we are trading flexibility against complexity. In addition there are different levels of server connectivity required for these different architectures, and successful applications depend not only on the technical infrastructure but also the social context within which issues such as pricing come into play. Therefore deciding on suitable software architecture for a specific Mlearning system depends not only on technical factors but also an analysis of the user context. While we may see that smart client architecture with server connectivity can provide us with the richest mobile learning environment, alternative architectures may prove easier to install, more robust in use, more easily deployed to a larger range of devices and cheaper for the learner to maintain. Therefore the most important aspect of designing an M-learning system architecture is to consider all aspects of the user context rather than just focus on the technical platform.

References

The use of palmtop computers for learning A review of the literature – Learning and Skills development agency.
The use of Computer and Video games for learning – Alice Mitchell & carol Savill-Smith.
http://www.xenglobaltech.com/j2me-mobile-applications.html
http://learning.ericsson.net/mlearning2/project_one/wap_article.html
Software architectures for mobile learning D. PARSONS & H. RYU
Chang, C. and J. Sheu (2002). Design and implementation of ad hoc classroom and eschoolbag systems for ubiquitous learning. IEEE Int. Workshop Wireless and Mobile Technologies in Education.


Predictive Tamil Short Messages
for Handheld Devices

Dept of Computer Science & Engg, CEG, Anna University, Chennai – 25.
Email: abirami_mr@yahoo.com, vilva.mca@gmail.com, baaski@annauniv.edu

Abstract
This paper aims at providing ease and flexibility to Tamil SMS users by suggesting the commonly used message templates with few key strokes, based on the context of letter typed. This system intends to predict short messages in Tamil, as similar to T9 dictionary, but the scope is not restricted to the prediction of a single word. Rather this attempt to predict the short messages (sentences) communicated in our day-to-day life.

Initially, keypad of the mobile phone has been configured with Tamil Vowels and Consonants using Tamil keypad layout standardized for hand held devices. To start with, this predictive SMS system allows the user to type an initial letter in Tamil to predict the possible nouns and verbs starting with the letter typed. Based on the user selection, initial word gets composed. If the initial word appears to be a noun, probable actions (i.e. Verbs) associated with the noun with proper case endings could be predicted. Depending on the needs of the user, verbal actions could be enriched with suitable adverbs too. If the initial word appears to be an adverb, suitable verb could be appended with it.

Possible templates (nouns, verbs and adverbs) are stored in Unicode text format in a file (records). Depending on the context (taking nouns, case markers and adverbs into account), suitable records would be retrieved. Few English message templates are available in standard mobile phones to reduce the typing hindrance involved in messaging. Moreover, these templates are not applicable to Tamil language. As a result, this paper has been motivated to introduce a Predictive SMS system which intends to reduce the number of key presses required to frame a message in Tamil. In addition, this system covers most of the commonly used Tamil sentences and not restricted to fewer templates, thereby attempting to provide a full fledged predictive Tamil short message service.

Introduction
Now-a-days, SMS through mobile phones have become an integral part of human communication. To make it easier and innovative, it is important to have the facility to enter, send, receive and read SMS’s in Tamil. SMS in Tamil are already a reality. This paper is therefore, we believe, the first attempt to come out with a possibility to send SMS’s in Tamil, as comfortably as one sends an SMS in English. To add flavor to it, we try to make it more easily than English.
**Tamil language**

Tamil is a South Indian language spoken widely in Tamil Nadu in India. Tamil has the longest unbroken literary tradition amongst the Dravidian languages. The earliest available text is the Tolkāppiyam, a work describing the language of the classical period. There are several other famous works in Tamil like Kambar Ramayānam and Silapathigaram but few supports in Tamil which speaks about the greatness of the language. For example the Thirukural is translated into most other languages due to its richness in content. It is a collection of two sentence poems efficiently conveying and few other things in a hidden language called Slaydai in Tamil. Tamil has 12 vowels and 18 consonants. These are combined with each other to yield 216 composite characters and 1 special character (aayatha ezhuthu) counting to a total of \((12+18+216+1) = 247\) characters.

**Vowels**

Vowels in Tamil are otherwise called UyirEzhuthu and are of two types short (Kuril) and long(Nedil).

**Consonants**

Consonants are classified into three classes with 6 in each class and are called Vāllinam, Idaiyinam, and Mēllinam.

**Tamil Unicode**

The Unicode Standard (http://www.unicode.org) is the Universal Character encoding scheme for written characters and text. It defines the uniform way of encoding multilingual text that enables the exchange of text data internationally and creates the foundation of global software. The Tamil Unicode range is U+0B80 to U+0BFF. The Unicode characters are comprised of 2 bytes in nature. For example, the Unicode for the character \(\text{இ}\) is 0B85; the Unicode for the character \(\text{ல}\) is 0BAE+0BC0. The Unicode is designed for various other Tamil characters.

**Functional Block Diagram**

[Diagram with steps: Press Mobile Keys → Map the Key Code to Unicode Text → Display the Tamil Letter & Predict Relative Data → Select the Needed Data & Continue]

**Taking Key Codes As Input**

When the user presses the keys of the mobile phone, the key press event generates unique key codes for each key. Using those keys codes and the time interval between key press events, the appropriate Tamil letter is mapped and displayed on the screen.

**Keyboard Mapping**

Mobile Phone Keyboards have a limited number of keys (roughly 12). However, the use of the mobile phone as a messaging instrument (using SMS) has risen in the recent past. Currently, predictive texting schemes exist for English and many European and other languages. However, such a scheme does not exist for most Indian languages, including Tamil.

Moreover, the mobile phone keyboard for the English Language is generally arranged in the alphabetical order, that is the key for the number 2 contains the letters A,B,C; that for number 3 contains D, E, F; and so on. Following a similar arrangement is a relatively inefficient
scheme for Tamil; hence, we have come out with a more efficient arrangement that does not club frequently used letters on the same key. Our mapping assigns, in general, two vowel and two consonants to each key.

