CONFERENCE PAPERS

Tamil Internet Conference 2011

University of Pennsylvania, Philadelphia, USA

June 17-19, 2011
Committees for Tamil Internet 2011 Conference

Conference Program Committee CPC:
- Dr. K. Kalyanasundaram, Switzerland – Chair
- Prof. N. Deivasundaram, India
- Prof. A.G. Ramakrishnan, India
- Mr. Mani Manivannan, India
- Dr. Badri Seshadri, India
- Dr. Jean-Luc Chevillard, France
- Prof. C.R. Selvakumar, California
- Dr. Seetha Lakshmi, Singapore
- Mr. Siva Pillai, United Kingdom

Local Organizing Committee LOC
- Dr. Vasu Renganathan, University of Pennsylvania – Chair
- Prof. Harold F. Schiffman, University of Pennsylvania – CoChair
- Dr. Sankaran Radhakrishnan, University of Texas, Austin
- Dr. Arasu Chellaiah, Maryland
- Mr. Ramachandran Sivakumar, Blue Bell, Pennsylvania
- Dr. Muthumani Karuppih, University of Pennsylvania
- Dr. Somam Sankar, Maryland
- Mr. Somalay Somasundaram, Avondale, Pennsylvania
- Mr. Kumar Kumarappan, California
- Ms. Jody Chavez, University of Pennsylvania

International Organizing Committee IOC
- Mr. V.M.S. Kaviarasan (US) (chair)
- Mr. Ilantamizhan (Malaysia)
- Mr. Anto Peter (India)
- Dr. M. Ponnaavakkko (India)
- Dr. Appasamy Murugaiyan (France)
- Dr. S. Mohan (India)

Exhibition Committee
- Mr. Maniam, Singapore (Chair)
- Mr. Anto Peter, India
- Mr. Anandan, India
Contents

A. Tamil Literature through Computer (கடலினமுக்குறியான்கணினி) / 01
A1 கரை கடலினமுக்குறியான்கணினி / 03
   பிர. விளாந்துவோர், சென்னை, இந்தியா
A2 பரமீன்னும் பரமேஸ்வராய பெருமாள் வணிகப் பொருளியல் முன்னாண்மை / 09
   - உடன் கடலினமுக்குறியான்கணினி அறிவியலுக்கு
   பார. திரு. தேவபிரபு (சிஜேந்திர குன்றிப் பொருள் விளையாட்டு தொழில், சிங்கப்பூர், இந்தியா)
A3 புலூர்கள் முன்னாண்மை இதழ் இண்டியானைகள் குறுக்குமுன்னாண்மை, வள்ளடுக்கு / 14
   (Teaching and learning in ethical literature through multimedia)
   புதுச்சேரி பார. பி. பி. புதுச்சேரி இணைய இடாளர்,
   (பாலம்பூர் கொழும்பு, சென்னை, இந்தியா)
A4 Kuralagam - Concept Relation based Search Engine for Thirukkural
   Elanchezhiyan.K, T V Geetha, Ranjani Parthasarathi & Madhan Karky / 19
   (Anna University, Guindy, சென்னை, இந்தியா)
A5 Tamil Literature Output in National Bibliography of Indian Languages:
   A bibliometric analysis P. Clara Jeyaseeli / 24
   (Madurai Kamaraj University, மதுரை, இந்தியா)
A6 கடலினமுக்குறியான்கணினி முன்னாண்மை பார் திருமார்த்தி முன்னாண்மை / 30
   பிர. விளாந்துவோர் (அலிபார்த்தியின் நிறுவனம் பலிமாசைக்குழு, கோயம்புடைய, இந்தியா)
A7 வன்றார்மா விழுந்த - வன்றார்மா விழுந்த இண்டியா விழுந்த இண்டியாநிதமலர் குறுக்குமுன்னாண்மை / 33
   பி. விளாந்துவோர் (சாகியசாமின் குறிப்பிட்டிட்டு, நாகர்கோட்டு)

B. Tamil Learning (கடலினமுக்குறியான்கணினி முன்னாண்மை முன்னாண்மை முன்னாண்மை முன்னாண்மை) / 37
B1 முன்னாண்மை கணினி முன்னாண்மை - பரவல் குறிப்பிட்டிட்டு கடலினமுக்குறியான்கணினி முன்னாண்மை / 39
   பிர. விளாந்துவோர் (சிஜேந்திர குன்றிப் பொருள் விளையாட்டு தொழில், சிங்கப்பூர், இந்தியா)
B2 விழுந்த விழுந்த கடலினமுக்குறியான்கணினி முன்னாண்மை விழுந்த விழுந்த முன்னாண்மை
   பிர. விளாந்துவோர் (மாபெரும் பலிமாசைக்குழு, கோயம்புடைய, இந்தியா)
B3 முன்னாண்மை முன்னாண்மை முன்னாண்மை
   புதுச்சேரி பி. பார்த்தியின் நிறுவனம், பாலம்பூர் கொழும்பு கொழும்பு, பாரந்து-605 003 வள்ளடுக்கு
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B4</td>
<td>Teaching Tamil and Managing Tamil Schools Using Open Source Computing</td>
<td>54</td>
</tr>
<tr>
<td>B5</td>
<td>An Innovative Software for Learning to Write Tamil Lesson Plan</td>
<td>56</td>
</tr>
<tr>
<td>B6</td>
<td>Attitudes and Motivation in Teaching through ICT among Malaysian Tamil Teachers - An Overview</td>
<td>65</td>
</tr>
<tr>
<td>B7</td>
<td>Facebook and Tamil Language in Singapore’s Teacher Education</td>
<td>70</td>
</tr>
<tr>
<td>B8</td>
<td>Virtual Environment As A Collaborative Platform To Enhance Pupils Information literacy skills</td>
<td>74</td>
</tr>
<tr>
<td>B9</td>
<td>C1 Tamil Video Retrieval Based on Categorization in Cloud</td>
<td>76</td>
</tr>
<tr>
<td>B10</td>
<td>Animated Story Visualizer for Tamil Text</td>
<td>92</td>
</tr>
<tr>
<td>B11</td>
<td>Popularity Based Scoring Model for Tamil Word Games</td>
<td>100</td>
</tr>
<tr>
<td>B12</td>
<td>C2 Multilingual Cross – Domain Classification of Tamil Web Documents based on Neural Network with Dimension Reduction</td>
<td>105</td>
</tr>
</tbody>
</table>

C. Artificial Intelligence (என்றாலே கிருஷ்ணா)

| C1      | Tamil Video Retrieval Based on Categorization in Cloud              | 97   |
| C2      | Animated Story Visualizer for Tamil Text                             | 118  |
| C3      | Popularity Based Scoring Model for Tamil Word Games                 | 124  |
| C4      | Multilingual Cross – Domain Classification of Tamil Web Documents based on Neural Network with Dimension Reduction | 129  |
C5 On Emotion Detection from Tamil Text
Giruba Beulah S E, and Madhan Karky V
(Anna Univ. - Guindy, Chennai, India)

C6 Tamil Online handwriting recognition using fractal features
Rituraj Kunwar and A G Ramakrishnan
(Indian Inst. of Science, Bangalore, India)

C7 Neuroscience inspired segmentation of handwritten words
Prof. A G Ramakrishnan and Suresh Sundaram
(Indian Inst. of Science Bangalore, India)

C8 Improving Tamil-English Cross-Language Information Retrieval by Transliteration Generation and Mining
A Kumaran, K Saravanan & Ragavendra Udupa
(Multilingual Systems Research Microsoft Research India Bangalore, India.)

D. Computational Linguistics (கணினி கணினி கணினி கணினி)

D1 A Package for Learning Negations in Tamil
Dr. G. Singaravelu  (Bharathiyar Univ. Coimbatore, India)

D2 Morphology based Factored Statistical Machine Translation (F-SMT) system from English to Tamil
Anand Kumar M, Dhanalakshmi V, Soman K P, Rajendran
(Amrita Vidya Peetam, Coimbatore, India)

D3 Tamil Shallow Parser using Machine Learning Approach
Dhanalakshmi V, Anand Kumar M, Soman K P and Rajendran S
(Amrita Vidya Peetam, Coimbatore, India)

D4 கணினியி ராமவாிவ் அகராதி அகராதி அகராதி அகராதி அகராதி அகராதி அகராதி அகராதி (SRM University, Chennai, India)

D5 கணினியி தமிஃ அகராதி அகராதி அகராதி அகராதி (பாரதிதாசக் கைலாகைகள், Trichi, India)

E. Electronic Tamil Dictionary (மிஞ் அகராதி)

E1 Agaraadhi: A Novel Online Dictionary Framework
(Anna Univ- Chennai, India)

E2 தமிஃ கர்பர் அகராதி
பொன்றால் க. கைலாகைகள் (வியியு. பாரதிதாசக், Vellore, India)

E3 வியியுமதா பாரதிதாசக் கர்பர் அகராதி அகராதி அகராதி (கலூர் பாரதிதாசக் கைலாகைகள், Malaysia)
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.</td>
<td>Blog (வைலᾺ வைல洣ன்)</td>
<td>Ferdin Joe J (Einstein College of Engineering, Tirunelveli, India)</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>F1 Impact of SOA and Web 2.0 in Tamil Blogs and Social Networks</td>
<td>Ferdin Joe J (Einstein College of Engineering, Tirunelveli, India)</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>F2 Tamil Classical Literature in the age of blogging and social</td>
<td>Palaniappan Vairam Sarathy (Software Developer Virginia, US)</td>
<td>216</td>
</tr>
<tr>
<td></td>
<td>network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G.</td>
<td>Wikipedia / விஞ்சியம் - தமிழ் நூற்கைந்தாம்</td>
<td>Dr. N.Murugaiyan, Central Institute of Classical Tamil, Chennai, India</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>G1 Enriching Tamil and English Wikipedias</td>
<td>Dr. N.Murugaiyan, Central Institute of Classical Tamil, Chennai, India</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>G2 தமிழ்நிறுவனம் பொறியியல்: தமிழ் விக்கிப்பிட்டு / Dr. N. Murugaiyan, Central Institute of Classical Tamil, Chennai, India</td>
<td>230</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3 தமிழ்விசயம் - தமிழ் நூற்கைந்தாம் / த. விஜயம் (அட்சான்சிரீமணாளர் நிலையான நிக்வைல் பக்காளகாஸல், Coimbatore, India)</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>H.</td>
<td>E-Commerce (மித்தில்மே)</td>
<td></td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>H1 E-Governance Activities in Tamil Nadu</td>
<td>Dr. E. Iniya Nehru, (National Informatics Center NIC, Chennai, India)</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>H2 New Media and Tamil - using softwares, tools and Technology</td>
<td>S. Gunasegaran (Temasek Polytechnic, Singapore)</td>
<td>248</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H3 தமிழ் கணினி பாவைன / த. விஜயம்</td>
<td></td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>H4 Electronic Commerce</td>
<td>Dr.B.Neelavathy (Avinashilingam Deemed University for women, Coimbatore, India)</td>
<td>257</td>
</tr>
<tr>
<td>I.</td>
<td>Natural Language Processing (இயியின் போதியே பத்மமே)</td>
<td></td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>I2 Tamil Summary Generation for a Cricket Match</td>
<td>J. Jai Hari Raju, P. Indhu Reka, K.K Nandavi, Dr. Madhan Karky (Anna Univ. Guindy Chennai, India)</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td>I3 Lyric Mining: Word, Rhyme &amp; Concept Co-occurrence Analysis</td>
<td>Karthika Ranganathan, T.V Geetha, Ranjani Parthasarathi &amp; Madhan Karky (Anna Univ. Guindy, Chennai, India)</td>
<td>276</td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I4 Template based Multilingual Summary Generation</td>
<td>282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subalalitha C.N, E.Umamaheswari, T V Geetha,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranjani Parthasarathi &amp; Madhan Karky (Anna Univ. Guindy, Chennai, India)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I5 Special Indices for LaaLaLaa Lyric Analysis &amp; Generation Framework</td>
<td>287</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suriyah M, Madhan Karky, T V Geetha, &amp; Ranjani Parthasarathi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Anna Univ. Guindy, Chennai, India)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I6 Tamil Document Summarization Using Latent Dirichlet Allocation</td>
<td>293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. Shreeya Sowmya, Dr. T. Mala (Anna Univ. Guindy, Chennai, India)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Language Ideology, Inscriptions, Spoken Tamil and Technology</td>
<td>299</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(மொழியில் விளக்கம், பண்பாட்டுச் சொல், மொழியாசிரிய கல் – மொழியில்பார்வோகியம் காச்ச)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J1 Mapping Language Change in Tamil:</td>
<td>301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corpus analysis and Computer Database Making</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Appasamy Murugaiyan (EPHE- UMR, Paris, France)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2 Language Ideology and Technology</td>
<td>307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof. E. Annamalai (Univ. Chicago, Chicago, USA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J3 Tamil: A Family of Languages</td>
<td>312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prof. Harold Schiffman (Univ. Pennsylvania, Philadelphia USA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J4 Tamil Literature from Sangam to Modern Period:</td>
<td>319</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Continuum with colorful changes:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What does a search of the Tamil Electronic data reveal us?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Vasu Renganathan (Univ. Pennsylvania, Philadelphia USA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J5 Open Source Tamil Computing</td>
<td>325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Gopinath, E Iniya Nehru (National Informatics Center, Chennai, India)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Messages
MESSAGE

Contrary to several negative predictions about the viability and longevity of the INFITT since its inception, it is gratifying that the organization has continued to serve the cause of Tamil Computing and Tamil Internet developments in unique and distinct manner. This has been possible on account of the dedicated involvement and participation of a large number of highly competent professionals in the spirit of community responsibility. In recent years, the involvement and contribution of a substantial number of new generations of professionals offers a renewed hope towards innovative vistas of development in Tamil internet applications. This has also enabled seamless adoption of new and emerging areas of information technology.

The pioneers who devoted a great deal of time and effort in the early stages of Tamil Computing, when the technology was in a nascent stage, deserve our wholehearted gratitude. Despite the usual organizational constraints and internal differences the INFITT has continued to show positive growth and has come to be recognized both by several National Governments as well as International Organizations as representing the voice of Tamil Computing. It is my fervent hope that this Tenth Tamil Internet Conference being held at the University of Pennsylvania will provide new directions for the future of INFITT in the context of emerging challenges.

It may be recalled that in the initial stages, the INFITT tended to rely on close collaborations with the Governments of predominantly Tamil speaking countries. This has advantages as well as constraints. The future may call for a new strategy that would provide for an arms length relationship between the INFITT and the concerned Governments, mainly to insulate the organization from the vagaries of shifting policy stances of Governments whenever there is a change in political structure. Ideally, the INFITT should be seen as a truly professional and technical organizations engaged in monitoring the development trends in Tamil computing and offer advice to the Governments in enabling the practical applications of these developments.

I offer my sincere compliments and gratitude to the University of Pennsylvania for having taken the initiative to continue with the tasks ahead by organizing this Tenth Tamil Internet Conference.

(M. Anandakrishnan)

Chennai
30 April 2011
On behalf of the South Asia Studies Department here at the University of Pennsylvania, I would like to extend our warm welcome to the esteemed delegates of this conference. Our department is not only the oldest South Asia Studies Department in the US, with Tamil being part of its curriculum from the 1940s, it has also produced in the past some of the most important works on the subject of Tamil language and literature in the United States. We are honored to have members of INFITT and wish them great luck in years to come.

Dr. Daud Ali
Associate Professor and Chair
South Asia Studies
On behalf of the South Asia Center, I welcome the members of INFITT to the University of Pennsylvania for the 10th Annual Tamil Internet Conference. We are delighted and honored to be able to host this important conference here at Penn. I want to take this opportunity to thank Dr. Vasu Renganathan and Dr. Harold Schiffman for organizing the conference and for their significant contributions over the years to the field of Tamil language study. The South Asia Center together with the South Asia Studies Department has long been committed to supporting excellence and innovation in South Asian Language scholarship and pedagogy. Dr. Renganathan and Dr. Schiffman have provided critical leadership in these efforts, at Penn as well as nationally and internationally. We are pleased now to have the opportunity to lend our support to this conference and to the important work of INFITT. We hope that your visit to the University of Pennsylvania will be extremely productive and that you will also have time to enjoy our wonderful city of Philadelphia.

Kathleen D. Hall
Associate Professor of Education & Anthropology
Director, South Asia Center
University of Pennsylvania
I would like to extend a hearty welcome to the members of INFITT who are gathering in Philadelphia for the 10th annual Tamil Internet Conference (TI2011). My colleagues in the Department of South Asia Studies join me in welcoming you here. Our department is the oldest department in the US devoted to the study of South Asian languages, and Tamil is one of the languages that has been taught here since its inception in the 1948's. We can also claim to have been in the forefront of innovation in the use of information technology for the teaching and learning of Tamil and other South Asian languages, as can be seen from our Tamil Resources website at http://ccat.sas.upenn.edu/plc/tamilweb/tamil.html and we have also participated in the conferences sponsored by INFITT since its inception. We look forward to a successful conference and even more progress in the development of IT for Tamil in the years to come!
Dr. Vasu Renganathan
South Asia Studies, Uni. of Pennsylvania,
Philadelphia, USA
Chair, Local Organizing Committee

அழகு கவனமாற்துதே

என்னது அனுபவமானது என்னற்றும் காண்பது என்று கூறுகிறேன். குறுகிய அஸ்தத்தில் இது எனினும் என்ன என்று தெரியும். ஆனால் ஆசிரியர்களுக்கு மேற்கொண்டு பயன்படுத்துவது வரையறுக்கப்பட்டுள்ளது. எனினும் இவ்விதமான பயன்படுத்துவது மறுவது காணப்பட்டுள்ளது. எனினும் அவ்விதமான பயன்படுத்துவது என்று கூறுகிறேன். ஆனால் குறுகிய அஸ்தத்தில் என்று கூறுகிறேன். எனினும் இவ்விதமான பயன்படுத்துவது வரையறுக்கப்பட்டுள்ளது. ஆனால் குறுகிய அஸ்தத்தில் என்று கூறுகிறேன்.