While pressing a key, the first letter assigned to that key is mapped. If the same key is pressed again, the time between the subsequent key strokes are calculated. If the time is below the quantum time assigned, instead of the first letter, the second letter assigned to that key is mapped. The same concept is applied for all the letters assigned to the keys.

When a consonant letter is pressed and subsequently if a vowel is keyed in, its corresponding consonant letter is mapped. For example, ma+e=mi.

**Stored Records**

The Possible Sentences used often are stored in the records, as far now, it is limited so as to make the work easier. The sentences are stored in Unicode format, thus making the size of records minimal.

**Predict**

As the keys are pressed and the corresponding Tamil characters are displayed in the screen, the system searches for the matching sentences in the Records and if any matches, they are displayed as choices. When the first letter is typed, the system searches for sentences starting with that letter in the record. And displays the possible sentences, the user can select one from that choice or can go for a different choice of typing in the second letter. Now the records are searched for sentences starting with that two letters combined. And the possible sentences are displayed. Now the user can select one from that choice or can go on for typing in the next letter. The same procedure follows for every letter typed.

**Future work**

This paper limits the sentences in Records to a meaningful limit, we can increase the records in future for a better system. The user interface is also restricted to our convenience, it can be changed to a standard format.

**Conclusion**

This system encompasses the Tamil typing and prediction of the sentences as and when typing. This eliminates the typing and time to a considerable extent. Also it helps a person with less knowledge in Tamil typing and spelling.
16

dவிளை உண்மை
Status of UNICODE
in The Indian Tamil Publishing Industry

P.Chellappan
Partner, Palaniappa Bros
chellappan@vsnl.com

The Publishing industry in general including the Tamil Publishing industry is at a crucial juncture today. The world of publishing is moving away slowly but surely from the traditional paper based publication towards e-book publication. This movement is more significant in English books than Tamil books. Nevertheless the Tamil publishing industry has to gear itself as soon as possible, so as not to fall behind and get lost in the technological revolution. The Tamil Publishing industry has traditionally been using 8 bit encodings just like all the other Indian languages. But looking at the future, it has to move over to 16 bit encodings.

The purpose of this study is to ascertain the current practices in the Tamil publishing industry so that suitable steps can be taken to help in this transition.

Methodology

A questionnaire was prepared and distributed to some of the major Tamil books, and newspaper publishers. The questionnaire mainly relates to the pre-press activities such as receipt of manuscripts, data entry, page layout, and graphic designing. The feedback was then analyzed.

Summary of analysis

In all 30 Book Publishers and 10 Magazine & Newspaper publishers were asked for their feedback. But only 11 responses were received from Book Publishers and 9 were received from Magazine & Newspaper publishers. It can be safely assumed that there will not be a significant difference in the findings even if the remaining feedback was received.

The data was analyzed and statistics are presented in the tables given below. In these tables (B P) refers to Book Publishers and (M & N) refers to Magazines and Newspapers.

<table>
<thead>
<tr>
<th>Receipt of Manuscripts</th>
<th>Hard Copy</th>
<th>Soft Copy</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>84.00%</td>
<td>16.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>4.00%</td>
<td>96.00%</td>
</tr>
</tbody>
</table>
It can clearly be seen that in the BP segment manuscripts are received mostly as either hand written or typed documents. But in the M&N segment it is quite the opposite. 96% of the manuscripts are received as soft copies. These soft copies are again received in various font encodings. Fortunately, in most cases a single manuscript comes in a single encoding. So encoding conversion operations are reduced to a minimum, unless of course it is absolutely essential like in the case of manuscripts received in Unicode.

**Type Setting:**

<table>
<thead>
<tr>
<th></th>
<th>In-house</th>
<th>Outsourced</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>23.00%</td>
<td>77.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TAM</th>
<th>TSCII</th>
<th>UNICODE</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>44.00%</td>
<td>12.00%</td>
<td>19.00%</td>
<td>25.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>67.00%</td>
<td>-</td>
<td>-</td>
<td>33.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Page Maker</th>
<th>Indesign</th>
<th>Quark</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>57.00%</td>
<td>10.00%</td>
<td>25.00%</td>
<td>25.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>10.00%</td>
<td>20.00%</td>
<td>10.00%</td>
<td>60.00%</td>
</tr>
</tbody>
</table>

Type Setting operation is one of the first major operations done by any publisher. Most of them have only limited in-house capability and the rest is outsourced. As can be seen in the tables above, the BP segment has 23% in-house capability but the M&N segment has 100% in-house capability. The major advantage of having in-house capability is in the tight control over font usage. In the case of outsourced operations, the publishers have little control over the font encoding used. Some times they are even forced to accept documents that use different font encoding since the DTP operator has only that option.
Some of the publishers are either ignorant or do not even bother about the font encodings, as they archive the books only as hard copies.

**Page Layout:**

<table>
<thead>
<tr>
<th></th>
<th>In-house</th>
<th>Outsourced</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>27.00%</td>
<td>73.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TAM</th>
<th>TSCII</th>
<th>UNICODE</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>50.00%</td>
<td>14.00%</td>
<td>-</td>
<td>36.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>67.00%</td>
<td>-</td>
<td>-</td>
<td>33.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Page Maker</th>
<th>Indesign</th>
<th>Quark</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>52.00%</td>
<td>23.00%</td>
<td>10.00%</td>
<td>15.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>11.00%</td>
<td>78.00%</td>
<td>11.00%</td>
<td>-</td>
</tr>
</tbody>
</table>

Page Layout is an important aspect in the publication of any book, magazine or newspaper. It can be seen that in the BP segment 73% of this operation is outsourced. The reason is that they would like to leave it to the professionals. However in the M&N segment 100% of the page layout is done in-house. Newspapers and magazines have very little time on hand before the matter goes to print. Hence it makes sense to do this important job in-house.

One important aspect here is that UNICODE is not used in this final stage. This is attributable to the applications used for page layout. Page Maker, Indesign and Quark do not support complex script and hence cannot be used for Tamil, or for that matter any Indian language.

**Graphic Design:**

<table>
<thead>
<tr>
<th></th>
<th>In-house</th>
<th>Outsourced</th>
</tr>
</thead>
<tbody>
<tr>
<td>B P</td>
<td>34.00%</td>
<td>66.00%</td>
</tr>
<tr>
<td>M &amp; N</td>
<td>100.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
All the observations for Page Layout are exactly applicable for Graphic Design also. Here also the applications used viz. Photoshop, Illustrator and CorelDraw do not support complex script rendering. Hence Unicode encoded text cannot be used in this step.

**Conclusion**

The survey clearly indicates that the entire publishing industry uses only 8 bit encoded fonts like TAM and TSCII, with TAM being used more widely than the other encodings.

It also indicates that even when the original manuscript is received in Unicode, it is being converted into one of the 8 bit encodings for further use. This is mainly due to lack of support of complex script rendering in the applications that are used in the publishing industry.

The lack of awareness of UNICODE amongst the book publishers is also another contributing factor for non-usage of UNICODE. The reason for this lack of awareness can also be attributed to the non-usability of this encoding.

However all the publishers in the M&N segment are aware of UNICODE and in fact about 55% of the publishers in the M&N segment use UNICODE for their on-line editions. They refrain from using UNICODE in the print editions only due to the lack of support for complex script rendering in the applications that are used by them.

In the coming days when use of multiple languages in a single publication becomes more common, it is better to migrate from the legacy 8 bit encodings to 16 bit encodings. The inhibiting factor as already stated above is the lack of support for complex script rendering. Since a definite time frame is not available for provision of this support, usage of an all character 16 bit encoding like TACE16 will help in immediate migration to the 16 bit environment. One additional advantage of TACE is that even off-the-shelf e-book readers will be able to render Tamil e-books embedded with a TACE font.
Problems of using Unicode in software components and in NLP

Dr V. Krishnamoorthy
(Former professor Of Anna University)
5, Srivatsa Apartments
A1-11, 23 rd Cross Street
Besant nagar, Chennai 600 090
Email: prof.vkrish@gmail.com

Abstract

To speed up software development and reduce cost, using specialised software components is an established practice. Tamil editing software developers have been using a component called “text control”, to include functionalities like justification and tables. These components do not support complex scripts. Hence they are not usable for using Unicode Tamil in the software. It is not known when such support will be provided. Economic considerations may indicate that it may not come through in the immediate future.

The processing of letters, words, sentences, paragraphs and passages to ‘understand’ what they mean, and, produce answers to the required questions is the basic aim of Natural Language Processing (NLP). This has string processing as the core. The encoding of Tamil in Unicode is not very conducive to this string processing is shown by taking a few simple fundamental operations. The complex behavior is the result of the variable length encoding and unnatural way of representation of letters in Unicode.

1. Introduction

It is now widely known that Unicode Tamil is not supported in high-end publishing software. But less known is the fact that some software components also do not support Unicode Tamil. This may pose restrictions on developing specialised Tamil software in the long run. Here we highlight this problem by giving a specific example. We find that the reason for not supporting Unicode Tamil is that Tamil in Unicode needs level 2 implementation.

There are many special editing software for Tamil, which provide many other special features not available in Word etc. All these software use software components called “Text control”. This text control component is essential to provide the functionalities like, justification and tables. It is not possible to create such components on our own, for our own use, in view of the complexity involved. Tamil software developers, like many other developers world wide, use these components, to provide rich functionalities within their software.

When Unicode is used in these components, the following problems arise. Complex script support is not available in text controls, which are available today. Absence of this leads to incorrect alignment and undesirable and unacceptable editing experience. As such these components cannot be used for Tamil
Unicode. The component developers are not willing to specify when they will provide support for complex scripts. It is seen that these components work with the Tamil All Character Encoding (TACE).

In the second part, the differences in dealing with the current Unicode and the all character encoding, in the context of natural language processing is brought out. Specifically how the fundamental operations are done easily and quickly in the all character encoding is studied. This shows that keeping the data in the all character encoding will speed up the NLP in Tamil.

2. Using Tamil Unicode in components

Software components called Text Controls are used in many Bilingual text editors. Currently they do not provide full support for Tamil. They open the file with Tamil Unicode characters. But the visible cursor position does not correspond to the “actual” cursor position. If one tries to insert a letter or string at a particular position, especially near the end of a line, it gets inserted somewhere else. An example is shown in Figure 1.

Here the cursor is placed at the end of the third word in the first line, which is ‘aduththum’, and the letter ‘vee’ is inserted using the ‘paste’ operation. The letter gets inserted in the middle of the third word.

It seems that this is due to the wrong calculation of the width of the letters. Here it seems that the letters are treated as having constant width. This can be seen by keeping the cursor at the left and tapping the right arrow key repeatedly. Also if ‘select all’ is given, the selected portion is shown wrongly. See Figure 2 for the left aligned text. The mismatch is clearly visible. Similar mismatch is seen in the case of the justified text also.

It is obvious that no commercial software can be sold with such grave shortcomings.
Figure 2

Figure 3 shows that the text and the selected area match correctly in the case of text in all character encoding. The widths are correctly calculated and the insertion is done correctly.

Figure 3
When contacted, the Text Control company has not given any definite answer about providing support to Tamil. Software components are not sold like mass products. Only software developers buy these. So the number of buyers will not be very large. When it comes to a particular language, this number may be very small. This may discourage many of these component vendors, who are not big companies, from providing support for a complex script, since it may involve considerable cost.

This indicates that no Tamil application using Tamil Unicode can make use of many good components in the near future. Whereas most of them may work without any difficulty with the all character encoding.

3. Problems in NLP

Comparing two strings is a fundamental operation, used heavily in NLP. For example, when a new word has to be added to an existing dictionary, the new word has to be put in the proper place. Otherwise it will lead to linear search, which is time consuming. When we compare two strings, it is natural to expect that the natural ordering of letters in Tamil be followed. This is essential, as the sorted sequence of words may be required to be printed for the human perusal, or stored for use in some other application.