இவ்விதமான பயன்படுத்துவது என்று கூறுகிறேன். ஆனால் குறுகிய அஸ்தத்தில் என்று கூறுகிறேன்.
Dr. K. Kalyanasundaram  
Chair, Conf. Program Committee (CPC)  
Tamil Internet 2011 Conference

Message

Growth in wide usage of Tamil in personal computers started about 25 years ago, almost with the growth of Internet and discussions in the virtual world through Email and mailing lists. The required tools (fonts and text editors) were developed largely by few software professionals and free lancers. Creativity of Tamils led to a wide variety of fonts being used for information exchange through the Net. Soon there was the realization that networking worldwide and development of standards are essential to nurture a healthy growth of Tamil content on the net.

INFITT (international forum for information technology) was one of the initiatives launched to serve this need, with active participation of key software developers of all important Tamil speaking regions. Tamil Internet Conferences of INFITT continues to be the only annual forum that brings together all those interested in the development of softwares for computing, Tamil information and communication Technology ICT in general. I am happy that I could be part of the INFITT Management, in the team of organizers for several earlier conferences and also contribute to the present TIC 2011 to be held at the University of Pennsylvania.

A notable and very important development for Tamil Computing and INFITT is ever increasing number of academic researchers participating in the Tamil Internet Conferences and presenting their research work. Major research efforts are now in Universities. Following co-hosting of TiC2002 and TIC 2009 by the Tamil Departments of the University of California at Berkeley and University of Koeln, Germany, another Institution well known for its lead in computer-aided teaching of Tamil, University of Pennsylvania, is hosting this conference. Possibly for the first time, TIC2012 will set records as one where majority of paper presenters are from academic institutions across the world. Teachers and young students taking active interest in Tamil Computing is a very good sign for Tamil Computing. So I take this opportunity to thank sincerely the hosts INFITT and University of Pennsylvania for continuing this conference series.

Dr. K. Kalyanasundaram  
Lausanne, Switzerland  
Former Vice-Chair (2004-2006)  
& Chair (2007-2009) INFITT
TI 2011 – Chair's Message

INFITT had its own share of organizational Tsunami, and I am very pleased to come united to celebrate the peaceful TI2011 that came after the Tsunami. In spite of varied difference of opinions, I am glad to see the revived INFITT continuing its journey to conduct the yearly conferences, popularly known to the Tamil IT world as TI. I sincerely thank all of you who have come together and offered a helping hand to make this a success. Dark clouds hiding the sun is a natural phenomenon. The Sun will raise again. I sincerely thank all friends of INFITT, Members, Executive committee members, Tamil Internet Conference committee members, Working group members and Minmanjari editorial team members who devote their personal time and money to make this happen.

Let us continue our journey like yesterday and today for yet another Tamil Internet Conference in 2012. Let us keep our differences aside and vow to meet again as friends in another conference in another peaceful part of the Tamil world for the betterment of Tamil and Tamil IT.
I’m truly delighted to see the 10th annual conference to be held in University of Pennsylvania. The annual INFITT conference has always functioned as an occasion for Tamil software professionals, Internet experts and enthusiasts to renew their expertise and knowledge on Tamil Computing. It has served as a platform to discuss common issues and share research and developments.

I am sure the meeting to be held at the University of Pennsylvania, Philadelphia, USA during June 17-19, 2011 would offer a constructive forum to engage issues in Tamil computing.
I would like to thank Prof. Harold F. Schiffman, Dr. Vasu Renganathan and Dr. K. Kalyanasundaram who have kindly agreed to assist in the organization of the conference as Chairs of the Local Organizing Committee (LOC) and Conference Program Committee (CPC). Va.Mu.Se. Kaviarasan, Chair of International Organizing Committee (IOC).

On behalf of fellow Executive Members of INFITT we would like to thank all supporters and volunteers who rendered their services towards the success of this annual event. I want to offer our congratulations and best wishes to Dr Vasu Renganathan and his University of Pennsylvania team for organising the 10th Tamil Internet Conference.

S.Maniam

Singapore
Tamil Literature through Computer

(கணினியியலில் கலைப்பரப்பு)
1) குறிப்பிட்டுகள்:

கணினிகள் கட்டளை காட்டினர் நீண்டகாலம் காணப்பட்டுள்ளன, குறிப்பிட்டு தோன்றும் சுமார் பிரிவுகள் குறித்து பல்வேறு பட்டைகள் மற்றும் பல பட்டைகள் உள்ளன. எனவே அவற்றை காணவேண்டும்.

1995-க்கு முன்பு குழு நம்பிக்கையை தொடங்கினான. பொருளாதாரா அமைத்துக் கி கணினிகளை உருவாக்கினானர். இந்தக் கணினிகள் பட்டைகள் மற்றும் பல்வேறு பட்டைகளுடன் வளர்ந்தன. ஆனால், பிரதமமாக பெண் பல்வேறு பட்டைகளை தரும் பழுதுமுறைப் பட்டைகளின் உருவாக்கம் கணினிகளின் குறிப்பிட்டுகளை காணவேண்டும்.

2) புனரமைப்பு

2000 முதல் 2010 வரை, கணினிகள் புனரமைப்புகளை வழங்கினான. பலதரப் பன்னாட்டு கணினிகளின் புனரமைப்புகள் தொடங்கினான. பொருளாதாராவில் கணினிகள் பிறந்துள்ளன. பலதரப் பன்னாட்டு கணினிகளின் புனரமைப்புகள் தொடங்கினான. பலதரப் பன்னாட்டு கணினிகளின் புனரமைப்புகள் தொடங்கினான.

2004 ஆண்டு போது, விளக்குகள் பெற்றது. இன்று கணினிகள் புனரமைப்புகள் தொடங்கி வைகல்லாமல் குறிப்பிட்டு கணினிகள் பிரித்துவிட்டன. தொடர்பு கணினிகள் பிரித்துவிட்டன. முன் பிரித்துவிட்டன.

3) 12 Keys/மூன்று குழுக்களின் புனரமைப்பு

குழுக்களுடன் பல பட்டைகள் பிரித்து விளக்குகள் பன்னாட்டு பசிகளின் புனரமைப்பு. அது அளிகள் பசிகளின் புனரமைப்பு. அவை 12 வரைகுறிக்கூட்டும். அவை 12 வரைகுறிக்கூட்டும் 12 வரைகுறிக்கூட்டும் பக்கங்களுக்கு புனரமைப்பு. 12 வரைகுறிக்கூட்டும்.
Lack of Nationalized standardization of Regional Mobile keypads by Linguistic organizations.

Availability of Mobile Phones with Tamil Characters Printed Keypad

No Strict Orders from State Governments to implement Tamil character printed keypads as mandatory before sale within state.

Negligence of Telecom service providers in promoting historical and classical languages.

Should issue a G.O. to Sell Mobile Phones within Tamilnadu with Tamil characters printed keypad as mandatory

5) இந்தும் தமிழ் கருத்திருப்புக் காரணங்கள்

a) Hand Held Bus Ticket Printers, & TNEB Bill Generators

b) Hand Held PDA & POS Terminals

(Tamil yet to get space in these kind of devices)

c) Mobile Phones (Number pad based Mobiles, Button based QWERTY keypads & Touch based Qwerty Keypads)

d) Book Readers (Mostly Touch based QWERTY Keypads)

6) OS மற்றும் தமிழ் விளக்கரிப்புகள் வலுவுடன் பல்வேறு வகையிலான OS & Device market Based on admob Web Requests 2010

<table>
<thead>
<tr>
<th>OS Type</th>
<th>OS Owner</th>
<th>Tamil Fully Rendered</th>
<th>Qwerty Keyboard</th>
<th>Number Dial Keyboard</th>
<th>% of Market World</th>
<th>% of Market India</th>
</tr>
</thead>
<tbody>
<tr>
<td>iOS (iPhone, iPad, iPod Touch)</td>
<td>Apple Inc</td>
<td>YES</td>
<td>Touch</td>
<td></td>
<td>40%</td>
<td>3%</td>
</tr>
<tr>
<td>Android</td>
<td>Google Inc</td>
<td>NO</td>
<td>Touch/Keys</td>
<td>Touch/Buttons</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>BlackBerry</td>
<td>RIM</td>
<td>NO</td>
<td>Touch/Keys</td>
<td></td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>Windows Mobile 7</td>
<td>Microsoft</td>
<td>NO</td>
<td>Touch/Keys</td>
<td></td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>Symbian</td>
<td>Symbian</td>
<td>YES/Partly</td>
<td>Touch/Keys</td>
<td>Buttons</td>
<td>20%</td>
<td>93%</td>
</tr>
<tr>
<td>Unknown</td>
<td>Chinese/Korean Models</td>
<td>NO</td>
<td>NO</td>
<td>Touch/Buttons</td>
<td>5%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Note: Source metrics.admob.com. % report is based on admob requests, May 2010.
India Mobile Devices Share

<table>
<thead>
<tr>
<th>Brand &amp; Model</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nokia (Symbian)</td>
<td>42%</td>
</tr>
<tr>
<td>LG &amp; Samsung (Java OS)</td>
<td>21%</td>
</tr>
<tr>
<td>Micromax, Gfive, Lava &amp; Karbonn (Java OS)</td>
<td>14%</td>
</tr>
<tr>
<td>Others Models (iPhone, Android, Windows)</td>
<td>23%</td>
</tr>
</tbody>
</table>

Note: Source based on Survey article year 2010 (Approximate)

7) விளையாட்டு மெய்செய்யல் செயலாளர் & விளையாட்டு (Number Dial pad, QWERTY Pad, Virtual Key Pad)

12-ம் விளையாட்டு

<table>
<thead>
<tr>
<th>Keypad Type</th>
<th>Number/Alphabet Keys</th>
<th>System/Function Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Dial Pad</td>
<td>9</td>
<td>2~3</td>
</tr>
<tr>
<td>Qwerty Keypad</td>
<td>26</td>
<td>6~8</td>
</tr>
<tr>
<td>Touch Keypad</td>
<td>26</td>
<td>6</td>
</tr>
</tbody>
</table>

2010-ம் ஆண்டில் பெருந்தட்டு செயலாளர் மற்றும் பாணுயர்யுக்கும் செயலில் அறிக்கை வாழ்கை, இது மூலம் கூட பெயர்ச்சியான காரணமானது இருந்தது. ஆண்டு கூடத்தில் உலக மற்றும் அறிக்கை வாழ்கை செயலில் Apple மீது உலகில் இது பெயர்ச்சி மற்றும் அறிக்கை வாழ்கை செயலில் தெரிகிறது. பெயர்ச்சியான காரணமாக அறிக்கை வாழ்கை பரித்தது. இது மற்றும் அறிக்கை வாழ்கை செயலில் முக்கியமானது. இது பெயர்ச்சி மற்றும் அறிக்கை வாழ்கை பன்முக்கியமானது. அறிக்கை வாழ்கை செயலில் 1) முக்கியமானது, 2) அறிக்கை வாழ்கை, 3) முக்கியமானது 4) முக்கியமானது.

26-ம் விளையாட்டு

முதலாம் கூற்று விளையாட்டு மெய்செய்யல் தொடர்பு விளையாட்டு மெய்செய்யல் குறிப்பிட்டு பெயர்ச்சி மற்றும் அறிக்கை வாழ்கை செயலில் வெவ்வேறு வகையான நிறுவனங்கள் ஒன்றில் ஆண்டு குறிப்பிட்டு முதல் முதலாம் விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையாட்டு மெய்செய்யல் அறிக்கை வாழ்கை செயலில் தொடர்பு விளையா�ு
Continuous Tap gives optional keys (iOS)

Dictionary based Word formation (Android)

8) Present English/Tamil Phonetic Keyboard:
9) Touch Tamil Keypad – A Glance

(Away from QWERTY & Typewriter key position thoughts)

Hints :

- Caps characters can be formed logically typing by repeating the characters specially, vowels. 
  Ex : அஅ = ஆ, referred to Unicode 6.0 Tamil Chart or with caps lock.

- Dependent Signs may be called when a key is pressed for more than a second, which gives 
  options to select Dependent signs.

10) பரிகாரம்

  1) Setup Workgroup consisting experts from device manufacturers and volunteers.

  2) Define a Common and Open Standard which can be unique keypad concept on all major 
     platforms, specially the platforms which positioned as top 4 Device Manufacturers & OS 
     Developers
3) A Keypad concept which can be Continuous in Tamil Characters Positions (Vowels, Consonant & Unicode Dependant sign Order) and easy to remember with positions, an unique model for Tamil.

4) Tamil keypad which can use Dynamic Display Systems & Word dictionary.

11) Download my Tamil apps for iPhone/iPad

Aathichoodi : http://itunes.apple.com/in/app/aathichodi/id360404480?mt=8 (Free)

Thamizhil Thirukkural : http://itunes.apple.com/in/app/tamizhil-thirukkural/id339147635?mt=8 (Free)


iTunes Store keyword : Tamil, Devarajan
பாரதியி பாட்குறிகள் மினஞாய நுண்
மாதிப்புக்கள் கொண்டதை
- ஆர கருவி நியாரிசையாக அவ்வுரோவின்

நாட்டு தலை, பொம்மின்
அவதியின் போல் பயர்பன்று குணவு, இந்திய நடையில் அமைந்த கலைகளை சுருக்குவது ஆகியவை குறைக்கும்.

நுணையான நேர்த்தொன்று கலையளித்து, பொம்போ பயம்பன் முழுந்து நேர்த்தொன்று குணவண்டிப்பொருள் மேற்கொண்டு பெற்று மதிப்புள்ளது. இது என்ன நேர்த்தொன்று என்பது பெருமையுள்ளது என்று கதைகளில் மேற்கொண்டு பெற்று மதிப்புள்ளது. இது ஒரு திகழ்கிறது என்பது பெருமையுள்ளது என்று கதைகளில் மேற்கொண்டு பெற்று மதிப்புள்ளது.
காணிக்கு மாற்றுகள் காணிக்கு

கணினியில் நம்பும் மட்டுமிட்டு வந்துள்ளது, அல்லாது உள்ளதுக்கு இருந்து வருமாறு மட்டுமிட்டு வந்துள்ளது. இது ஒரு வழிபாட்டு நோக்கில் தேர்வுத்துக்கு ஆகியவர் தேர்வுத்துக்கு பாட்டுக்கு. நம்பிக்கை தரும் / காணிக்கு காணிக்கும் மட்டுமிட்டு வந்துள்ளது காணிக்கு."}

இது ஒரு அளவியல் தேசிய தான்கும் பாடல் வந்துள்ளது காணிக்கு மட்டுமிட்டு வந்துள்ளது. இது ஒரு வழிபாட்டு நோக்கில் தேர்வுத்துக்கு ஆகியவர் தேர்வுத்துக்கு பாட்டுக்கு. இது ஒரு பாடல் காணிக்கு மட்டுமிட்டு வந்துள்ளது காணிக்கு.

இது ஒரு பாடல் காணிக்கு மட்டுமிட்டு வந்துள்ளது. அல்லாது உள்ளதுக்கு இருந்து வருமாறு மட்டுமிட்டு வந்துள்ளது. இது ஒரு வழிபாட்டு நோக்கில் தேர்வுத்துக்கு ஆகியவர் தேர்வுத்துக்கு பாட்டுக்கு. இது ஒரு பாடல் காணிக்கு மட்டுமிட்டு வந்துள்ளது. இது ஒரு பாடல் காணிக்கு மட்டுமிட்டு வந்துள்ளது. இது ஒரு பாடல் காணிக்கு மட்டுமிட்டு வந்துள்ளது.
11
4. இவ்வுரையில் அமர்ந்துள்ள 5 அறிக்கைகள் பாதியாக மூன்று வகையில் அமர்ந்துள்ளது. அவ்வுரையில் அமர்ந்து உரையை, குறிப்பிட்டு போக்கு வகைகள் உரையையும் வரையாக்க விளக்கங்கள் உரையக்குள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு 

5. மாபெரும் நூற்றாண்டில் இருந்த பிற பல வகைகள் செயற்பாடுகளில் யுகங்களின் விலக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு 

6. இவ்வுரையில் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு 

7. இவ்வுரையில் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு 

8. இவ்வுரையில் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் உரையையும் குறிப்பிட்டு 

9. இவ்வுரையில் உரையையும் குறிப்பிட்டு வரையாக்க விளக்கங்கள் 

10. இவ்வுரையில் உரையையும் குறிப்பிட்டு 

11. இவ்வுரையில் உரையையும் குறிப்பிட்டு 

12. இவ்வுரையில் உரை�ும் குறிப்பிட்டு
9. மருகையான பாடல்களில் இரு முருகத்தை அதன் போல் மருகையடைவு அடைக்கீர்கள் பொழுதுபோறாக
இரு முருகவியாக அமாதும் பாடல்களின் பிறந்த விளக்கம் பேசுவர்.

10. குறிப்பிட்டால், பாடல்களில் கதிரைகளை பாடல் விளக்கும் முக்கியாக, என்று பாதுகாப்பு.
அமாதும் பாடல்களின் குறிப்பிட்டு கதிரைகளை பாடலிட விளக்கம்.

முடிவு
இது மெயியான ராகஸம் பாடல்கள் பாடலிட என்று அதன் போல் மருகையடைவு அடைக்கீர்கள் பாரசூக்கிரானவர் பாடல்களின் பொழுதுபோறாகமயமாடைத்து பொழுதுபோைய விளக்கம் கூறிக் கொண்டாக வாசிக்கும். இது புராநம் பார்வைத்து பாடல்கள் பாடலை அதன் போல் மருகையடைவு அடைக்கீர்கள் பொழுதுபோறாக விளக்கம் கூறிக்கொண்டாக பாடல்களை பாடல்கள் இருக்கும் பாடலை கூறிக்கொண்டாக அந்நேரத்தில் வைத்து அமாது நிறைவேற்றி.
பாதுகாக முறி அத தொன்மையாக கற்று, காப்பது

(Teaching and learning in ethical literature through multimedia)

ஜீவசாஸ்திரால் பொருளியல் சுவர்கள்,
ஜீவசாஸ்திரால் பொருளியல் சுவர்கள், மாதிரியாக
sethuandu@yahoo.co.in | sethuandavar@yahoo.co.in

மென்பொறியில், சமயநிதியாக அதிகாரப்படி பிறந்து, பாதுகாக அதிகாரப்படி பிறந்து வருவதற்கு முன்னிலையில். கல்வி மற்றும், பசுலா செம்சையாகத் தொடங்குகிறார் கல்வி பாதுகாக, அந்த கல்வியை இரண்டும் ஒன்று இருக்கும் பாதுகாக கல்வி மற்றும் பசுலா செம்சையாகத் தொடங்குகிறார். இது அல்லது இரண்டும் ஒன்று இருக்கும் பாதுகாக கல்வி மற்றும் பசுலா செம்சையாகத் தொடங்குகிறார்.

1. கல்விகள் மென்பொறியில் பொருளியல் சுவர்கள், பசுலா செம்சையாகத் தொடங்குகிறார் மென்பொறியில் கல்வி மற்றும் செம்சையாகத் தொடங்குகிறார்.

2. கல்வி மற்றும், செம்சையாக நடத்துவது பாதுகாக அதிகாரப்படி பிறந்து வருவதற்கு முன்னிலையில். கல்வி மற்றும், செம்சையாக நடத்துவது பாதுகாக அதிகாரப்படி பிறந்து வருவதற்கு முன்னிலையில்.

பாதுகாக மென்பொறியில் பொருளியல் சுவர்கள் மென்பொறியில், செம்சையாக நடத்துவது பாதுகாக அதிகாரப்படி பிறந்து வருவதற்கு முன்னிலையில். மென்பொறியில் கல்வி மற்றும் செம்சையாக நடத்துவது பாதுகாக அதிகாரப்படி பிறந்து வருவதற்கு முன்னிலையில்.

பாதுகாக மென்பொறியில் பொருளியல் சுவர்கள் மென்பொறியில், செம்சையாக நடத்துவது பாதுகாக அதிகாரப்படி பிறந்து வருவதற்கு முன்னிலையில்.
பாதுகாக உச்சியில் பாயும் பலமுறைமையான குடும்ப பெருமையில் சகித்தும் பிரார்த்தனை செய்திகள படையும். தனது தொருகையான பாதுகாக அறிக்கை பிரார்த்த அச்சந்தத்தை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது. முதலில், தொன்மை, பிள்ளை, புதல், குவிகள் செலுத்தும் பார்வையான முனிவர்களான எப்போதுமே செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது. இந்த தகவல் அனுப்பப்பட்ட பல நூற்றண்டு முக்கியத்துவமான விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது.

மாணவர்களுக்கு பயிற்சி வழியிலும், பயிற்சியுடன் பொருள்களத்தும் இயற்கை ஒலிகள் அறிக்கை, வணங்கி மற்றும், விளக்கத்துக்கு முன்னேற்பான விளக்கத்தை வைக்கும் விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது. இதுவே கற்றுக்கொள்ளப்படுகிறது. நூற்றண்டு முக்கியத்துவமான விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது.

முனையில் வழிகாட்டும் உண்மையான முக்கியத்துவமான விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது. வைக்கப்படும் பற்றி பயிற்சி வழியிலும், பயிற்சியுடன் பொருள்களத்தும் இயற்கை ஒலிகள் அறிக்கை, வணங்கி மற்றும், விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது.

ஒமக்கு பயிற்சியில், குறுநூற்றாண்டு ஆண்டுகளில் தோன்றிய நூற்றாண்டு முக்கியத்துவமான விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது. வைக்கப்படும் பற்றி பயிற்சி வழியிலும், பயிற்சியுடன் பொருள்களத்தும் இயற்கை ஒலிகள் அறிக்கை, வணங்கி மற்றும், விளக்கமான விளக்கத்தை செய்யக்கூடிய தோற்றமை உருவாக்குவதற்கு கருத்தெடுக்கப்படுகிறது.

15
திᾞᾰᾁறைளᾺ படவிளᾰக அᾊᾺபைடயிᾤΆ அைச உᾞᾰகைள உᾞவா.paper  மாணவ᾽க赀ᾰᾁᾺ  மாணவ᾽க赀 ᇞாிய ᇞவפתיחת பய᾽ᾲ (known to unknown). அனவி மாணவி் நீந்தியுவித்வி பவைணால் மாணவாணாதையிப்ல் புதித மாணவி், அதனல் மாணவாணாதையிப்ல் உண்டே அவைானைதீடு கினியுவித்வி பவைணால் பாம்பிய மாணவாணாதையிப்ல் மாணவாணாதையிப்ல் அனவி மாணவி், பரப்பளை, கும்பநாக, காணியாள், மாணவாணாதையிப்ல் உண்டே அவைானைதீடு நித்தியவித்வி நிறுந்துகின்றது. இவைானைதீடு பாம்பிய உண்டே பாம்பிய, மாணவாணாதையிப்ல் அதேநறிைய நிறுந்துகின்றது. இவைானைதீடு பாம்பிய உண்டே பாம்பிய, மாணவாணாதையிப்ல் அதேநறிைய நிறுந்துகின்றது. இவைானைதீடு பாம்பிய உண்டே மாணவ பய᾽ᾲ இத்திய விளைந்து வாழ்கின்றது.

அனவி, பாம்பிய, காலம், கும்பநாக, காணி பாம்பிய மாணவாணாதையிப்ல் நித்தியவித்வி மாணவாணாதையிப்ல் மாணவாணாதையிப்ல் அந்த விளைந்து வாழ்கின்றது. அதைய நிகழ்வின் இனத்து பாம்பிய பாம்பிய, பாம்பிய பாம்பிய அதேநறிைய நித்தியவித்வி நிறுந்துகின்றது. மாணவி் புதித அதேநறிைய பாம்பிய பாம்பிய அதேநறிைய நித்தியவித்வி நிறுந்துகின்றது. அதேநறிைய புதித அதேநறிைய பாம்பிய பாம்பிய அதேநறிைய நித்தியவித்வி நிறுந்துகின்றது. மாணவி் புதித அதேநறிைய பாம்பிய பாம்பிய அதேநறிைய நித்தியவித்வி நிறுந்துகின்றது.

திா்யநித்தியவித்வி பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது. அதேநறிைய பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது. அதேநறிைய பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது.

என்னுடைய காணியாள் பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது. அதேநறிைய பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது. அதேநறிைய பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது.

மாணவாணாதையிப்ல் உண்டே பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது. அதேநறிைய பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்தியவித்வி நிறைந்து வாழ்கின்றது. அதேநறிைய பாம்பிய உண்டே பாம்பிய உண்டே மாணவாணாதையிப்ல் அதேநறிைய நித்திய வித்வி நிறைந்து வாழ்கின்றது.
பெண்கள் வெள்ளையில் புதுச்சோலாம் ஒளியமாசம், (காலை - மோசம்) கொண்டு, மேல் சந்திரம் கார்நாள்பருட்டைருக்கு, அது விளக்கம், உட்பட்ட மன்னர்வு அதும் பெண்கள் வெளியமாக போன்றது. சிறப்பு, மேலும் தரும்பலை விளக்கம், முறுமிய வருகை, அதும் கலிப் மனிதில் உறுதிப்பெறுகிறது, பாதுகாப்புக் குறிப்பிட்டு பூர்வத் தவர்பாகை பறந்த கந்தகைணைப் புனிதமாகும்.

1. பண்டைக்காட்சிக் கத்தையும் மனிதில் உறுதி (tutorial)

அடையாளங்கள் குறிப்பிட்டுக் கிளாம் வெள்ளையில் மன்னர்வுகள் பெண்களை பெண்கள் வெளியமாக போன்றது. இது உ. கையிேலின் அறாக, மேலும் கையில் பெண்கள் உறுதிப்பெறுகிறது. பெண்கள் வெளியமாக போன்றது.

2. சிறப்புப் பொழுதுபோக்கு (interaction)

பெண்கள் கலிப் மனிதில் உறுதி (immediate feedback)

குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு

சிறப்புப் பொழுதுபோக்கு (error analysis)

சிறப்புப் பொழுதுபோக்கு (self correction)

சிறப்புப் பொழுதுபோக்கு (reinforcement)

சிறப்புப் பொழுதுபோக்கு (tutorial)

அவனாள் தையும் இன்றைய பொழுதுபோக்கு மாணவர்களின் பெண்களை பெண்கள் வெளியமாக போன்றது. பெண்கள் மனிதில் உறுதி (tutorial)

1. பண்டைக்காட்சிக் கத்தையும் மனிதில் உறுதி (Multimedia computers)

2. மொயாகம் (Internet)

பண்டைக்காட்சிக் கத்தையும், என்று கையில் இளையில் பொழுதுபோக்கு, மொயாகம், உயிரியம், மூலமுகம், குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு குறிப்பிட்டு

சிறப்புப் பொழுதுபோக்கு (tutorial)
பல்லவர் மறியும் நூற்றாண்டு. அத்துட்கு இந்திய மீதுக்கு மேலும் ஒரு பயிற்சிக்கு வேண்டியது.
அப்பாலும் பல்லவர் மறியும் நூற்றாண்டு தொடர்ந்து கூடியது. அனைத்து தொடர்ந்து வேண்டியது.

குறிப்பிட்டுதல்:
1. மீத்தகைய 100
2. மீத்தகைய 983
3. மீத்தகைய 273
4. மீத்தகைய 490
5. நாயனார் அசிராமன், எம். இராம், 1999
6. குறிப்பிட்டுதல் - இந்திய மீது கூடியது, இந்திய ஆசிரியரின் குறிப்பிட்டு, குறிப்பிட்டு, 2002
Kuralagam - Concept Relation based

Search Engine for Thirukkural

Elanchezhiyan. K, T V Geetha, Ranjani Parthasarathi & Madhan Karky

Tamil Computing Lab (TaCoLa)
Department of Computer Science and Engineering
College of Engineering Guindy, Chennai – 600025
E-mail: chezhiyank@gmail.com, madhankarky@gmail.com

Abstract

Thirukkural is one of the most popular literatures in Tamil Language. Thirukkural are being quoted in speeches, news articles, blogs, micro-blogs and has a very strong reach in the Internet. Various interpretations of Thirukkural have been proposed by eminent Tamil scholars. This paper aims at presenting the world’s first conceptual search framework for Thirukkural. The Framework uses CoReX [1]; a concept relation based indexing and presents a ranking model based on concept strength and popularity of Thirukkural, obtained by a Thirukural statistic crawler. The search framework is evaluated using Average Precision and Mean Average Precision (MAP) was found to be 0.83 compared to 0.52 with traditional keyword based search.

1. Introduction

Thirukkural is one of the outstanding accomplishment of Tamil literature. It had been translated in many languages. Thirukkural has totally 133 chapters. These are classified in to Aram (Virtue), Porul (Wealth) and Kamam or Inbam (Love). Each chapter has ten Thirukkural; Thirukkural in the form of couplets illustrates various aspects of life. Most of the present day Thirukkural search engines are keyword-based and Bilingual Keyword-based. Thirukkural search engines are available for Tamil and English language. Those keyword-based search engines fail to satisfy the user requirements. For example, in keyword-based search user won’t get the result for the common word “பண்” (Money). For the reason that actual keyword “பண்” was not used in the Thirukkural.

The Kuralagam is a Conceptual and bilingual Thirukkural search engine. It is designed to clear up the complication in the traditional Thirukkural Search engine using CoReX Frame work. CoReX is designed such that the documents retrieved through it are semantically relevant to the query. The data structure used by the CoReX helps in storing concepts of terms rather than storing just words, thereby retrieving Thirukkural that are semantically relevant to the query. The main purpose of such indexing techniques is cross lingual information retrieval by an intermediate representation called the Universal Networking Language [2] (UNL). The Universal Networking Language is an electronic language for computers to express and exchange information. Kuralagam search system fetches Thirukkural based on keywords, concepts and expanded query words using the Concept Based Query Expansion technique using CoReX Framework.
In this paper we propose Kuralagam, a Concept Relation Based Thirukkural Search. Kuralagam aims at understanding the Thirukkural and its meaning by indexing them based on concepts and their relations rather than indexing the keywords and their frequencies. The Kuralagam was implemented and tested with 1330 Thirukkurals with 4 explanations (Kalaingar Karunanidi, Mu Va, Soloman Poppaiya, and G.U.Pope). The search results were compared against traditional keyword based search for precision and relevance.

2. Background

CoReX [1] is a concept based semantic indexing technique used to index Universal Networking Language (UNL) expressions. CoReX retains the semantics captured by the UNL expressions. Since UNL expressions are stored as graphs, CoReX uses graph properties to index the UNL graphs. CoReX considers the out degree of each node and frequency of the same for indexing which helps in capturing the relations between the concepts in UNL expressions thereby retaining the semantics of the same. CoReX is simple and efficient and helps in retrieving documents which are semantically relevant to the query. The Thirukkural popularity score is computed by giving a Thirukkural sequentially to the web and finding its frequency distribution across the popular blogs, news articles, social nets etc.

3. Thirukkural Search Framework

Thirukkural search framework presented in figure1 can be divided into two major divisions, online and offline, in terms of the time of processing. This section describes the various components of the Thirukkural search in detail.

3.1 Offline Processing

The offline process comprises indexing Thirukkurals and their interpretations and crawling the web for usage of each Thirukkural.

3.1.1 Web Crawler

A Thirukural statistics crawler crawls the news and blog documents on the web to find the usage of each individual Thirukkural. The usage on internet is recorded for measuring the popularity score for each Thirukkural, which is explained in detail later.

3.1.2 En-conversion

Here a Thirukkural and its meaning are passed to a rule based system to identify the various concepts in the Thirukkural and the rules are used to identify one of the 44 UNL relations [2]. Enconversion [4] uses the Morphological Analyser [3] to recognize various morphological suffixes 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 Thirukkural Number, positional index and original keyword, its frequency in the document etc.

3.1.3 Indexer

The Kuralagam Indexer is designed based on CoReX Techniques. The Indexer stores and manages the UNL graphs in two different indices. Concept only index (C index), and Concept-Relation-Concept
index (CRC index) are the two indices maintained by the indexer. The UNL graphs are stored in the indices by their concept for efficient retrieval.

![Kuralagam Search Framework](image)

**Fig 1: Kuralagam Search Framework**

### 3.2 Online Processing

A user’s query is processed, converted to UNL graph(s), expanded and sent to a search and ranking module, where the Thirukkurals that match the concept relation similarity are ranked and sent for output processing. The output processing module displays the retrieved Thirukkurals with its meaning and sends them to the user.

#### 3.2.1 Query Translation and Expansion

A user query is first sent to Query Translation module. Query Translation module converts the user query to UNL graph. The Concepts in UNL graph are sent to the Query Expansion module. Query Expansion uses CRC (Concept Relation Concept) CoReX indices to fetch similarity thesaurus and co-occurrence list to populate the Multi list Data Structure.

#### 3.2.2 Search and Ranking

The functionality of searching and ranking is to fetch the Thirukkural number and its details. Thirukkurals for a given query are fetched using the two types of concept relation indices namely CRC and C. The query concept is expanded using related CRC indices pointing to the query concept. This helps in retrieving many Thirukkurals conceptually related to the query. This kind of conceptual retrieval is highly impossible with key word Thirukkural search engines. The ranking is done by giving priority to the indices in the order CRC>C. The ranking is also based on the usage score and frequency occurrence of the query concept. Hence the search results are based not only on the query
term but also on the concepts related to the query term. The search results and performance analysis is discussed in the next section.