Comparing two Tamil strings in Unicode is a formidable task, since the ordering of characters in Unicode does not follow the Tamil ordering. Also, the letters ksha and sri, which we consider as single letters, are stored as a combination of two or more characters. In the all character encoding, this task is just as easy as saying 1,2,3. The pseudo code for this can be given below.

Let minLength = minimum ( length(string1), length(string2) )

For I = 0 to minLength-1
    If (string1[I] > string2[I]) then string1 is bigger. Exit
    // here the common part is same.
    If ( length(string1) > length(string2) )
        then string1 is bigger.
    else string2 is bigger.

Exit.

Writing the pseudo code in the case of Unicode is not so easy. I do not attempt it here, due to lack of patience on my part. But, note that the following things have to be taken care of while comparing.

The consonants with ‘a’, have to be rearranged according to Tamil sorting order.

The mei pulli should come before any of the vowel mathras.

A consonant with ‘a’, with the pulli coming next to it, precedes that consonant with ‘a’.

Ksha series letters should not be put in the ka series, though the first character is ka.

Sri has to be put in a separate place.

The length of the string is not the same as the number of letters. String length cannot be used straightaway.
One may understand the complexity and the time involved in writing this code. The time involved in executing these two codes is obvious.

Let us take another simple example.

Hyphenation is a problem faced in the printing industry. Let us consider a simple solution for this, with the following rules. Break only when there are 4 or more letters. At least 2 letters should on both sides of the cutting point. The second part should not start with a pure consonant. (This solution has been implemented by us already.)

In the case of all character encoding, the pseudo code would look like this.

If ( length ( word ) < 4 ) exit.

// length is 4 or more, here

// try to accommodate as much as possible in the first line

for I = length(word) -2 to 2 do

    if word[I+1] is not a consonant, and, first I letters fit in the first line,

        then split after the I-th letter. Exit.

Considering the same solution in Unicode, the following things need attention.

The length of the string does not give the number of letters. Hence the condition that at least two letters on both sides has to be taken care of separately. The boundary points of the ‘for loop’ does not ensure this.

The cutting may happen at the middle of the letter. So every time it should be assured that we do not cut at the middle of a letter. Special attention is needed for sri and ksha series letters.

All these will need more coding and more execution time.

We will close with one more example.

Searching for a string as it is, has limited scope in Tamil. In Tamil, words appear with many variations. For example the word ‘maram’ appears as ‘maraththil’, ‘maramaaka’ etc. To find out all the places in which ‘maram’ appears, it has to locate all the words with variations also. A simple solution will be to first search for the string ‘mara’, and then find out whether the word found is a required word. For this start checking whether the next letter is ‘im’ or ‘th’. The same logic works for both Unicode and TACE. But the difference is the number of false hits. In Unicode, when ‘mara’ is searched, all the words with ‘mar’, ‘mara’, ‘maraa’, mari’, … will be the results of the search. This will result in a large set which has to be filtered for the required words. But in the case of TACE, only words having ‘mara’ will be found out. Others will not appear in the searched words. This will save the time for processing the unwanted words to a great extent.
In the three simple examples provided, we have shown that in the case of Unicode, the coding may be complicated, and even when the coding may be the same, some unnecessary work may be done. In view of the above, we can conclude that the processing for NLP will be easier and faster, and the running time will be less when TACE is used, when compared with Unicode. Also the internal data, like dictionary, kept in TACE will make the processing easier and quicker.

Conclusion

It is shown that some useful software components, which are being used by Tamil editing software developers, do not support Tamil Unicode. The complexity of executing fundamental operations of NLP using Unicode is brought out. These are due to the level 2 implementation of Tamil in Unicode, and, the variable and unnatural way the encoding is done. It is shown that the all character encoding provides a simple natural way of encoding, and hence provides better results.

References

1. Unicode.org
Challenges of Publishing Industry and E-Governance with Tamil Unicode (TU) and possible remedies with Tamil All Character Encoding in 16-bit (TACE16)

A. Elangovan
email: elango_ ae@yahoo.co.uk, Chair-Infitt-WG08 & Founder President, Kani Thamizh Sangam MD-Cadgraf Digitals & Digiscape Gallery, Chennai, India.

Background - 21 years experience in Publishing and Printing, development of Multi-lingual Editorial Work flow system, Indian language fonts and interfaces for Win & Mac, Adobe plug-in development, Tamil hyphenation & spell checking

Introduction
India is emerging as the Publishing hub of the world with 45% of publications in English, the world’s third largest English publishing country after UK and USA. The remaining 55% is published in Indian languages. Today Print media is still the largest media in terms of readership, circulation, published pages and editorial manpower involved. Today publishing does not stop with print media but also extends to web, mobile, palm readers and broadcast media.

E-Governance in India is in its nascent stage. Indian language enabling is essential for the successful implementation of any E-Governance project. E-Governance will include all media including web, print, mobile and palm readers. The data should be viewable, portable and reliable across various media, operating systems and applications.

Multilingual publishing is still a major challenge for publishers in spite of the Unicode advantage of having all languages of the world under one unified encoding scheme.

Today most of the content available in the web is in Unicode and Tamil Unicode is widely used by all segments of people. Inspite of its popularity, the Publishing Industry and E-Governance segment face several challenges in implementing Tamil Unicode. This paper explains the specific areas of challenges and the possible remedial measures to some of them. It also forewarns the users about the pitfalls in the implementation of Tamil Unicode and how to avoid them.

The paper discusses some of the implementation issues faced in real life projects with Tamil Unicode and the remedial measures taken. The paper also covers some possible solutions which include Innovative methods of Unicode Font design, TACE16 based applications and Software Plug-in Tools and its benefits.
Challenges in Publishing with Tamil Unicode

1. Text Editing

Today the editors have a wide choice of sources for the news, stories, events and statistical data. The sources include wire services, news agencies, reporters, journalists including freelancers, web sites, blogs, TV channels, digital archives and printed archives. The text data exchange mostly happens in the digital form through email and in other formats like pure text, word, spread sheet and PDF. More and more content in Tamil Unicode (TU) is available in the web, email and word formats and is becoming a popular source of information. However the text need to be converted to 8-bit legacy encoding like TAB/TAM before it can be paginated as most of the publishing applications like Adobe, Quark and Corel do not support TU yet. Converters are widely available to convert TU in pure text format and the availability for other formats like word is limited. We face two problems while doing this conversion. First, if the TU text contains more than one language ie. Tamil and English, the text in English gets converted to junk. This requires manual intervention to identify the English text, apply the English font, proof read the text and correct all junk characters before they are sent to pagination. The second one being loss of all formatting done in other applications like word while doing this conversion. The typical text process flow is given in (Fig.1).

Solutions: One way to overcome the problem of multi-lingual text is to convert Unicode text to TACE16 (Tamil All Character Encoding for 16 bit) instead of 8-bit TAB/TAM. TACE16 can easily co-exist with other language text in Unicode. It is possible to retain the formatting features of the source document with help of plug-ins for conversion from within the respective applications. Eg. Indesign Encoding Converter plug-in.

2. Pagination

MS Word and Publisher are used to some extent in the entry level and do support Tamil Unicode with some limitations. PageMaker has been discontinued by Adobe few years back and anyway it does not support 16-bit encoding like Unicode. The most widely used pagination applications by
professional publishers like Newspapers, Magazines, Books, Directories, Government publications like budget documents, gazette publications, assembly documents are Adobe Indesign, Quark Xpress and Corel Draw. Unfortunately all these professional applications which are vital for publishing and printing are yet to provide TU support even in their latest versions (Fig. 2 A). And the definite date of their future support is also not yet known. Also about 30% of the publishing industry uses Apple Macintosh systems, where the TU support is incomplete.

Solutions: Interestingly TACE16, which is the standard recommended by TN Task Force is found to be working well in all the above professional applications (Fig 2 B). It is also supported both in Windows and MacOSX operating systems. Real life trials in production have proved not only its support but also its higher efficiency.

3. Illustrations

Graphic illustrations like logos, graphs, charts, cartoons, sketches, back grounds, banners and book illustrations form important elements of publishing for not only print but also for web and mobile. The popular applications used for creating graphic illustrations are Adobe Illustrator and Corel Draw. Unfortunately both of them are yet to provide TU support. (Fig. 3 A and 3 B)

Solutions: Fortunately TACE16 is found to be working well in both these applications in both Windows and MacOSX. (Fig. 3 C)
4. Image Editing

Photo editing, painting, colour correction, titling and labeling images are the next important functions of publishing. The leading application for image editing for print, web and animation is Adobe Photoshop. Corel Photo Paint is also used to some extent. Both these packages are yet to provide TU support.

**Solutions:** Once again TACE16 is found to be working well in both these applications in both Windows and MacOSX.

5. PDF Data Storage

Adobe PDF, the most popular Portable Document Format is almost the de-facto standard in print and publishing for creating the print ready format. PDF provides the portability of printing the documents across any printing device irrespective of the application and operating system in which it was originally created. It provides option for embedding the original fonts making it possible to print in any device without the need for having the original fonts at the printing end. PDF made it possible to electronically transfer print ready document across the world, paving the way for phenomenal growth of the growth of the e-Publishing (Electronic Publishing) industry. PDF is also the most popular format for storage of documents both for short term and long term. PDF retains the original text and graphics, which can be extracted and edited for future use. ISO standards for PDF are well defined and extensively used by the print and publishing industry while defining specifications and printing standards. Adobe PDF does not provide support for the complex rendering process in Tamil Unicode. The Unicode text is stored inside PDF in its own native format, different from that of the Unicode rendering order. When the Tamil Unicode text is extracted from the stored PDF documents, and placed back into MS Word, some characters with Egara, Eegera, Ugara, Uugara and Augara modifiers get reordered and sometimes even joins with the previous consonants, making the text meaningless (Fig. 4 A).

**Figure 4 A - TU PDF Extracted in Word**

**Figure 4 A - TACE16 PDF Extracted in Word**

Hence Tamil Unicode text stored in PDF standard is unreliable. The data integrity is totally lost in a round trip PDF storage and retrieval cycle even from MS Word. This is a major draw back for not
only print and publishing but also for all office and commercial applications where PDF is used as data storage and exchange format.

**Solutions** : Unlike Tamil Unicode, TACE16 stores all data as *Simple Script* without the use of the modifiers. All characters are stored as a single unit and there is no need for the OS to do the Glyph reordering process. Tamil text in TACE16 when stored in PDF, the internal storage order of PDF matches with the TACE16 ordering. On extraction of TACE16 from stored PDF documents, the data integrity is strictly maintained. TACE16 is found to be very reliable for round trip PDF storage and retrieval cycle. Production trials in real time publishing environment also proved its reliability with all applications like MS Word, Adobe, Quark and Corel in both Windows and MacOSX. (Fig 4 B)

6. Tamil Spell Checking

Automatic spell checking is a tool used extensively by not only the print and publishing but also by all office and commercial applications. Spell Checkers indicate not only spelling errors but also sandhi errors which is common in Tamil. Some of the text control tools used by the developers of Tamil Spell checkers do not support TU. Dual representation of certain characters in TU creates problems to the spell checking tools in identifying the words and its component parts. Most of the Spell checking engines use their own internal storage format for processing the text. This creates an additional burden in converting the Unicode text to internal encoding and after processing reconverting to Unicode format. This may affect the efficiency of the spell checking process to a great extent.