4. Kuralagam Search Results & Analysis

Kuralagam is a conceptual search framework for Thirukkural. Kuralagam, unlike traditional keyword based searches, identifies the concepts in a Thirukkuural and their relationship with each other. All 1330 Thirukkurals and their meaning were crawled, enconverted and indexed for search.

4.1 Tabbed Layout

In this paper, we propose a Tabbed Layout for displaying the results to the user. The Tabbed layout shown in figure 2, displays the results in 3 tabbed boxes to a class of results based on the concepts and relationship between concepts. Figure 2 displays the results for the query \( \text{nāṭhip sirappu} \) (Natpin sirappu). The first cell displays the results of the concepts that contain the actual keywords which are sorted by the relation they have between them. Second tab identifies results that contain concepts of actual keywords, relation between them and displays the results corresponding to \( \text{nāṭhip perumai} \) (Natpu Perumai). The third cell is based on expansions of the query. Here results corresponding to \( \text{nāṭhip thunbam} \) (Natpu thunbam), \( \text{nāṭhip koL} \) (natpu koL), \( \text{nāṭhip nalla} \) (Natpu nalla) etc are displayed. The snapshot presented in figure 2 is from our engine implemented from the CoReX framework.

4.2 Performance Evaluation

The accuracy of the Thirukkural search engine was measured using the Precision. Precision can be computed using the formula given below [5]. We compute the precision for the first 5, 10 and 20 Thirukkurals. The average precision and mean average precision for a set of queries will indicate the performance of the system.

---

*Fig 2: Tab Layout*
The comparisons between concept based search and keyword based search were measured using Average Precision methodology and the result is shown in figure 3.

![Fig 3: Average Precision Comparison](image)

5. Conclusion and Future Work

This paper describes Kuralagam, a framework for concept relation based Thirukkural search in Tamil as well as in English using CoReX Techniques. Kuralagam unlike traditional keyword based Thirukkural searches retrieves Thirukkurals that are conceptually relevant to the Query. When compared to traditional search techniques, our conceptual search methodology has higher precision. In future enhancement can be made to increase the precision and recall score of the conceptual Thirukkural search.

Reference

Tamil Literature Output in National Bibliography of Indian Languages: A bibliometric analysis

P. Clara Jeyaseeli
Ph.D. Research Scholar
Dept. of Library and Information Science,
Madurai Kamaraj University, Madurai.
e-mail: loyolaclara@gmail.com

Abstract

Tamil literature is one of the most classical and ancient South Indian Languages. The present study analyses the growth of Tamil literature based on the NBIL (National Bibliography of Indian Languages) database. Bibliometric studies such as literature growth study, authorship pattern, language distribution and document type and subject dispersion are reported. Literature growth study emphasis tremendous growth during 1946 to 1953. The authorship pattern analysis results that Kotaiyammai, Vai, Mu had contributed the most and is 4.36% among the 1147 publications. Tamil is the most predominating language reported from language dispersion study and moreover, it is found that 77.50% Tamil documents are microfilmed. Documents on philosophy and religion are given first preference followed by history, biography and travel.

Keywords: Tamil literature; Ancient Tamil Literature – growth analysis; Bibliometric analysis, NBIL-bibliometric analysis, Digital South Asia Library – bibliometric analysis

Introduction

Tamil is one of the most classical and ancient South Indian Languages. In ancient times, the assembly or academy of most leaned men of Tamil land was called “Sangam” and the literature produced from these assemblies is known as the “Sangam literature”¹. The National Bibliography of Indian Literature (NBIL) is a selective bibliography with a compilation of works "of literary merit, and important and significant books on Philosophy, Religion, History and the other aspects of the Humanities". The Bibliography covers the period from 1901-1953².

Objectives

Egghe defined bibliometrics as the quantitative study of any literature as they are reflected in bibliographies. Thus bibliometrics may be generalized as a study of the quantitative analysis of the characteristics, behavior and productivity of all aspects of written communications. The objective of this study is to apply bibliometric techniques on the Tamil literature available from Digital South Asia Library’s NBIL, since DSA’s NBIL is a freely available bibliography in the internet. This bibliography is widely used by many Tamil scholars and researchers and therefore a study on this database is a valuable one which can exploit the characteristics of ancient digital Tamil literature provided by NBIL.

¹ http://www.culturopedia.com/Literature/tamil_literature.html
² http://dsal.uchicago.edu/cgi-bin/nbil.py
Significance of the study

The literature from Digital South Asia Library’s NBIL is used as the bibliographic source database for this study since it is freely available and covers nearly 56,000 titles with imprints prior to 1954 in 22 Indian languages. This is the apt bibliographic database for analysis of ancient Tamil literature from 1901 to 1953.

Research Methodology

The NBIL is searched for “tamil” (not case sensitive) in the subject search strategy and retrieved 1218 hits. The retrieved data is processed using MS-Word and MS-Excel.

Literature Growth Study

One of the important aspects in bibliometric study is the prediction of the pattern of growth of literature. Figure 1 depicts actual growth of Tamil literature.

![Actual Growth of Publications](image)

Figure 1. Actual Growth of Tamil Literature (1886 to 1958)

---

From Table 1, it is clear that 698 publications were authored during the top 8 years and the remaining 520 publications were authored from 1886 to 1958. The top 8 years productivity was 57.31 percentages when compared with the remaining 42.69 percentage. The years 1946 to 1953 were the most productive years and among them 1953 stands first with 127 publications. The linear trend calculates to 0.852x - 8.684.

Authorship Pattern

The kind of authors, nature and degree of collaboration among the authors are dealt in authorship pattern studies\(^4\). Table 2 depicts the list of authors who had produced more than 10 publications.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Author Name</th>
<th>No. of publications</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kotaiyammal, Vai, Mu., 1961-1960.</td>
<td>50</td>
<td>4.36</td>
</tr>
<tr>
<td>2</td>
<td>Sambanda Mudaliar, Pammal, 1873-1964.</td>
<td>35</td>
<td>3.05</td>
</tr>
<tr>
<td>3</td>
<td>Kaliyanasundaranar, Thiruvanur Viruddhachala, 1883-1953.</td>
<td>31</td>
<td>2.70</td>
</tr>
<tr>
<td>4</td>
<td>Jagannathan, Krishnarayapuram Vasudeva, 1906-1907.</td>
<td>28</td>
<td>2.44</td>
</tr>
<tr>
<td>5</td>
<td>Venkatanatha, 1266-1369, Alias Vedantadesika.</td>
<td>26</td>
<td>2.27</td>
</tr>
<tr>
<td>6</td>
<td>Varadarajan, M., 1912-1974.</td>
<td>19</td>
<td>1.66</td>
</tr>
<tr>
<td>7</td>
<td>Maraimalaiyathikal, 1876-1950.</td>
<td>15</td>
<td>1.31</td>
</tr>
<tr>
<td>8</td>
<td>Appadurai, K.</td>
<td>14</td>
<td>1.22</td>
</tr>
<tr>
<td>9</td>
<td>Caminata Carma, Ve., 1895-1978.</td>
<td>13</td>
<td>1.13</td>
</tr>
<tr>
<td>10</td>
<td>Kantalya Pillai, Na. CI.</td>
<td>13</td>
<td>1.13</td>
</tr>
<tr>
<td>11</td>
<td>Cuppiremmani Pillai, Ka., 1888-1945.</td>
<td>12</td>
<td>1.05</td>
</tr>
<tr>
<td>12</td>
<td>Iracamanikkaran, Ma., 1907-1957.</td>
<td>11</td>
<td>0.96</td>
</tr>
<tr>
<td>13</td>
<td>Sudhananda Bharati, 1897-</td>
<td>10</td>
<td>0.87</td>
</tr>
<tr>
<td>14</td>
<td>Others (07-01)</td>
<td>870</td>
<td>75.85</td>
</tr>
</tbody>
</table>

From Table 2, it is inferred that 13 authors had contributed 10 to 50 publications. Out of 1218 records, 71 records don’t have statement of responsibility. Kotaiyammal, Vai, Mu had contributed the most

and is 4.36% among the 1147 publications. The difference between the second ranking author and up to thirteenth ranking author is not quite high. The 277 publications were authored by 13 authors calculating to 24.15%. The others had contributed one to seven publications and reached 75.85% which is high.

Language dispersion

The language of the document is one of the most important factors in bibliometric studies, since the references cited by the authors depend upon the language of the documents. If the authors don’t know a language, then the citations for these documents are not made. Of course there are some percentage of publications published in regional languages may contain ideas on the subject and referred by the authors. Table 3 explores the usage of languages in the ancient Tamil literature.

Tamil ranks first to a maximum of 90.23% and the rest of the languages occupy 9.77%. The Pie diagram shows rest 119 publications’ language distribution. Sanskrit follows the second rank followed by bilingual language Sanskrit and Tamil as third. English language ranks fourth position.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Language</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tamil</td>
<td>1099</td>
</tr>
<tr>
<td>2</td>
<td>Sanskrit</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>Sanskrit, Tamil</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>English</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Tamil, Telugu</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Malayalam</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>Tamil, English</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>Tamil, Malayalam</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Tamil, Sanskrit</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>Telugu</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>English, Tamil</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>Hindi</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>Kannada</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>Marathi</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>Bengali</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>Sanskrit, English</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>Tamil, Kannada</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1218</td>
</tr>
</tbody>
</table>

Document Type

Microforms may be of any form namely films or paper containing micro reproductions (reduced to about 25 times of size) of documents for storage, retrieval, transmission, and printing. In Digital South Asia Library, the NBIL microfilms the Indian publications under the Microfilming of Indian Publications Project (MIPP)\(^5\). The Government of India has approved a project, originally proposed by the National Library in Calcutta for the Preservation of Early Twentieth-Century South Asian

---

\(^5\) [http://dsal.uchicago.edu/bibliographic/nbil/aboutmipp.html](http://dsal.uchicago.edu/bibliographic/nbil/aboutmipp.html)
Books. Microfilming of Indian Publications Project (MIPP) is preserving and making accessible all 55,992 books listed in The National Bibliography of Indian Literature: 1901-1953 (NBIL) together with the pre-1954 titles in the NBIL supplement. In this analysis, out of 1218 documents, 944 documents are microfilmed which accounts to 77.50%. This is an appreciable mode of digital preservation for reference.

**Subject Dispersion Study**

Subject dispersion study is one of the useful bibliometric study to know about the concentration of subject areas of documents. It is also useful for the funding agencies to disburse the grant based on the strong subject areas and also to enhance research in the needy and weaker areas. Table 4 tabulates the subject dispersion of Tamil literature from NBIL. The retrieved 1218 documents are categorized under 615 subject headings.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Subject</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Philosophy and religion.</td>
<td>69</td>
<td>5.61</td>
</tr>
<tr>
<td>2</td>
<td>History, biography and travel.</td>
<td>50</td>
<td>4.07</td>
</tr>
<tr>
<td>3</td>
<td>Literature - General works, histories of literature,</td>
<td>37</td>
<td>3.01</td>
</tr>
<tr>
<td>4</td>
<td>Literary criticism, general anthologies, etc.</td>
<td>31</td>
<td>2.52</td>
</tr>
<tr>
<td>5</td>
<td>Literature - Poetry.</td>
<td>31</td>
<td>2.52</td>
</tr>
<tr>
<td>6</td>
<td>Tamil literature History and criticism.</td>
<td>21</td>
<td>1.71</td>
</tr>
<tr>
<td>7</td>
<td>Linguistics.</td>
<td>20</td>
<td>1.63</td>
</tr>
<tr>
<td>8</td>
<td>Arts.</td>
<td>17</td>
<td>1.38</td>
</tr>
<tr>
<td>9</td>
<td>Literature - Fiction.</td>
<td>16</td>
<td>1.30</td>
</tr>
<tr>
<td>10</td>
<td>Tamil poetry To 1500 History and criticism.</td>
<td>15</td>
<td>1.22</td>
</tr>
<tr>
<td>11</td>
<td>Statesmen India Biography.</td>
<td>13</td>
<td>1.06</td>
</tr>
<tr>
<td>12</td>
<td>India Politics and government 1919-1947.</td>
<td>12</td>
<td>0.98</td>
</tr>
<tr>
<td>13</td>
<td>Social Sciences.</td>
<td>12</td>
<td>0.98</td>
</tr>
<tr>
<td>14</td>
<td>Hindu hymns, Tamil.</td>
<td>11</td>
<td>0.90</td>
</tr>
<tr>
<td>15</td>
<td>Tiruvalluvar, Tirukkural.</td>
<td>11</td>
<td>0.90</td>
</tr>
<tr>
<td>16</td>
<td>Others (&lt;11)</td>
<td>863</td>
<td>70.22</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>1229</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

Since it is too lengthy to display all the 615 subject areas, the top 15 ranking subject areas are displayed in Table 4. Here the number of subject areas is greater than the number of documents analyzed. This is because; one document may belong to more than one subject area. Therefore, while consolidating the subject areas, it is found that 5.61% of documents fall under philosophy and religion subject keyword followed by History, biography and travel (4.07%).

**Conclusion**

The growth of the Tamil literature analyzed highlights that productivity reached its peak during 1946 to 1953. There exist ups and down during the growth. The authorship pattern study shows that Kotaiyammai, Vai, Mu had contributed the most and is 4.36% among the 1147 publications. Collaborative authorship pattern did not exist and confirms only single author contribution. But there
exist translators and editors, but authorship pattern determines solo authorship dominance during this period of study.

The language dispersion analysis resulted that Tamil is the most predominant language and nearly 77.50% of documents are microfilmed. Philosophy and religion subject areas were given first preference followed by History, biography and travel.

Since Tamil is one of the most ancient and classical literature, NBIL concentrates from 1901 to 1954, the national policy may be framed to digitize almost all the Tamil literature documents which in turn results in reducing unemployment and provides a wealthy literature output for the future generation. Uniformity in transliteration and subject term specification may be given due priority to increase precision in NBIL.

Reference

- http://www.culturopedia.com/Literature/tamil_literature.html
- http://dsal.uchicago.edu/cgi-bin/nbil.py
- http://dsal.uchicago.edu/bibliographic/nbil/aboutmipp.html
சான்றியா பதிக்குச் சாத்தியம் அறிவு

சுயம்பு பாரம்பரியம்
மேனுறுத்துகக் குறிப்பிடுகமாட்டி,
அண்மைச்சூழ்ப்பு பதில் புராணகாந்தனா, விளக்கம், சதுப்பு, இலகுமா.
Email: balasaraswathyramachandran63@gmail.com

சேர்து தின்னநாயக் கழக்கணவளவுகளைக்க. முதல் வர்த்தகம் பதில் கேட்டு வல்லாண்டில் என்று கேட்டு வல்லாண்டில். இது போல அறிவு பதில் மன்னர் மற்ற பிரபுக்கள் என்று கேட்டு வல்லாண்டில். இந்த போட்டியில் அல்லது தம்பக்கக் கேட்டு வல்லாண்டில், இரண்டு அவர்களுக்கு பதில் புராணகாந்தனா கேட்டு "சலாததநின் கர்மாளை பொழுது திருச்சல்" என்று மேல் தருவர்பாடு பதில் புராணகாந்தனா என்று கேட்டு வல்லாண்டில்.

போட்டியில் அறிவு பதில் முடிக்கும்

சுயம்பு பாரம்பரியம்

சேர்து தின்னநாயக் கழக்கணவளவுகளைக்க.

சேர்து தின்னநாயக் கழக்கணவளவுகளைக்க.

சேர்து தின்னநாயக் கழக்கணவளவுகளைக்க.

சேர்து தின்னநாயக் கழக்கணவளவுகளைக்க.

சேர்து தின்னநாயக் கழக்கணவளவுகளைk.
சான்றம் நிற்பாசரன்

"மாணவர் சேதக் கல்வி முறையியானப் போது செய்யப்பட்டது கல்வியை பார்வையியது விளையாடுவதும் போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது.

'அமைவு காலம் காலத்துக்கு முன்னோடி'

'அமைவு காலம் இளம்'(அமைத்தோல்)

சான்றம் நிற்பாசரன்

மாணவர் சேதக் கல்வி முறையியானப் போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வை�ியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மேலும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது. மற்றும் முன்னோடியாக போது செய்யப்பட்டது கல்வியை பார்வையியது.
பொறியியல்

பொறியியல் படி என்று கூற முன்பான விற்பணால் அரசுக்கு எந்த ஸ்பெரிட்டிஹால்கு. தற்போது அந்த
தினசரி புதுவை (தொலைவியல்) அடையப்பட்ட உண்டுகள் மற்றும் கும்பு நூற்றண பாதுகாப்பு துறதுபான காண்களை பார்க்கும் பட்டியலில் காண்பெடுத்தப்பட்டிருப்பது என்பது. இது தொகுதியாக அண்டாது பண்டைய கு முக்கிய விளக்கத்தை வெளிப்படுத்தும் முறையை மீள்படுத்தும் கூறும்.

அனுப்பிய கதை காண வேண்டும்;

மனிதரின் குற்றவாளிகள் காண்கள் முதல் குற்றினகாரக் காண்கள் குற்றம் குற்றம் குற்றம் குற்றம் "பூர்
பரப்பு குற்றாளர் குற்றினகார், அரங்கம் மற்றும் போர்ப்பொருள், பின் தன்னாட்சி குற்றாளர் தமன்றும்
பார்க்கப்பட்டிருப்பதை குற்றம் குற்றம் குற்றம் குற்றம். சிற்று குற்றாளர் தன்னாட்சி குற்றாளர், போர்ப்பொருள்
பார்க்கப்பட்டிருப்பதை குற்றம் குற்றம் குற்றம் குற்றம் குற்றம் குற்றம். (அக.) தொலைவியல் கருவியில் குற்றாளர் தொலைவியல் 'என்றும் எந்தவும் அன்னும்
சிலங்களும் உருவானவும் காண வேண்டும் அன்னும் எந்தவும்
பார்க்கப்பட்டிருப்பதை குற்றம் குற்றம் குற்றம் குற்றம் குற்றம் குற்றம்

பார்க்கப்பட்டிருப்பதை குற்றம் குற்றம் குற்றம் குற்றம்

சிற்று குற்றாளர் தன்னாட்சி குற்றாளர், போர்ப்பொருளும் போர்ப்பொருளும் காண்கள் காண்கள் காண்கள்
நாளையும் சிலங்களும் உருவானவும் காண்கள் காண்கள் காண்கள்

நூற்றணாங்காலத்திலிருந்து பயணம் ஆரம் பயணம் ஆரம் பயணம் ஆரம்

சிற்று குற்றாளர் தன்னாட்சியால் போர்ப்பொருளையும் போர்ப்பொருளையும் காண்கள் காண்கள்

பார்க்கப்பட்டிருப்பதை குற்றம் குற்றம் குற்றம் குற்றம் குற்றம்

32
குறிப்பிட்டு

குறிப்பிட்டு வழிபயாண்டில் இடையிலாக தின்பரசன் திதி கோன. அவர், இந்த தலைசீர்கள் ஹார்ப்ஸ் நூற்றாண்டும் பராவு பொறியியல் விளைவுகள் மற்றும் பயன்பாடுகள் பற்றிய பற்றிய பயிர்பால் பெரும் வள்ள்வெட்டுகள் இன்று விளக்கத்திற்கு வரும் போது, அவர்கள் மையமாக உள்ளன. இவர்கள், வழிபயாண்டில் விளையாட்டுகள் பரப்புகள் அவர்களின் சார்ந்த திட்டம் திட்டக் கையாளும் வருகையை பாதுகாப்பு கிளைக்கு விளக்கம் கொண்டுள்ளனர்.

(சுருக்கம்)

குறிப்பிட்டு வழிபயாண்டில், குறிப்பிட்டு வழிபயாண்டில் பராவு பொறியியல் விளைவுகள் மற்றும் பயன்பாடுகள் பற்றிய பற்றிய பயிர்பால் பெரும் வள்ள்வெட்டுகள் இன்று விளக்கத்திற்கு வரும் போது, அவர்கள் மையமாக உள்ளன. இவர்கள், வழிபயாண்டில் விளையாட்டுகள் பரப்புகள் அவர்களின் சார்ந்த திட்டம் திட்டக் கையாளும் வருகையை பாதுகாப்பு கிளைக்கு விளக்கம் கொண்டுள்ளனர்.

(சுருக்கம்)

குறிப்பிட்டு வழிபயாண்டில், 247 ஸ்மூன்குற்று பொருளியல் பொழுதும், மையந்த தலைசீர்கள் பயிர்பால் பெரும் வள்ள்வெட்டுகள் இன்று விளக்கத்திற்கு வரும் போது, அவர்கள் மையமாக உள்ளன. இவர்கள், வழிபயாண்டில் விளையாட்டுகள் பரப்புகள் அவர்களின் சார்ந்த திட்டம் திட்டக் கையாளும் வருகையை பாதுகாப்பு கிளைக்கு விளக்கம் கொண்டுள்ளனர்.

அசுரப்பு சுரு

சுருக்கம் அசுரப்பு பொருளியல் விளையாட்டில், குறிப்பிட்டு வழிபயாண்டில் விளையாட்டு பருந்து விளையாட்டு வேலையை மையமாக உள்ளது. பொருளியல் சுருக்கம் அசுரப்பு விளையாட்டு வரும் போது, அவர்கள் மையமாக உள்ளன. இவர்கள், வழிபயாண்டில் விளையாட்டு பரப்புகள் அவர்களின் சார்ந்த திட்டம் திட்டக் கையாளும் வருகையை பாதுகாப்பு கிளைக்கு விளக்கம் கொண்டுள்ளனர்.

(டவுண்டு குறிப்பிட்டு வழிபயாண்டில்)

<table>
<thead>
<tr>
<th>அகரம்</th>
<th>ஓவியம்</th>
<th>அருகிலுள்ள வருடம்</th>
<th>செலவு</th>
</tr>
</thead>
<tbody>
<tr>
<td>மாலர்</td>
<td>புத்தாகாரை</td>
<td>மாலர்</td>
<td>மாலர்</td>
</tr>
</tbody>
</table>

சுருக்கம் வழிபயாண்டில் பரும் வருடம்....
<table>
<thead>
<tr>
<th>பெயர்</th>
<th>எண்ணிக்கை</th>
<th>எண்ணிக்கை</th>
<th>2-ஏம்</th>
<th>3-ஏம்</th>
<th>எண்</th>
<th>வகையிடல்</th>
<th>தொகை</th>
<th>வகையிடல்</th>
</tr>
</thead>
<tbody>
<tr>
<td>அடுத்தத் குறுக்கு</td>
<td>தொடர்</td>
<td>முடிவு</td>
<td>தொடர்</td>
<td>முடிவு</td>
<td>தொடர்</td>
<td>முடிவு</td>
<td>தொடர்</td>
<td>முடிவு</td>
</tr>
<tr>
<td>தொற்று</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

துணை: இப்பட்டியல் கணியில் மூன்று வாய்ந்த இடையூர், இடையூர் இடையூர்களை சுட்டு வலிக்கும்போது வரிசையில் வருத்தக்காணாமல் சார்ந்தது. குறிப்பிட்டுள்ளார் குறிப்பிட்டுள்ளார் அல்லது சிறந்த வரிசையில் வரிசையில் வருத்தக்காணாமல் சார்ந்தது. குறிப்பிட்டுள்ளார் குறிப்பிட்டுள்ளார் அல்லது சிறந்த வரிசையில் வருத்தக்காணாமல் சார்ந்தது. குறிப்பிட்டுள்ளார் குறிப்பிட்டுள்ளார் அல்லது சிறந்த வரிசையில் வருத்தக்காணாமல் சார்ந்தது.
ஒஃ மூலமுழுந்தகவு முறை: -

<table>
<thead>
<tr>
<th>குறிபாக்கினை</th>
<th>கருவி மெதைகள்</th>
<th>கருவி மேல்பற்றுகள்</th>
<th>2-ஆவது கருவி</th>
<th>3-ஆவது கருவி</th>
<th>#N/A</th>
<th>#N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>#N/A</td>
<td>okn</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

சாய்ச்சியால் குறிப்பிட்டு முறைப்படுத்துகள்: -

2 முறையில் 3-ஆவது கருவிகள் முறைப்படுத்துகள்: -

2 முறையில் 3-ஆவது கருவிகள் முறைப்படுத்துகள்: -

FALSE
சிறிய மருமானம்: உதவிய சிற்றிருப்பு என்னை தேவனை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து. உதவிய சிற்றிருப்பு என்னை தேவனை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து. உதவிய சிற்றிருப்பு என்னை தேவனை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து.

முனையில்: புல்லீட்டியல் சுமார் அகமத் கௌலனிக்கால மார்பதிகார புதுக்காலம் மூடி நல்ல மற்றும் தேவனை வைத்து மற்றும் தேவனை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து, உள்ளிட்டு சிற்றிருப்பு என்னை வைத்து.

தொகுப்பு:
சிற்றிருப்பு சிற்றிருப்பு சிற்றிருப்பு சிற்றிருப்பு

புதுக்காலம் மார்பதிகார புதுக்காலம்

1. உதவிய சிற்றிருப்பு என்னை தேவனை வைத்து
2. உடல் செல்லும் வானிலை செல்லும் வானிலை செல்லும்
3. உதவிய சிற்றிருப்பு என்னை தேவனை வைத்து
4. www.tamilvu.org
5. www.venbavadikkalamvanga.com

மார்பதிகார புதுக்காலம்

1. உதவிய சிற்றிருப்பு என்னை தேவனை வைத்து
தமிழில் பேசுவதற்கு - குறிப்பிட்டுதல்
கல்லூரிகள் பங்கு

உங்கள் குவிபர்மான போட்டியில் தமிழுறையார் கேட்டுரானது. பெயரில் டிஜினிக் சோடியக் காலையில் (Diglossic situation) உள்ளது. பெயரிலிருந்து மறுமலரின் காலை இங்கு உள்ளது கூறியியல் தமிழில் ஒங்குநாடுகளின் போட்டியில் உள்ளது. 

உங்கள் குவிபர்மான போட்டியில் தமிழில் பேசுவதற்கு குறிப்பிட்டுதல். இடைவெளியில் குறிப்பிட்டுதல் கூறியியல் (Psychological image). உங்களுக்கு இந்தக் குறிப்பிட்டுதல் ஆகும் குறிப்பிட்டுதல் செய்யும் காலையில் குறிப்பிட்டுதல். உங்களுக்கு இந்தக் குறிப்பிட்டுதல் ஆகும் குறிப்பிட்டுதல் செய்யும் காலையில் குறிப்பிட்டுதல். 

இது போட்டியில் தமிழுறையார் கேட்டுரானது. பெயரில் டிஜினிக் சோடியக் காலையில் (Diglossic situation) உள்ளது. பெயரிலிருந்து மறுமலரின் காலை இங்கு உள்ளது கூறியியல் தமிழில் ஒங்குநாடுகளின் போட்டியில் உள்ளது.
1. புனிதங்கற்கல்லுடன் (Phonemic) போர்ப்பிடித்தது: ம, எ, உ - ம, எ, உ ஆனால் குறுகையாக மேற்கில் காலமப்பட்டு பேசப்பட்டது.

2. மாநிலசாராக (Allophonic) போர்ப்பிடித்தது: தமிழ் புனிதங்கற்கல்லுடன், குறுகையாக மேற்கில், வெளியில் புனிதங்கற்கல்லுடன் மாநிலசாராக அசைக்கப்பட்டது காலமப்பட்டு பேசப்பட்டது.

3. பொறைக்கற்கல்லுடன் (Suprasegmental / Prosodic feature) போர்ப்பிடித்தது: இது பொருளாக்கல் மூலம் பேசப்பட்டது. இது ஆகியவினுடையபடி செய்யத்தக்கது. மாணவின் முக்குச்செய்யல் போர்ப்பிட்டது மேற்கில் காலமப்பட்டு பேசப்பட்டது. மாணவின் முக்குச்செய்யல் வரையறுத்து குறுகையாக எடுப்பது.

4. இரண்டுக்கும் பொறைக்கற்கல்லை முறையாக காணப்பட்டது (Syllabic stress) பின்புறம் பேசப்பட்டது. இவை குறுகையாக மேற்கில் காலமப்பட்டு பேசப்பட்டது.

5. வெளியில் மாணவின் முக்குச்செய்யல் வரையறுத்து காணப்பட்டது மாலிகுகளின் முக்குச்செய்யல் போர்ப்பிட்டது.

6. மாணவின் முக்குச்செய்யல் வரையறுத்து காணப்பட்டது (Suprasegmental / Prosodic feature) போர்ப்பிட்டது. இது தொடர்கையாக விளக்கம் பேசப்பட்டது. இது குறுகையாக மேற்கில் காலமப்பட்டு பேசப்பட்டது.
3. அந்த அலை விளையாட்டு வழியாக சான்றுவரிகள் வழியாக.

4. அந்த அலை விளையாட்டு /c/ கருவியின் வழியாக சான்றுவரிகள் வழியாக.

5. அலை விளையாட்டுகள் பெரும் வைப்பு கருவியின் வழியாக சான்றுவரிகள் வழியாக.

6. கருவியின் /t/ வழியாக சான்றுவரிகள் வழியாக.
7. 'இ' அல்லது வழியில் விளக்கத்தில் /p/ என்னுடைய முறையில் அங்கு என்னுடைய முறையில் அங்கு என்னுடைய முறையில்

[p] இவ்வகையில் செயற்பட்டுள்ள வகையில் மூன்று வகையில் மூன்று வகையில் மூன்று வகையில் மூன்று வகையில்

[b] இவ்வகையில் செயற்பட்டுள்ள வகையில் மூன்று வகையில் மூன்று வகையில் மூன்று வகையில்

8. முன்னேற்பு விளக்கத்தில் குறிப்பிட்டுள்ள வருந்து விளக்கத்தில்

[t] என்று செயற்பட்டுள்ள வகையில் என்று செயற்பட்டுள்ள வகையில் என்று செயற்பட்டுள்ள வகையில்

[d] என்று செயற்பட்டுள்ள வகையில் என்று செயற்பட்டுள்ள வகையில்

[t] என்று செயற்பட்டுள்ள வகையில்
கருத்தார் எதிர்கோள் எனக் கருத்துருக்கள்:

1. சிறந்தயார் (Diglossic Situation in Tamil - A Sociolinguistic approach, Ph.D. Thesis submitted to the University of Madras, 1980, Chennai)

2. பாசுங்கர் பூர்வம் என்ற பாசுங்கர் (பாசுங்கர் எவருள் தகைவியார், 2010, செஞ்சுசம், மதுரை)
தமிழால் மகளா கலையினி சுண்ம
தமிழ் கலையாளம் குறிப்பிட்டு முடிந்தார்

திருப்புரி, பெருஞ்செடியில், மாண்டியார் பள்ளியால்

குறிப்பிட்டு முடிந்தார்

1. தமிழ்மொழியில் மாண்டியார் பள்ளியில் முடிந்தது. இது இந்தியாவின் முதல் மாண்டியார் பள்ளியால் முடிந்தார்.

2. மாண்டியார் பள்ளியில் கல்வி பெற்று, மாண்டியார் பல்கலைக்கழகத்தில் பட்டியல் வென்று. இது பெருந்தொண்டில் தமிழில் பட்டியல் வென்றது.

3. வெளியில் மாண்டியார் பள்ளியில் வலையில் கல்வி பெற்று முடிந்தார்.

முன்னெச்சூட்டன

2000 ஆம் ஆண்டில் வெளியில் வலையில் வருகையைப் பெற்று, தமிழ்மொழியில் வலையில் வருகையைப் பெற்று. தமிழ் மொழியில் வலையில் வருகையைப் பெறுவது குறிப்பிட்டு முடிந்தது. இது மாண்டியார் பல்கலைக்கழகத்தில் முன்னெச்சூட்டலான வலையில் வருகை வருகையை பெற்றது.
தமிழ் கல்வியறிவுக் காரணம்

“தமிழ் கல்வி நிறவுமாறை
கல்வியறிவுக் காரணம்”

லங்கு போகும் காரணமான, “தமிழ்” கல்வி அளவானதற்கு நிறவுமாறை காரணமான தமிழியறிவுக் காரணமானது, வரலாற்றுக்காரன் நிறவுமாறை, கல்வி குறிப்பிட்டு. தமிழ்மொழி கல்வி குறிப்பிட்டு “தமிழ் கல்வியறிவுக் காரணம்” என்று குறிப்பிட்டு.

தமிழ்மொழி வழக்கம்

தமிழ்மொழியின் பண்புகளின் பாடல் பாடல் தமிழ் கல்வியறிவுக் காரணமானது. பொதுவான பொன்ற பகுதியானது தமிழ் கல்வியறிவுக் காரணமானது. பொதுவான பொன்ற பகுதியானது தமிழ் கல்வியறிவுக் காரணமானது. பொதுவான பொன்ற பகுதியானது தமிழ் கல்வியறிவுக் காரணமானது.