**Solutions** : TACE16 with all Tamil characters encoded with one to one mapping and without any complex rendering process is supported in all text control engines. In TACE16, there is no dual representation of any Tamil character. Hence TACE16 can be used as the internal encoding for processing also. The vowels and consonants are easily identified from their code values, which enhances the text processing speed considerably. Tamil Spell checking tools based on TACE16 not only work well with all current applications but also found to be far more efficient as compared to tools based on Tamil Unicode. Processing speed is very critical in production environments like newspapers.

7. Tamil Hyphenation and Justification:

Hyphenation and justification for Tamil text is far more complex than English. For English as the Hyphenation dictionaries are generally built into most of the professional applications, the hyphenation process is simple and automatic. For Tamil, there are no standard hyphenation dictionaries available. In a multi-column newspaper or magazine publication, hyphenation and justification is very critical. Improper word breaking at the end of each line will give very awkward look. Normally it is essential to do the manual hyphenation break at appropriate places. In real life production environment, reformatting and repagination are quite common. At each reformatting stage, the hyphenation breaking needs to be redone. All the earlier hyphenation breaks need to be removed as word breaks in middle of the lines is not allowed. With TU some of the complex characters can break in between the characters resulting in illegal breaks.

**Solutions** : Auto hyphenation for Tamil is normally achieved by plug-ins to applications like Indesign, etc; A series of hyphenation rules are built into the plug-ins, which automates this tedious
manual process. While applying these hyphenation rules, it is important to ensure that illegal breaks do not happen in between complex characters. With Tamil Unicode, the bit length varies from character to character and hence this process is little more complex. Whereas with TACE16, all characters are having uniform bit length, the process is much simpler. Also the possibility of illegal character break is totally eliminated in TACE16, as all characters are represented by a single code point.

8. Challenges in E-Governance with Tamil Unicode

Most of the e-governance projects require people involvement and interaction at grass-root level. Local language enabling in an efficient method is essential for the successful implementation of e-governance projects. An end to end e-governance project will involve multiple media which include web, print, mobile, palm readers, kiosks and e-books. The data should travel reliably across multiple operating systems and databases. The documents should be viewable, portable, text extractable for local processing and printable as and when required. The system should provide for reliable and efficient long term storage of data, which are independent of operating systems or applications. The system should provide a reliable and easy to use search and retrieval of the documents.

Tamil Unicode implementation in E-Governance is faced with a series of challenges. TU support is lacking in several media like print, mobile, ebooks, palm readers and broadcast. Wherever the documents need to be published in printed form, the documents need to be converted back to legacy 8-bit encoding. This increases the risk of errors and necessitates manual proof reading and correction. Integrating mobile, which is the most widely used device in the grass-root level, poses several limitations with TU due to lack of support for complex rendering in all devices uniformly.

Also reliable data portability across operating systems and applications is doubtful, as the TU support is incomplete in several applications. Dual representation of certain complex characters poses a problems to database sorting and indexing, resulting in inefficient search and retrieval of documents. Long term storage of portable documents with TU in PDF format is unreliable, as data integrity on retrieval can’t be ensured. Also the stored data in text and other formats can’t be retrieved without the help of the respective rendering engine (like Uniscribe in MS) provided by the respective operating system. This is a threat to data security in case of discontinuity of support for particular rendering engine by the OS provider after some years.

Solutions: TACE16 was tested by the National Informatic Center(NIC), GoI, as part of the testing process of TN Task Force on All Character Encoding in many of the E-Governance applications and confirmed its usability. Hence TACE16 could be used safely in all E-Governance applications where Unicode support is lacking or doubtful.

9. Font design for Multilingual Publishing

Unicode provides code points for all languages of the world including Indian languages. This give the provision for keeping multiple language scripts in a single font. (eg.Arial Unicode). One can design a font with both TU and TACE16 in the same font, which will give many operational advantages where both are required. This will help the user to read the TU text and convert to TACE16 when necessary for printing without the need for changing the font. Publishers of multilingual dictionary, commercial applications which require multi-lingual user interface and
government departments which need to create documents in multiple languages can create a font with the required language scripts to suit their special needs. This is one of the major advantages of Unicode as compared to legacy 8-bit encoding.

10. E-Paper Publishing

In today’s condition Newspapers and magazine publishers who are using professional publishing applications will be forced to use TACE16 until TU support is available. Publishers who are publishing E-Paper version of their publication in the web (Electronic Paper in Newspaper look and feel) can use TACE16 encoding itself without the need to convert to TU. This gives the advantage of better access speed to the web readers. They also provide archived older newspapers for their readers. The extend of old archived E-papers one can provide depends on the online storage space provided in the Web Server. With TACE16, one can provide much longer period of newspaper archives online with available space.

11. Mobile Publishing:

Today’s newspapers do not stop with print medium, but extend to mobile for quicker delivery of hot news and headlines. Tamil Unicode support in mobile devices across brands and models is still a distant reality. This is due to the difficulty in accommodating the overload of complex rendering necessary for TU. TACE16 being a simple script without any complex rendering, can be used easily and efficiently in mobile publishing.

12. E-Book Publishing

Book publishers may be using TACE16 due to lack of TU support in publishing applications. Just like in mobile devices, E-Book readers also lack the capability to handle the complex rendering required for TU. Here again TACE16 has proved its usefulness in field testing with few manufacturers.

13. Broadcast Media

Many media houses today are planning to integrate their newsroom for print with broadcast studios in order to provide cross media news coverage. They use special purpose titling and editing tools, which are yet to provide TU support. In these situations, TACE16 will be useful, as all these broadcast applications support simple scripts like TACE16.

Conclusion

Publishing and E-Governance, though faced with several challenges in implementing Tamil Unicode, they can be overcome to a great extent by use of the TACE16 encoding wherever TU support is lacking and by using suitable plug-ins for encoding conversion, spell checking and hyphenation tools.
A study on Tamil Script in digital media

N. Anbarasan
Chief Executive Officer
APPLESOFT
#39, 1st Floor, 1st Cross, 1st Main
Shivanagar, W. C. Road, Bangalore - 560010, INDIA
Tel : 23386167, Telefax : 23357167, Mobile : 9448053137
e-mail - n.anbarasan@gmail.com, aplesoft@yahoo.com

Synopsis
Tamil script is one of the earliest recognized script amongst Indian scripts to get implemented on Computers, Pagers, Mobile phones, Dot Matrix Printers, Display boards etc and has been adopted for various applications like Desk Top Publishing, Messages, Web pages, Teaching/learning software, Billing software, Video sub-titling, News readers etc. This paper presents the possible basic requirement, which enabled implementation of Tamil script in digital media.