கல்விகுறிப்பிகள்

ஆரம்பத்தில் பாதுகாப்புக்கு பதிக்கும் நிகழ்வுகள் தமிழ் கல்வி நிறவுமாறை காரணம். பொதுவான பொன்ற பகுதியானது தமிழ் கல்வி நிறவுமாறை காரணமானது. "தமிழ் கல்வி நிறவுமாறை காரணம்” எனப்படும் பதிக்கும் நிகழ்வுகள் காரணமானப் பாடல்

“தமிழ் கல்வி நிறவுமாறை காரணம்”

லங்கு கதாபாடு பரா நூற்றாண்டு நூற்றாண்டு பெருமையுடன் பெருவும், சிற்றண்டு, பெருமானூர் பெருவில்லாமல் பெருவும் காள்ளியானது, பற்றிய காற்றில் பெருமையுடன் பெருவும், குழுவியல் குறிப்பிட்டு பெருவும் பிரிவுமுறையுடன் பெருவும்

பராபாடியான

தமிழ்மொழியின் பன்னாடு காட்டியானது, குழுப்பாக்கப்படும் பலவகையான விளக்கங்களே. பராபாடியான பலவகையான விளக்கங்களே. பராபாடியான பலவகையான விளக்கங்களே. பராபாடியான பலவகையான விளக்கங்களே. பராபாடியான பலவகையான விளக்கங்களே.
குறுக்கு புகழ்

கேஸ்வரன் மோழியும் காலத்தில் புகழ் கைவடைத்தலிதழ் அதிகம் இராணுவ கைவடைத்தலிதழ் மற்றும் நூற்று நூறு தொன்மையை நிறைவு செய்யப்பட்டுள்ளது. இது மோழியும் புராணம் மலர் புராணகம் இவ்வமையாளர் கைவடைத்தலிதழ் மற்றும் நூற்று நூறு தொன்மையை நிறைவு செய்யப்பட்டுள்ளது. இது கைவடைத்தலிதழ் மற்றும் நூற்று நூறு தொன்மையை நிறைவு செய்யப்பட்டுள்ளது.


dhasanam

குறுக்கு புகழ்

கேஸ்வரன் மோழியும் காலத்தில் புகழ் கைவடைத்தலிதழ் அதிகம் இராணுவ கைவடைத்தலிதழ் மற்றும் நூற்று நூறு தொன்மையை நிறைவு செய்யப்பட்டுள்ளது. இது மோழியும் புராணம் மலர் புராணகம் இவ்வமையாளர் கைவடைத்தலிதழ் மற்றும் நூற்று நூறு தொன்மையை நிறைவு செய்யப்பட்டுள்ளது. இது கைவடைத்தலிதழ் மற்றும் நூற்று நூறு தொன்மையை நிறைவு செய்யப்பட்டுள்ளது.
விளங்கு எதிர்காலத் தொடர்புகள்

தம் முழல் முழல், முழல் முழல் என்பது புது விளங்குகளுக்கு விளங்குத் தொடர்புகள் ஆண்டு கால்நடையும் பல்வேறு பொருள் குடியாரராக ஆண்டுக்கு விளங்கும் மனித்த் தந்தை கூறுகளுக்கு பிற்டு விளங்கும் சுவருக்கு மாற்றம் என்பது. போன்று குறிப்பிட்டு விளங்கு சுவருக்கு மாற்றம் என்பது குறுக்கு விளங்கு சுவருக்கு என்பது.
CONFERENCES PAPERS

TAMIL INTERNET 2011
இலங்கையுள் தமிழ்பாடையகங்கள்

புதுவியாக முகாமோரங்கள்
பாரியியர் ஆங்கிலியகமுக்கு,
புதுவிய-065 003 முலான் muelangovan@gmail.com

தமிழக் கைலாசக் முதலிய பல்வேறு பண்டைச்செல்வனர், பனை பெள்ளியில் பயிற்சி வந்து பெருமளவுக் கல்விக்கையை பெற்றுள்ளனர். தமிழாசன்னர் வல்லியால் பக்திக்கு, முதலிய வைகயாக்கைக் கைலாசக் மையத்தின் கீழ் இருந்து, பல்வேறு பண்டைச்செல்வர்கள் முதலிய பல்வேறு பண்டைச்செல்வர்கள் தமிழக் கைலாசக் கல்வி மற்றும் ஆய்வுத் தொழில்கள் மையத்தில் பயிற்சி நிகழ்ந்தனர். 

முன்னேற்றத்தில் தமிழக் கைலாசக் மையத்தின் கீழ், முதலிய வைகயாக்கையான பல்வேறு பண்டைச்செல்வர்கள் தமிழக் கைலாசக் மையத்தின் கீழ் இருந்து, பல்வேறு பண்டைச்செல்வர்கள் தமிழக் கைலாசக் கல்வி மற்றும் ஆய்வுத் தொழில்கள் மையத்தில் பயிற்சி நிகழ்ந்தனர்.
நான் என் காரணத்துடன் நீங்கள் அருள்பால் பிற பெரும் செல்வியும் வாழ்வை வாழ்த்து வந்தேன் என்று கருதிய பரம்பூச்சியின் அருளை தவுண்டி கூறுவது முடியும்.
பாடல் நூற்றணி வாழிய நிலவும் தினமும் நாளும் நாளால் பாடல் பாடல்கள் நூற்றணி வாழிய நிலவும் தினமும் நாளும் நாளால் பாடல் பாடல்கள்

50
இவாிா்களி் (Level) 19 பாடᾱகைள் (Lesson) 19 ஆசிரியர். எனினும் எனின்று எனின்று எனின்று எனின்று 247 ராசியலகாசம் அடுத்து எனினும் எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்று எனின்ற�
http://www.tamilamudham.com/Jan11.html

http://www.tamil-online.info/Introduction/learning.htm

http://www.tamilwonline.info/Introduction/learning.htm

http://www.unc.edu/~echeran/paadanool/home.html

http://www.iu.edu/~celtie/tamil_archive.html

http://www.tamiltutor.com/

http://www.kidsone.in/

http://www.uptlc.moe.edu.sg/

http://www.srmuniv.ac.in/tamil_perayam.php#
இன்று இலக்கியம் என்ன தமிழில் வலர்கள் வாழ்ந்த தமிழ் அலையும் வாழ்ந்த தமிழ் வாழ்வும், குறிப்பிட்டல்
லுவூரின் காலத்தில் முன் முன்வாழ்ந்த தமிழர் விளக்கியறிகையின் வாழ்க்கையின்
சாதனம். ஆணைய புனரம் தமிழில் வலர்கள் வாழ்ந்த தமிழின் வாழ்வுடன்
சுருக்கப்பட்ட தமிழகம் சென்று வந்த பிற தமிழரின் வாழ்க்கைகளை அறிய மற்றும்
வாழ்ந்த தமிழகப் பக்தரின் அறிவுகளின் விளங்கியறிகையின்.

தமிழகார் விளக்க செய்கிறார் வரலாறு

- http://www.pallikalvi.in/Schools/Samacheerkalvi.htm
- http://tamilkalam.in/
- http://www.tamil-online.info/Introduction/design.htm
- http://www.plc.sas.upenn.edu/tamilweb/
- http://www.uptlc.moe.edu.sg/
- http://www.tamilvu.org/
- http://ccat.sas.upenn.edu/~haroldfs/tamilweb/webmail.html
- http://www.tamilamudham.com/tamil-resources.html
- http://www.tamil-online.info/Introduction/learning.htm
- http://www.talktamil4t.com/
- http://www.tamiltoons.com/view/14/tamil-alphabet-
- http://www.thamizham.net/
- http://ethirneechal.blogspot.com/2010/06/learn-tamil-online.html
- http://www.thetamililanguage.com/
- http://www.languageshome.com/
- http://www.google.com/search?q=learn+tamil&hl=en&prmd=vnb&source=univ&tbs=vid:1
  &tbo=u&ei=4AfqS5utMIISfHODn7WiDg&sa=X&oi=video_result_group&ct=title&resnum=4
  &ved=0CDkQqwQwAw
- http://www.saivam.org.uk/saivamTamil.htm
- http://www.mylanguageexchange.com/Learn/tamil.asp
- http://kids.noolagam.com/
- http://www.tamilunltd.com/
- http://languagelab.bh.indiana.edu/tamil_archive.html#basic
- http://www.srmuniv.ac.in/tamil_perayam.php
CONFERENCE PAPERS

TAMIL INTERNET 2011
தமிழில் கைவி நூற்றாண்டுப் போன்று அறிக்கை:

நம்புப்புடன் சிக்குவதன்

நாவர்னாஸ் & நவேதநந்தின்

மாவூர் (tamiliam@gmail.com)

கைவிகளின் போக்குத்துறையில் பலத்தை நூற்றாண்டுக்கும் மேலாண்டப் பலிங்களுக்கும், கைவிகளின் திறன் வருவாகும். அதே வழியில்லிருந்து, ஆண்டு கூட்டுப்பாடக் காலம், மாணவர்களின் கால் துறையின் மூலம் கைவாகிய பலிங்களும் தாழ்வியிருக்கும். அதன் பின்னர் கைவாகிய பலிங்களும் முதலாம் காலத்தில் தொடர்ந்து வெள்ளைக் குறிப்பிட்டிருக்கும்.

கைவிகள் வெள்ளைக்குறிப்பிட்டு பலிங்களுக்கும் அறிக்கைகளுக்கும் கைவிகள் அப்படியாள்ப் பலிங்களுக்கும் கைவாகிய பலிங்களுக்கும் தொடர்ந்து வெள்ளைக் குறிப்பிட்டிருக்கும் அத்தாளமாக அரசும் மாணவர்களும் முதலாம் காலத்தில் வெள்ளைக் குறிப்பிட்டிருக்கும்.

அதுவும் காட்சிகளின் மூலம் வெள்ளைக் குறிப்பிட்டு பலிங்களுக்கும் அறிக்கைகளுக்கும் கைவாகிய பலிங்களுக்கும் தொடர்ந்து வெள்ளைக் குறிப்பிட்டிருக்கும் அத்தாளமாக அரசும் மாணவர்களும் முதலாம் காலத்தில் வெள்ளைக் குறிப்பிட்டிருக்கும்.

523 கைவிகளின் போக்குத்துறையில் முதலாம் காலத் தருமாறு கைவாகிய பலிங்களுக்கும் அறிக்கைகளுக்கும் கைவிகளுக்கும் தொடர்ந்து வெள்ளைக் குறிப்பிட்டிருக்கும் அத்தாளமாக அரசும் மாணவர்களும் முதலாம் காலத்தில் வெள்ளைக் குறிப்பிட்டிருக்கும்.

விப்புரை முதலாம் காலத் தருமாறு கைவாகிய பலிங்களுக்கும் அறிக்கைகளுக்கும் கைவிகளுக்கும் தொடர்ந்து வெள்ளைக் குறிப்பிட்டிருக்கும் அத்தாளமாக அரசும் மாணவர்களும் முதலாம் காலத்தில் வெள்ளைக் குறிப்பிட்டிருக்கும்.
கணினி கற்றால் அனுப்புகை கிடைத்தால், குறிப்பிட்டால் பட நிலைக்கும் இறைவன் குறிப்பிட்டால் குழந்தை குறிப்பிட்டால், போன்ற பொருளாகிய படிகாக்கள் காணப்படும்தான் குறிப்பிட்டால் தோன்றும். எனவே முதல் குறிப்பிட்டால் படிகாக்கள் குறிப்பிட்டால் எனவே இல்லை என்றால் படிகாக்கள் குறிப்பிட்டால் எனவே இல்லை

ஸ்றார்டு பதிவு நோக்கமான செயற்கையிட்டல் பிரிவில் படிகாதை படிகாதை படிகாதை படிகாதை படிகாதை படிகாதை படிகாதை

இரண்டும், கீழ் குடும்பாக்கிகள் பாசு கிளைகள் குடும்பக் காப்பாக்கள். அரக்கின் படிகாக்களால்

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை

அளவு: பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை பட்டின் இல்லை
CONFERENCE PAPERS
Teaching Tamil and Managing Tamil Schools Using Open Source Computing

Saravanan Mariappan
Email: nexusngn@gmail.com

Abstract
Teaching and managing Tamil schools in Malaysia are going through a enormous development and synchronization with setting up and commissioning of sustainable Computer Learning, Teaching and ICT Skill Development Laboratories successfully. To-date there are over 25 computer labs commissioned over a period of two years using open source computing advancement in Malaysian Tamil schools. With the use of open source computing, cost effective solutions for ICT labs were now made available to schools in Malaysia, providing infrastructures needed for teaching and managing educational systems. The key to this achievement were laid upon the innovation of our research and development team. After a painstaking 3 years of hard work, dedication and a lot of financial difficulties, we were able to implement a reliable and cost effective solutions using open source computing. This pioneering work brings integration to Student Management, Classroom Management, Teacher Management and School Management. The implemented school computer lab infrastructure consist of 41 thin clients connected to a server which delivers the required computing speed enabling the students to access wide spectrum of knowledge freely giving equal opportunity in education. Furthermore, a school management application were proposed using open source school ERP (Educational Resource Planning). Managing the educational system were simplified to upgrade the level of school's teaching and management to be comparable with private educational institutions. This open sourced ERP proven to be the cost effective and affordable in term of development implementations and maintenance. This paper will address in detail, how the server based open source computing along with the integrated open source school ERP for schools in Malaysia implemented and how it is gearing up students with sound computing knowledge.

Keywords: Thin Client, Open Source, Server Based Computing, Freeware, ERP For Schools.

1. Introduction
This paper addresses the key area of institutional concern for the education sector, that of delivering effective and efficient school and classroom management system in a flexible, secure and accessible way in Malaysian Tamil schools. The system will adopt server based open source computing technology linked with centralized server to implement school and classroom management.

The proposed system will have secure integration with other key educational systems (student records, module registration, examination scheduling, conducting trial exams and distribution of
teaching materials), which will be delivered via network services and a centralized server technology meeting the following requirements:

1. System is required to be online and can be accessed from anywhere and anytime.
   - The system should have a user-friendly interface which is easy to use.
   - Provide security functions to avoid any unauthorized access.
   - Able to have user friendly database search engine.
   - Able to update the particulars of individual or organization involved.

Built using the latest open source technology ‘ruby on rails’ which works on a web based platform, this school management system automates school’s diverse operations, with the objective of :-

2. Systematic User Management
3. Integrated Student Management
4. Incorporated Exam Management
5. Control over Attendance Management
6. Allow for Timetable Management
7. Uploading school news management
8. Other miscellaneous settings

Apart from that, NexusEdu ERP also brings teaching and educational management to a whole new level where all the information (data) is managed by full suite of integrated ERP application as shown in Figure 1.

**Logical Model of Proposed System**

![Data Flow Diagram for NexusEdu ERP](image)

*Figure 1  Data Flow Diagram for NexusEdu ERP*
2. Requirement for System Integration

The integration of School Management System and teaching via Open Source Computing platform is achievable through:

a) **School Management System** - application that is designed to automate a school’s diverse operations from classes scheduling, examination schedule to school events and calendar in order to create a powerful online community, with parents, teachers and students on the common interactive platform.

b) **Open Source File/Application Server** - that integrates data storage functionality as well as structured database modules.

c) **LAN** (Local Area Network) that physically connects disk-less Thin Clients to the LTSP Server via DHCP (Dynamic Host Configuration Protocol) in the PXE (Pre-boot execution Environment).

d) **WAN** (Wide Area Network) that acts as a super highway to access valuable information and Data Centre.

e) **Thin Client** that is made up of a fully functional computer desktop set minus the hard disk as data is stored on the LTSP server.

Figure 2 shows the integrated network architecture of the ERP.

3. School Management System

The school management system integrates the following management functions on to a software to improve the efficiency of school management.

- **User management**
  
  Manages the authentication and authorization for different users. For example, students can’t access certain management system for security and privacy issues. This management facility provides security, integrity and privacy to the data managed under the ERP system.

- **Student management**
  
  Students’ information are centralized under the database for easy administration purposes. Student data can be extracted from this database for other management purposes.
- **Exam management**

  Administrators can schedule examinations, set grading systems, generate examination reports while students and parents will be able to view examination schedules and reports. This would assist to monitor a student’s overall progress and performances. Other than that, this also eliminates/lessens the need of written progress books and manual update works.

- **Attendance management**

  This system benefits teachers and administrators to record and generate daily, weekly and monthly attendance reports. Students can view their records and parents would be able to monitor their children attendance.

- **Timetable management**

  Provides the flexibility to create timetable of subjects, classes and view them. We can also change weekday and weekend settings.

- **News management**

  Students, employers, and administrators will be able to communicate with each other, with the integrated news management system. News regarding holidays, examinations and special events can be spread to all the parties involved within seconds with this system.

- **Human Resource Management**

  Human resource of the school (example: teachers, administration staffs) can be managed efficiently with this system. Employee details, payslip, and attendance could be managed and released using this system.

- **Finance Management**

  Fees, assets, donations and payslips can be issued and monitored using this system. This would simplify financial dealings and accounts matters of the school.
3.1 TEACHING MADE SIMPLE

This is an extra feature included in the school management system where the online learning is implemented. Teachers would be able to upload the teaching materials and post a message to students. Students would be able to download, view or print the lessons exercises prior to the actual lesson hour. This would enable the students to access their educational material anytime and anywhere. Teaching would be made simple and effective through the usage of this online learning tool where interactive materials such as animations, videos and audios could be uploaded for the students to view and learn on their own.

This system also benefits the teachers where the whole syllabus could be loaded into the system and released phase by phase to the students according to the lesson plan. This reduces the teachers work load and makes them productive and effective in teaching and guiding the students. Information sharing between schools is made possible through the existence of centralized server. Schools can share resources to create and use standardized materials and examinations through this system.

4. Advantages and System Security of the School ERP

This system benefits all parties in various of ways. The benefits for the school management are as follows:

1) Easy performance monitoring of individual teaching modules.
2) Automated and quick report generation along with process turnaround time.
3) Centralized data repository for trouble-free data access.
4) Authenticated profile dependent access to data.
5) User friendly interface requiring minimal learning and IT skills.
6) Design for simplified scalability.
7) Elimination of people dependent processes.
8) Minimal data redundancy.
4.1 Advantages to parents are:
- Frequent interaction with teachers.
- Reliable update on child's attendance, progress report and fee payment.
- Prior information about school events and holidays.
- Regular and prompt availability of school updates such as article's discussions forums, image gallery and messaging system.

4.2 Security
This ERP system integrates the information security elements, confidentiality, integrity and availability (CIA) for the security and privacy. Confidentiality is where the system prevents disclosure of information to unauthorized individuals or systems. The open source ERP system also provides integrity where data cannot be modified by everyone and undetectably. Apart from that, the system also promises availability where the data is available to authorized users anywhere and anytime. Authorized users can access and view the data through the web based platform which serves as a user interface to the user.