Whenever the ever growing technology has posed limitation in adopting Tamil script, its simplicity enabled implementation into such newer devices. But, in certain cases the number of glyphs required to implement Tamil script is definitely posing difficulty. This author analyses the difficulties posed by the Tamil script in implementing Tamil script on power hungry digital devices.

In order to implement Tamil script on digital devices, standards are required for input, storage and display. Eventhough, there are certain standards made available for Tamil script by concerned bodies, this author experienced contrasting nature of the standards prescribed for Tamil script. This author presents the experience gained while implementing the various standards as developer and as an user.

As a way out of possible seamless implementation of Tamil script and to have feedback for users while inputting Tamil text, the author suggests the possible script reform, the encoding standard and an input method. This paper also presents the possible fallout in adopting the script reforms and its possible implications on the existing implementations.

Introduction
Publishing being an earliest application, for which early hardware and software solutions were developed first to support Tamil script and still continue to be demanding. The next application identified was Word processing. It is evident that Publishing and Wordprocessing were driving the widespread use of computers for Tamil script related works. Also, some indigenous software were developed for publishing and word processing and some add-on software were also developed to enable to use the existing or familiar software developed for English. Over the years, as the capabilities of the Operating Systems were increasingly providing better facilities, software were developed using Tamil script for various requirements such as Teaching/Learning, Games, Video sub-titling, Billing, News readers etc.,.
Technologies behind the possibilities

The switchover from Vector graphics monitors to Pixel graphics monitors has opened revolutionary possibilities for the graphics based computer applications such as DTP, digital graphics, Video editing etc. The Pixel based graphics enabled development of various font technologies, which has thus resulted into soft fonts and then advanced from TTF to OTF. Apart from the Operating System based font technologies, the developers have also developed their own font technologies to provide support for Tamil script in their own software. Also, some developers provide the rendering engines to make the software, platform independent.

Basic requirement

In order to provide support for a script in any software, a well defined standard or atleast its well established implementation called industry standard is required. In the absence of a standard, the propriety script implementations leads to incompatibility of the data crippling data exchange. In Operating Systems, which provide support for scripts by means of fonts, the scripts of languages intended to be used have to be encoded as glyphs in the fonts with the total number of glyphs for a script not exceeding the usable or available code positions of the code page. As the standards established to provide language support on computers are based on the scripts of the language, the standards are expected to provide well defined rules to provide feedback for various typing layouts and letter formation.

Standards prescribed for Tamil letters

Govt of India, in its efforts to promote usage of languages on computers have recommended standards for storage and keyboard layout for typing based on the recommendations of the committee constituted and these recommendations were announced as Bureau of Indian Standards, known as Indian Script Code for Information Interchange (ISCII). But, ISCII has never been implemented on popular Operating System such as Microsoft Windows series as an encoding for general use due to the difficulties associated with its implementation. However, due to the need arising to use Tamil on computers, the local developers have started providing font based solutions, which resulted into creation of non-portable data. In order to encourage and enable data portability, the Govt of Tamilnadu have prescribed an ordered set of glyphs as standard based on the Tamil script. Further, the Govt of India have also prescribed glyph based standards for TTF fonts.

For the power hungry digital devices such as Pagers, Govt of India has prescribed a standard called Indian Standard Code for Language Pagers (ISCLAP). As the font technology has advanced to handle pre-defined character combinations, positioning of glyphs etc, Unicode standard got implemented on Operating Systems and application software. Even though, Unicode for Tamil is also based on the Tamil Script, it is getting implemented in Operating Systems and Application software as an International standard, which has enabled the shrink wrapped software developers to develop methods to handle world languages for their multilingual software products.

Standards prescribed for typing Tamil letters

There are Three standard typing methods prescribed for Tamil script 1. Typewriter, 2. Inscript and 3. Tamil ’99. While Typewriter layout is based on letter formation using glyphs, Inscript and Tamil ’99 are based on the phonetic letter formation.
Typewriter layout

Eventhough, the typing is largely based on the appearance of the letters, for some letters it is not so. For example, the zero width glyphs such as pulli, vowel signs ஐ and ஐ are typed first before typing the base consonant. As per the Unicode standard, the vowel signs have to come (to be typed) after the base consonants. But, while typing Tamil text, some vowel signs like ஐ and ஐ have to be typed first but have to be displayed to right side of the base consonant. For some other vowels like ஐ and ஐ, the vowel signs have to be typed first and then the base consonant to be typed. However, as per the Unicode standard, the vowel signs have to be placed after the consonant. For vowels like ஐ, ஐ and ஐ, the vowel sign have to be split and typed on left and right side of the base consonant.

Inscript and Tamil '99 layout

As these typing layouts are based on the phonetic concept, the vowel signs are typed after typing the base consonant. However, while editing, typing vowel signs contradicts with their visual appearance. For example, eventhough the vowel sign for vowel ஐ appears to the left of the base letter, the typing of vowel sign while editing has to take place towards right side of the base letter.

Difficulties in implementing tamil script

Tamil script has 247 Tamil letters for writing text in Tamil script. Obviously, it is not possible to accommodate all the Tamil letters on the computer keyboard for want of more keys. In order to implement Tamil script on computers, one has to think of possible means of combination of characters or glyphs to form all the letters. In some keyboard layout, such as Typewriter keyboard layout, it is not possible to provide feedback for all the keys to the typist while maintaining the use of aesthetic glyphs for forming vowelised consonants of ஐ, ஐ, ஐ, ஐ.