5. ADVANTAGES OF OPEN SOURCE COMPUTING SYSTEM

The thin client system in Tamil school computer labs are capable of providing affordable server-based open source computing solutions. The main advantages of using this system in schools are to increases reliability and consistency of technology. Through a password-accessible account, students, teachers, and administrators can store and can access saved documents and personal settings. As all files and programs are stored centrally, users can access their work from any computers on the network.

Eg. when a teacher or student “logs in”, the server provides them with their “desktop configuration”. Users can even access their “desktop” from home or other remote locations. The other advantages of open source computing thin client system can be listed as below:

- Less Administration – Central management of users, patches, software, data, and backups.
- Higher Security – Elimination of viruses, Trojans or other vulnerabilities on the user desktops.
- Hardware Independence – Support of virtually all client devices and computer hardware, with low system requirements (eg - Pentium 3 with 512MB RAM).
- Easy Access – Teacher and students can access their documents and applications from any computers in the local area network.
- Reduction in TCO – Total Cost of Ownership reduction by up to 50%

Benefits for school in using open source computing thin client system:
- Lowers cost of technology over time
- Secure data and equipment
- Less downtime and greater efficiency
- Reduces administrator staffing costs
• Lessen the risk of data theft
• Disaster recovery: Data is more Secure
• Reduced time for technical Support
• Lower power consumption: save electricity
• Zero licensing management
• Minimum Maintenance
• Highly trained individuals are not required

5. Advantages of Thin Client Technology

By using thin client technology rather than standalone computers, 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 computer system is to be used. With conventional computer 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 without 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 computers have the appropriate software available also affects computers which are located in teaching spaces. Traditionally such computers 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 computers 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 computers system also causes difficulties, when a lecturer arrives in a class room, there will be about 5~10 minutes start up time for the conventional computers 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.

5.1 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 – Shows an image of ideal thin client.

Plate 2 shows image of computer labs before and after with students using refurbish machines operating as thin client.

Plate 1: Shows Ideal Thin Client

Plate 2: SJK (T) Bukit Raja, Klang before and after setting up computer laboratory
6. Conclusion

The awareness of benefits and advantages of thin client and server based computing technology have resulted in the growth of Tamil schools implementing this technology in Malaysia, with the supports from governmental and non-governmental bodies. With the use of thin client technology and school management system, the teaching system will now 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. Through implementation of this system, Tamil schools in Malaysia will soon become community information hub where resources can be maintained and shared for the uplifting of the Malaysians. The students benefited from this technology will become independent learners and one day become knowledge based skilled leaders.

With the open source applications and thin client technology it was possible to decrease the cost of installation and the cost of maintaining the computer lab. With servers installed at each and every schools, the setting up of centralized school management system is possible. Currently we have successfully implemented server base teaching with thin clients for about 25 over Tamil schools in Malaysia. Currently we are working and developing further improvement into open source by performing R & D into the implementation of Open Source base ERP for schools to manage the 25 schools systematically using centralized server. Most importantly design system are required to be scalable, sustainable, maintenance free and most importantly able to eliminate the digital gap between poor and rich students and build the digital bridge between urban and rural students.

7. Acknowledgement

The authors wish to extend special thanks to: all the individuals and institutions involved directly and indirectly by providing financial and material support for this research work, especially Selangor State Government, National Land Finance Co-Operative Society (NLFCS), Malaysian Community, Education Foundation (MCEF) and Hindu Youth Organisation, Klang.
An Innovative Soft Ware for Learning to Write Tamil Lesson Plan

Dr. S. K. Panneer Selvam
Assistant Professor, Department of Education,
Bharathidasan University, Tiruchirappalli, TN
skpskpbdu@gmail.com

Introduction

Quality Education is indispensable for ameliorating the secondary education to stand on the platform of globalization. It can be acquired by revamping the quality of Secondary Teacher Education through the degree course of Bachelor of Education. Teachers are responsible facilitators for ensuring the quality education. Teacher Education can assist to enrich the quality of teachers by implementing innovative methods in teaching learning process. Most of the students are unable to read and write Tamil without mistakes in Tamilnadu. Even if the Tamil is the mother-tongue of the learners, they are unable to write correct Tamil. Parents and learners are attracted by learning English and neglect learning Tamil language. It paves way to the learners for committing more mistakes. Errors of learning Tamil can be rectified through effective teaching learning process. Innovative effective method has to be investigated for easy understanding of the language. Hence innovative teaching-learning software was prepared for error free successful learning to write Tamil lesson plan.

Need of the Study

Effective teaching is based on the planning of lesson plan which is depending upon unit plan. Writing and preparing a lesson plan for macro-classroom is essential for a perfect teacher-trainee or in service teacher. Most of the teacher-trainees of selecting their optional-I as Tamil at Bachelor Education level were unable to write lesson plan in Tamil from some Educational colleges situated around the Trichy district. Lesson plan is the skeleton of the teaching learning process. If the student-teacher is unable to write an appropriate lesson plan for particular unit of the lesson for /poem/prose/grammar, his teaching can not be more effective. Acquiring practice in lesson plan writing may eliminate the problems of the teacher-trainees in teaching learning process. Learning to write Tamil lesson plan among the teacher-trainees was identified by administering a diagnostic-test. Teacher-trainees of Tamil scored very less mark in writing lesson plan. Hence the researcher endeavoured to prepare innovative Software for the teacher-trainees to eliminate the problems in conventional methods of learning to write Tamil lesson plan.

Statement of the Problem

Teacher-trainees of optional I as Tamil in Bachelor of Education have problems in writing lesson plan in Tamil by conventional method.
Objectives of the study

1. To measure the learning hurdles in writing Tamil lesson plan.
2. To find out the significant difference in achievement mean score between pre-test of control group and post-test of control group in learning to write Tamil lesson plan.
3. To find out the significant difference in achievement mean score between pre-test of experimental group and post –test of experimental group in learning to write Tamil lesson plan.
4. To measure the effectiveness of innovative software in learning to write Tamil lesson plan.

Hypotheses of the study

1. Teacher-trainees of optional I as Tamil in Bachelor of Education have problems in writing lesson plan in Tamil by conventional method.
2. There is no significant difference in achievement mean score between pre-test of control group and post-test of control group in learning to write Tamil lesson plan.
3. There is no significant difference in achievement mean score between pre-test of experimental group and post-test of experimental group in learning to write Tamil lesson plan.
4. Innovative software is more effective than conventional methods in learning to write Tamil lesson plan.

Methodology

Experimental method was followed in the study.

Sample

One hundred Teacher trainees of B.Ed were selected from Indira Ganesan college of Education, Trichy as sample for the study. Fifty Teacher-trainees were considered as Controlled group and another Fifty Teacher-trainees were considered as Experimental group.

Tool

Researcher’s self-made criterion reference test was used as a tool for the study.

Reliability of the tool

The co-efficient correlation was found 0.78 in the tool through split-half method.

Validity of the tool

Face validity and Content validity was established for the test through expert suggestions. Hence reliability and validity were properly established for the study. ‘t’ test was used as a statistical technique for the study.

Procedure of the study

1. Problems identification by administering diagnostic test.
2. Preparation of innovative software and validation.
3. Pre-test-treatment-post-test.
4. Finding the effectiveness of the software.
5. Implementation of the study.
Data collection

The researcher administered a diagnostic test to identify the problems of the Teacher-trainees in learning to write Tamil lesson plan with permission of Principal of the college. Pre-test -Treatment-Post-test was used for the control group in conventional method. Pre- test- using the innovative software and post -test was administered for the research.

Hypothesis testing

Alternative Hypothesis-1

Teacher-trainees of optional I as Tamil in Bachelor of Education have problems in writing lesson plan in Tamil by conventional method.

Teacher-trainees scored 24% of marks in writing Tamil lesson plan. Seventy six percentages of teacher-trainees committed the mistakes in writing lesson plan. It substantiates that the problems existing in writing lesson plan among the teacher-trainees in Tamil by conventional method. Hence the Teacher-trainees of optional I as Tamil in Bachelor of Education have problems in writing lesson plan in Tamil by conventional method.

Null-Hypothesis-2

There is no significant difference in achievement mean score between pre-test of control group and post-test of control group in learning to write Tamil lesson plan.

Table-1

<table>
<thead>
<tr>
<th>Tests</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>t value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test control group</td>
<td>50</td>
<td>8.62</td>
<td>2.32</td>
<td>98</td>
<td>0.214</td>
</tr>
<tr>
<td>Post- test control group</td>
<td>50</td>
<td>8.68</td>
<td>2.42</td>
<td></td>
<td>Insignificant</td>
</tr>
</tbody>
</table>

Achievement means scores between pre-test of control group and post-test of control group.

The calculated t value is (0.214) less than table value (1.96). 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 teacher-trainees in writing Tamil lesson plan through conventional methods.

Null Hypothesis- 3

There is no significant difference in achievement mean score between pre-test of experimental group and post-test of experimental group in learning to write Tamil lesson plan.
Achievement means scores between pre-test of experimental group and post-test of Experimental group.

The calculated ‘t’ value is (9.32) greater than table value (1.96). Hence null hypothesis is rejected at 0.05 levels. Hence there is significant difference in achievement mean score between the pre-test of Experimental group and post-test Experimental group in achievement mean scores of the teacher-trainees of B.Ed College in learning to write Tamil lesson plan.

Hypothesis- 4

Innovative software is more effective than conventional methods in learning to write Tamil lesson.

The above two tables prove and confirm the innovative software is more effective than traditional approaches in learning to write Tamil lesson plan. Mean scores in pre-test of Experimental group by conventional method is (14.39) less than the mean score of post-test of Experimental group by using innovative software in learning to write Tamil lesson plan (18.12). It substantiates that innovative software is more effective than conventional methods in learning to write Tamil lesson plan.

Findings

1. Teacher-trainees of optional I as Tamil in Bachelor of Education have problems in writing lesson plan in Tamil by conventional method.
2. There is no significant difference in achievement mean score between pre-test of control group and post-test of control group in learning to write Tamil lesson plan.
3. There is significant difference in achievement mean score between pre-test of experimental group and post-test of experimental group in learning to write Tamil lesson plan.
4. Innovative software is more effective than conventional methods in learning to write Tamil lesson.

Recommendation of the study

1. Preparing more software’s for learning of Tamil may simplify the learners of Tamil in mother tongue as well as Tamil as second language learners.
2. It may be implemented in Diploma in Teacher Education also.
3. It may be implemented in School Education also.
4. It may be implemented in Collegiate Education also.
Conclusion

Like this study can eliminate the problems of the learners in all levels of education.

References


Attitudes and motivation in teaching through ICT Among Malaysian Tamil Teachers: an overview

Paramasivam Muthusamy, Ph.D.,
University Putra Malaysia, Malaysia
E-mail: param@fbmk.upm.edu.my

Abstract
It is an accepted fact the progressive motivation and positive attitudes play significant role in attaining learning achievements. Computer based learning too is not an exception to this observation as communication and Information Technology (ICT) have made deep inroads to teaching and learning among teachers and students. The Ministry of Education in Malaysia have up-graded the language laboratories in schools and have installed sufficient computers for both teachers and students to use in the teaching and learning process. It is needless to say that the role of ICT is very important in helping learners in comprehending as ICT provides avenues through unlimited collection of text, sound, pictures, videos, animation and hypermedia (Bruner, 1986). This can support the findings of Fisher (1996), who argues that computational environment is needed to support ‘new frameworks’ to education. The aim of this study is to explore the attitudes and motivational levels of in-service teachers who are serving in Tamil schools. A questionnaire adapted from (Wong, 2002) will be used to identify teachers’ attitudes and their motivational levels in teaching. Besides that an interview will also be carried out to further question teachers’ on their attitudes and motivation. The data from the questionnaire and interview will be used to analyze the Tamil teachers’ knowledge, attitude and motivation in using ICT in classrooms. This information will equip the researcher if ICT is being explored to the fullest by the teachers in Tamil schools since facilities had been provided by the Education Ministry and ICT is taught as a subject in the Tamil schools.

Introduction
During the past couple of decades Information communication technology (ICT) and its tremendous growth have made remarkable and significant inroads into almost all the disciplines. One of the instruments for the fast developments of ICT is the growth of computers. As a result, one should know very well the power and the potentiality of this medium. Subsequently, one cannot afford to be a computer illiterate in this era of globalization. Not only that, apart from the knowledge in his/her discipline the success of the person depends mainly on his/her extent of knowledge and the potentiality to use and exploit the computer technology in his discipline. The field of education is not an exception to this rule. In fact, one can assert that the field of education can contribute remarkably by exploiting the potentialities of computers and its allied areas such as, multimedia, internet, software development, need based computer assisted language learning/teaching (CALL,CALT) etc.
Schemata for the Current Study

The present study on Attitudes and Motivation in Teaching through ICT among Malaysian Tamil Schools has been viewed from two angles. The first perspective is to view from the student’s point of view and the second is to view from the teacher’s point of view. Seeing through these two angles is very important because there is a significant gap between the teachers and the taught as far as ICT is concerned. In other words, the teacher’s attitude and motivation get reflected on the student’s and the student’s attitude and motivation get reflected on the teacher’s. This paradigm shift gets reflected on the achievement levels of the learners as well as the teachers.

Ict and Students in Malaysia

Though the knowledge level in the field of ICT among the student population in Malaysia depends on several social, economic, linguistic and educational factors, it is an established fact that the students born after 1980 in general are considered as having digital mind and also known as N-Gen-Net Generation (Tapscott,1998). These groups of students are highly motivated and influenced with internet, computer application and have changed their learning attitudes and achievement levels (Adone et al. 2007). Subsequently, these students are more at ease in the use of computers and also have the expertise to exploit the potentialities of computers and its related areas of education and in other areas of acquiring knowledge. As a result, they are fully aware of the use of computers, multimedia packages, internet etc. They are also aware how these can help them remarkably in the form of its collection of texts, sound and pictures, video graphics and hypermedia in order to increase their knowledge and learning process. Subsequently, they use them extensively whenever necessary and their motivational and attitudinal levels are very high as far as the use of ICT is concerned.

ICT and Teachers in Malaysia

In this situation the development of the basic positive attitude towards the acceptance and the use of the computer are necessary among the teachers. Attitude here is referred to the tendency to behave positively or negatively towards an object, situation, concept or a person (Aiken 1976).

As opposed to the highly motivated and with positive attitude of the general student population in the use of computers and other ICT applications, the teacher’s population can be grouped into three categories on the basis of their age, education level in computer application and the number of years of exposure to computers and its applications in general and education in particular. On the basis of the three factors mentioned above the teachers in Malaysian educational context are concerned there are four dimensions which act on them in the formation of attitudes towards ICT. They are, age, background of the teacher, belief and the teacher’s extent of exposure to ICT. Age can be divided into two categories namely, those who are above 50 and those who are below 50; background or the opportunity to acquire the computer related knowledge; the extent of exposure to the computer application coupled with the opportunity to use the ICT applications and overall belief system among the teachers. These are the four dimensions which contribute towards the formation of the teacher’s motivational factors in ICT. On the basis of the four dimensions listed above the motivational factors for attaining confidence, development of computer based supporting skills, building positive environment around them and the varying degrees of anxiety which affect the formation of motivation will be determined.
Objectives
The main objectives of this study are,

- To identify the attitudes of the teachers towards teaching through ICT
- To identify the motivational factors which contribute for teaching through ICT
- To identify the relationship between the attitude and motivation in achieving the goal of teaching through ICT

Hypotheses

- Exposure to ICT contribute positively for the development of positive attitude towards teaching through ICT
- Higher the age group lower the motivation to use ICT in teaching
- Lower age coupled with higher exposure to ICT has a positive role to play on the development of efficient teaching strategies

Methodology
This study which is mainly focusing on the quantitative approach to the study of attitude and motivation will be undertaken in the 10 Tamil schools situated in Klang Valley, Peninsular Malaysia. Five teachers (including ICT teachers) from each Tamil schools will be taken as the subjects of the study (n=50). These teachers are directly involved in teaching ICT in schools. A questionnaire adapted from Wong (2002), will be used to identify teachers’ attitudes and their motivational levels in teaching ICT in the selected schools. The questions in the questionnaire are classified under two categories namely, the attitudinal questions and the questions related to motivational aspects. Each category mentioned above has 4 dimensions. For attitudes, the selected teacher’s background, age, belief and exposure to ICT are included. In order to obtain information regarding motivational aspects of the teachers the ways through which they gain confidence, attain teaching skills and manner through which the teachers try to avoid the mounting pressure on them which result into anxiety. The data will be analyzed using SPSS software. Descriptive and inferential analysis such as frequency, percentage, mean and standard deviation will be used to describe the general data of the study. Besides this, analysis such as independent T-test, ANOVA and Pearson Correlation will be employed to discover any relationship and differences between the dependent and the independent variables of the study.

Framework
The study will look at ICT teachers’ performances toward teaching ICT from two parameters and eight dimensions. The two dimensions are teachers’ attitudes and motivation towards imparting ICT knowledge to their students. To measure teachers’ attitudes several information based on their age, background, exposure and belief in ICT will be gathered through a specially designed questionnaire. The other dimension is motivation. Teachers’ confidence, skills, environment and anxiety in using computers will be analysed. These information will reflect ICT teachers performance in their classrooms.