Such typing of vowel signs imposes difficulties while editing the text. In order to correct the spelling mistakes introduced due to placement of vowel signs by the Unicode standard, the developer has to make use of Private Use Characters (PUA). For such vowel signs, to avoid formation of composite letters is combination with the previous unintended base consonants. While editing the Tamil text, backspace deletes the vowel sign of the composite letter when the user intends to delete the consonant. For example, to delete ஐ in ஐ, when the user presses backspace, instead of ஐ gets deleted the vowel sign ஐ gets deleted. Such behaviors of the software annoy the user. As a result the editing process becomes cumbersome.

While implementing ISCII and Unicode, the basic requirement is to process combination of characters to form composite letters. While ISCII for Tamil has never been implemented in any of the Operating Systems, Unicode is being implemented in many Operating Systems such as Windows XP, Windows Vista etc. However, ISCII for Devanagari was implemented by IBM and APPLE Computer Inc.

Difficulties faced with Tamil '99 keyboard layout

The key feature of Tamil '99 keyboard layout is auto pulli. Eventhough, the auto pulli feature is an added advantage in reducing the number of keystrokes required to type the given Tamil text, it introduces errors in text while typing. It is learnt from the users that the users are comfortable in typing pulli than getting the same automatically. However, when the auto pulli has to be avoided in words like ஐ, ஐ, ஐ inflected forms of ஐ, the auto pulli feature introduces errors. When the user wants to avoid
auto pulli, he has to type ஞ immediately after the first letter. While typing being a subconscious skill, the users are forced to be conscious of words or letter combinations being typed.

Script reform

It is accepted by the archeologists involved in script deciphering that the evolution of script started first from logographic writing system then improved over to syllabic writing system and finally to phonemic writing system. Tamil writing is neither syllabic nor phonemic but it has graphemes for phonemes and orthographic syllables. The secondary forms of vowels are written to right, left and both sides of the base consonant. Vowel signs of ஞ, ட, ஢, ண combine with base consonants to form vowelised consonants, which are designed to be a single glyph. Such orthographic formation leads to difficulties in implementing editing text.

Writing system largely depends on the material used for writing and the writing instrument. When Tamil script implementation is considered for computer, the material used for writing becomes the storage device where Tamil letters are stored in its encoded form. The writing instrument to be considered for computer is keyboard. Even otherwise, if the writing system using pen is considered, it would be convenient to have a representative glyph to represent the Vowels ஞ, ட, ஢, ண.

The history of writing establishes that revision of writing taken place in 1. Adopting different script than the existing script as in Malay, 2. Corrections introduced in the existing letters as in Chinese and Malayalam, 3. Increasing or Decreasing the number of letters as in Kannada.

Reform in Tamil script writing is required to improve the efficiency of composing Tamil letters and editing Tamil text. As the script reform proposes to introduce graphic shapes as graphemes for vowel signs of ஞ and ண, it enables to provide feedback while typing Tamil text and allows to edit the text like deleting visible vowel signs instead of deleting invisible vowel signs.

Required vowel signs for script reform

In Tamil script, additional letters or signs are formed, sometimes by adding a closing circle at the end of the stroke as in ஞ, ஞ ஞ signs and letter ஞ or by adding additional stroke to mark elongation as in letter ஞ. It is also to be noted that the signs added to mark vowel signs of ஢, ண for the Grantha letters is contradicting and thus leads to confuse the users. As a result, the ஞ sign is normally mistaken for ஞ sign because of the rounded circle used at the end of the stroke in ஞ sign.

In order to achieve script reform in Tamil, vowel signs are required to represent the vowels ஞ, ட, ஢ and ண. For ஞ and ட the existing signs ஞ ஞ could be modified to use it in standalone form. Vowel signs for ஢ and ண have to be based on the familiar sign already used to from vowelised consonants.

Suggested vowel signs

In order to achieve the script reform in Tamil, the existing signs for ஞ and ட could be used with little modification as ஞ and ஞ so that these signs could be written to the right side of the base consonant. On the basis of hand movements used for writing vowel signs and the kind of stroke already familiar to Tamil users, the following vowel signs are suggested for vowel signs of ஢ and ண respectively:

![Vowel Signs](image-url)
The above vowel signs are written with the same clockwise hand movement followed for writing other vowel signs like உ, ஊ, ஋ and ஌. Also, such type of vowel signs are used for some vowelised consonants formed with the combination of ஛ and ஝ vowels.

**Benefits of script reform**

The script reform adds the following benefits to Tamil script usage:

1. Reduces the number of glyphs by 72, which are required to form the vowel consonants with உ, ஊ, ஋ ஌ vowel signs.
2. Makes typing and editing Tamil text easier.
3. Enables feedback for typists.
4. Since the number of glyphs required for Tamil script reduces to 1/3, it would be easier to implement Tamil script in every digital gadgets wherever English is implemented.
5. It would enable development of application software and games as available in English.

**Suggestions**

- Vowel signs have to be allowed to be typed in standalone form inspite of its dependent vowel status. Vowel Signs have to be allowed to appear without its preceding dotted circle when typed as a standalone characters.
- The overhead of placing the vowel signs to the right of the base consonant has to be moved to the application software through normalization and convenience of typing the vowel signs as per their appearance has to be enabled in the application software. This is suggested as a feature for Tamil enabled software and this feature has to be considered while testing the software prior to certification of the software.
- Backspace key has to delete the immediate visible letter, instead of deleting the vowel signs of the composite letter.
- Auto pulli feature has to be removed from the Tamil ’99 keyboard layout to enable error free typing.
- Two new symbols are suggested to represent vowel signs of ஛ and ஝ to enable script reform.

**Conclusion**

An attempt has been made to study the present status of Tamil script in digital media and the associated issues. As a remedial action script reform has been discussed and some suggestion are made for consideration. It is believed that the factors discussed and he suggestions made will help to improve the implementation of Tamil script in digital media.
இதலையில் மேற்பாற்றும் சமூகம்

www.tamilinternetconference.org