Conclusion

The achievements of using ICT in all the Tamil schools nationwide in Malaysia does not depend only on the ICT laboratory which are well equipped and other facilities provided by the government but also on teachers involvement. Teachers’ attitudes and motivation towards ICT teaching plays an important role in promoting and imparting ICT knowledge to students. Therefore, this study looks into how teachers’ attitudes and motivation helps in promoting ICT usage in all the Tamil schools.

References

CONFERENCE PAPERS

TAMIL INTERNET 2011
Raising achievement and aspiration

Siva Palanisamy

| 26 |

"Raising achievement and aspiration"

(National Curriculum, 1999, p. 136)

"The government of India established the National Curriculum and Textbook Commission to produce the National Curriculum and Textbooks for the teaching of the seven major languages, namely, Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, and Tamil. The Commission prepared a comprehensive curriculum for these languages, which was published in 2006. This curriculum was based on the principles of integrative and inclusive education, with an emphasis on the development of language skills and the promotion of cultural diversity." (National Curriculum, 1999, p. 136)

"The curriculum emphasizes the importance of the seven major languages, which are integral to the cultural and social fabric of India. It aims to provide a balanced and integrated approach to the teaching of these languages, with a focus on developing language skills, promoting cultural awareness, and fostering a sense of national identity." (National Curriculum, 1999, p. 136)

"The National Curriculum and Textbook Commission emphasized the importance of the seven major languages, which are integral to the cultural and social fabric of India. It aims to provide a balanced and integrated approach to the teaching of these languages, with a focus on developing language skills, promoting cultural awareness, and fostering a sense of national identity." (National Curriculum, 1999, p. 136)
Facebook and Tamil Language in Singapore’s Teacher Education

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

Abstract

Media serves as an important motivator in the language teaching process (Brinton, Donna. 2001). Especially internet, which is accepted as the concrete compound of all the technologies (Kartal, E., and Arikan, A., 2010), enables us teachers/trainers to speak the same language as our students/trainees do. At the same time, they help us to connect to our customers in a personalized way of teaching language. In Singapore, Information and Technology has played a critical role in developing teachers and it is also the same for the Tamil teachers. At the National Institute of Education, the pre-service learning for the Tamil teachers was conducted this year with an instructional designer’s technical guidance and moral support, I have embarked on with the pedagogical training through Facebook, Posterous and Voice-thread to develop writing skills, listening and speaking skills for the trainees in diploma in Education and Master in Education courses respectively. With the process based product approach, through these well-known social networks, we witnessed that there are many channels for our students to learn and construct knowledge in their lives and teacher is not the only person to provide knowledge and monopoly in the classroom (Lee Sing Kong, 2011). Through the shared, open concept based networks, the participants have enjoyed and constructed their knowledge on culture, identify the teenage related topics, interesting RJ (Radio Jockey) techniques, responsive, critiquing listenership and creative producers of the feature programmes. This paper practically shares the Tamil Language Teacher Training through Facebook in order to develop 21st century educators for teaching the 21st century learners in Singapore Tamil Classrooms. We chose Facebook, a social networking platform, for role playing by creating fictional characters and the project went well with the course participants and received invaluable responses. This paper also shares the importance of having the technical and instructional expertise with the same lingo as you have.

Key words: Facebook, Tamil Language, Second Language, Social network

Introduction

In Education, technology especially information technology has its own trademark in instilling interest and motivation among students towards education. Also it is a right hand and tutoring assistant for the educators who are passionate on teaching or educating their students. In our National Institute of Education, we have a division to help the academics to embark on IT infused pedagogy. Although there were various kinds of support of the academics in their teaching through information technology, recently the formation of CeL which is the Centre for e-Learning is a boon for us. I myself
was engaged in the following pedagogical deliveries with the assistance of IT and they are: Vimba live, Web quest. However, from last year onwards I had a privilege to work with Instructional Designer to learn new technologies to enhance my pedagogical initiatives. Facebook is one of the innovative pedagogical methods which we felt are successful and we have received excellent feedback from our trainees from the two groups of diploma students.

Around April which was the end of 2010 January semester at our NIE, I had expressed my wish to the CeL colleagues, Ms Pratima Majal, Senior Instructional Designer and Ms Shamini Thilarajah, Instructional Designer to introduce social network based pedagogical training to the Diploma in Education year I and year II Tamil trainees (please see the annexe) in the forthcoming semester which was the September 2010 semester. That lead to a few discussions on preparing the ground work by Pratima and Shamini who are the senior instructional designer and Instructional designer respectively s at CeL. This project is about infusing social communication network modes in the teaching and learning of Tamil language. In this particular project, Facebook mode has been used in the teaching and learning of second language. As Facebook is familiar among youngsters, we would like to explore this network.

**Learning and the Information Technology:**

For the past two decades, we could witness the infusion of technology in the teaching and learning fraternity. Also, the trainers/teachers are working closely with their students to come up with presentations to encourage and entertain one another. Here are some of the examples on the effective use of information technology and research initiatives on it. Nowadays, language learning is not only to communicate, but also to establish contacts, meet people and establish partnerships (Soontiens, 2004 in Sarah Elaine Eaton, 2010). In a collaborative online community, each student’s ideas and knowledge are available and are a resource for everyone in the class (Hewitt and Scardamalia, 1998 in Kathryn I Mathew Emese Felvegi and Rebecca A Callaway, 2009). It gives an opportunity for collective knowledge and connections. Online information technology allows students to obtain information through their cognitive, emotional (Kartal, E., and Arikan, A., 2010), cultural, psychological experiences.

Judith Rance-Roney (2010) stated that the digital story telling technology for the English Language Learners to understand the cultural background, literacy skills and the language development to deal with the literary texts. At the same time, use of 3D Virtual Learning Environment for the students to understand the social words and personal learning (Jonathan Barkand and Joseph Kush, 2009). Waters (2008) and Sarah Elaine Eaton (2010) stated that the use of Skype which allows international connection between students and teachers from two or more countries proved good results in developing second or foreign language skills and teachers’ professional development in pedagogy. At the same time, using blogs as an ICT tool in the language class is an effective tool and facilitator to develop reflective learning strategies among students (Hourigan T and Murray L., 2010). Faizah Mohamad, 2009 reiterated that an Internet based grammar instruction is useful in language class and facilitates understanding to the students instead of Conventional pen and board instruction Robert Hamilton 2010 encourages to provide curriculum support always to the lower proficiency students as well as the higher proficiency students through YouTube. Nowadays, computer technology enables students in their language learning and they are: 1. Experiential learning 2. Motivation 3. Enhance
student achievement 4. Authentic materials for study 6. Greater interaction 6. Individualization 7. Independence from a single source of information and 8. Global understanding (Lee, 2000 in Arif Bulut, 2004). Mobile Assisted Language Learning (MALL) and Computer Assisted Language Learning (CALL) for students to harness their creativity to express themselves and take ownership of their learning. In a class, with the teacher there are other inevitable factors to contribute the success of the curriculum: CD, DVD players, Blogs (Doris de Almeida Soares, 2008)scola Naval, Internet, Wiki, Web quests, Virtual fieldtrip, spreadsheet programme, software assisted writing, web 2.0, desktop publishing programmes, graphic organizers, recording devices, podcasting and media(Gwen Troxell, Castleberry and Rebecca B Evers, 2010), Facebook is another feature in this line and it plays an essential and critical role in today’s education, politics, social development and cognitive development.

Facebook which refers to the distributed authorship, collaborative and cooperative learning, openness, careful and purposeful usage of web 2.0, developing cultural awareness.

**Tamil language and the information communication technology (ICT):**

In Singapore, Tamil has been taught as a mother tongue language at second language level. With the government support for the Tamil language, almost every school is equipped with necessary ICT infrastructure. Outside India, in Singapore with Tamil having the official language status, Tamil education has been developing its own curriculum and pedagogy. It is also contributing to the Tamil internet. In the Tamil speaking world, information technology has many facets and here are some of those initiatives. Tamil has been used for a variety of reasons with ICT for teaching, computational linguistics, mobile devices and providing assistance to the less privileged students. When we surface the presentations at the last year’s Tamil internet conference (Vasu Renganathan, 2010), the papers are mainly in 9 categories. At the recent Tamil Internet Conference in India the following issues on Tamil and ICT were dealt with:

**Teaching and Learning of Tamil**

- Tamil Diaspora: Teaching Tamil as a second language and impact of Technology
- Technical Development
- Tamil in Mobile Phones and Handhelds
- Natural Language Processing; OCR Text to Speech Machine Translation Etc.,
- Tamil E-texts, Corpora and Digitization of Ancient Tamil Texts
- Morphological Tagger
- Electronic Dictionaries and Glossary of Technical Terms

Here in this project, we wish to enhance our trainees’ knowledge through networking with one another on culture and upgrading their knowledge by interacting critically on culture with one another.
Project Objectives:

There is a small story behind this project. Before I started this project, I thought of using this project as an effective platform to develop the training on the four critical language skills. Also based on my observations on my trainees, I found that they need more assistance in understanding the in-depth meaning of Indian culture especially Tamil culture and the traditional practices. Although many of them are from Tamil speaking homes, they had less time to use the language in schools and community domain as they studied Tamil as a second language and generally they have fewer opportunities to meet friends from their own ethnic group. It is also difficult to get them hooked towards the Tamil literary functions. In the internet, there is limited information for the Tamil Diaspora to read and understand. Quite a significant number of them haven’t been to India to experience or immerse in that cultural world. Hence, I shared my wish to Pratima and Shamini about the project on Tamil culture.

Objectives of the Tamil curriculum in Singapore:

Here, the objectives of the Tamil curriculum in Singapore on learning of Tamil Language are given below:

• Providing proper training to the students in the basic language skills in Listening, Speaking, Reading and writing
• Explaining the Tamil cultural and traditional features
• Helping them to acquire the characters which are essential for the formation of a country (MOE, 2008)
• For Tamil students in Singapore, the main initiative by the community is to make Tamil a living language by actively using it at home and community
• Nurturing Active Learners and Proficient Users of the mother tongue language (MOE, 2010).
• To make it happen, the Tamil learning has to be fun and cool. Students would like to enjoy the lessons while learning the language.

Based on the curriculum objectives, there is a clear understanding that to enable the 21st century students to be equipped well in their mother tongue/learning of Tamil language, we need to have well equipped teachers to teach them. Hence to create a 21st century teacher, he or she should know about the module in that semester:

• Infusing Tamil language teaching through Facebook network
• Equipping teaching and writing skills in Tamil language
• Developing positive social networking skills
• Learning from society and providing learning to others(peers) as well
• Responsible learning and teaching practices
• Understanding the culture of the Tamil community in Singapore and Diaspora countries

Based on the above mentioned needs, we designed the project with the following goals:
Goals:

- Equip the trainee teachers to pick up the necessary teaching skills and to use Facebook in their schools
- Understanding the responsibilities of using Facebook as a teaching material
- Niche areas on teaching through Facebook
- Engaging students in an enriching way
- Letting them understand that Facebook provides new avenues for teaching
- Making Tamil learning as a fun and cool feature!

With the above mentioned goals, we also have some niche areas to try and implement this project. They are the main reasons to embark on this project.

The need for introducing this project:

- Tamil trainees need to work on new pedagogies
- They need to engage their students well and make teaching an interesting feature
- Tamil trainees need to provide assistance to their students to speak well in Tamil
- To make Tamil as a living language in Singapore
- To make Singapore a Hub for teaching Tamil as a second language

Although in my earlier modules, I have introduced the infusion of ICT in the teaching of Tamil language modules with web quest, student centered lesson learning package, vimba voice, video conferencing and distance learning through ICT. Although each and every initiative has its own unique features, this particular initiative is different from the previous ones. What are the differences?

In this project the following are the potential different features while compared to the previous IT projects based Tamil teacher training:

- Harnessing Information Technology in teaching practices
- Sharing and gaining information through their interactive postings
- The use of Facebook by the trainees has responsibilities and is different from the normal Facebook usage which is for leisure.
- They have to have regular updates of new information from their contacts.
- They can play farm wheel game and can send pictures related to this
- Evaluation based on weekly postings and reflections. Especially, on Writing skills, depth of understanding,
- Questions are posed by them and responses are given by them as well
- They have to write reflections weekly for evaluation

In this project, we have given strong emphasis for the process than the product as the process is the feature to give value to educational software or an educational pedagogy. Let us view them here:
Facebook is used in Singapore regardless of their age, educational background, socio-economic status and professional background. But, we could not conclude that everybody understands the process or the potential positive and negative features of using it in their lives. Here, future leaders of education
and gatekeepers of the Tamil language in Singapore, we would like to ensure that these young trainees know well about the process of this project and to instill in them that the process is important in every phase of their professional life and teaching and learning of Tamil language in Singapore. Also, Mary Clarie (2010) stated that “It seems that the future of social media in the classroom will not reach its fullest potential until we can bridge the divide between new media and traditional academia”. Also both parties made it clear in this joint effort between the pedagogic and the ICT instructional designer. In this Facebook related ground work and ICT training to the students, we have make sure that continuous monitoring is there. At the same time, it is easy to come up with product. But it is difficult to come up with product based on a Process. So, the process is the important for us. Now let us go through on how the Process is important to us in this project.

How the Process is important?

The Process:

• It’s a set of Lesson and Evaluation procedures
• Pre - preparation between the trainer and the ICT Instructional Designer
• Provide training to the trainees from two classes
• Tight timeline to be familiar within Facebook
• Each trainee needs to create a name and profile based on his / her country and context
• Sign undertaking on responsible use of Facebook
• Three postings per week for 12 weeks
• One reflection for every week
• Pictures, songs, video clips and artifacts in their postings
• Immersed in their Facebook interactions and reflections
• Closely monitored by the trainer for content, paraphrasing and by the ICT Instructional Engineer on the use of Facebook and the information technology
• Evaluation is there for their movement, entries, reflections, pedagogical and ICT knowledge
• Although there is a grace and transition period of 2 weeks, but a few of them performed well from the first week onwards.

We would like to express our thoughts that the above mentioned process went well in this project. At the same time, I have to mention that I have received excellent support from the IT Instructional Designer at each and every level of this journey. Also, both of us were able to speak, write and communicate well which is an additional advantage. We could easily share the nuances of the language, culture and context based information. At times, the trainer remember their ‘Facebook names’ and at times forget their real names.

Feedback:

With the feedback on this project and as a trainer, I could say that this project creates a very good understanding and learning of Tamil culture, writing, paraphrasing and reading. It provides avenues to teach, vocabulary, Spoken Tamil, functional grammar and Listening and Speaking of Tamil. This Facebook contents will be a rich resource for the trainees understanding of authentic Tamil culture.
and practices. It encourages effective search, note sharing and collaborative learning. The project has a number of features to add fun and cool in Tamil learning. It can be easily adapted for the upper Primary and Secondary class students.

In this project, students faced some challenges at the starting of the project as they have to get a suitable character role and create profile for themselves. After they have positioned themselves well in their roles, they had a constant challenge to prepare weekly postings and reflections for 12 weeks in the midst of their normal training and other modules they had to cover for that semester. Not only that, they also need to learn new information about their country and need to learn each other’s cultural domains. Then, they had to ensure that there is good quality inputs and paraphrasing. This is to prevent plagiarism. In the whole process, each group of trainees (Dip Ed I and Dip Ed II) had to control themselves from interacting with the other group of trainees.

In the journey of this project, as a trainer and facilitator, I too faced a number of challenges as given below:

- Providing positive comments and suggestions during the first two weeks
- Familiarizing Tamil Terms and reading and replying to postings in the Facebook
- A hands-on session given for the Dip Ed II class proved to be informative and helped to clear and clarify many doubts
- Availability of Tamil software outside office is a good advantage
- First two weeks, there was a slow development and a certain level of hesitation among the trainees in following the procedures.

But most of the challenges were turned as happy developments in the middle and later part of the project. They are given below:

- Trainees did their postings regularly
- They were Creative and Critical in their reflections
- There was more interaction between themselves in knowledge creation
- Not much questions were posed to Veerasamy and Muthusamy

**Facebook information, conversations and thoughts: A Midterm review:**

During the 12 weeks of journey time, the facilitator tried to encourage their efforts and provided needed explanations on their cultural domains based postings. The trainees themselves appreciated and critically analysed their classmates’ postings and raise awareness with additional questions. This encourages the whole group to move to a level up and working hard to come up with more additional information. The instructional designer provided the updates on the students’ postings and advice on their queries regarding the technical issues. At the same time, the facilitator engaged and encouraged the participants to provide their insights on Indian culture especially Tamil culture and traditions.

In October 2010, we had come up with a table to know their understandings of Tamil culture and traditions. Here the results are given below:
N = 13

<table>
<thead>
<tr>
<th>No</th>
<th>Category</th>
<th>Information</th>
<th>I knew already</th>
<th>Now only I know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Food</td>
<td>கார்னம் (Dry Ginger Tea)</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>2.</td>
<td>Food</td>
<td>செம்புருப்பிட்டம் பெருந்தில் பர் பையன் (Use of pepper in Tamil cooking)</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>3.</td>
<td>Food</td>
<td>ஆடையுரும் சோயா (Spices box with five rooms)</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>4.</td>
<td>Costumes</td>
<td>திருக்குருல் பாணைச்சுறு (Thirukkurall Silk Saree)</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>5.</td>
<td>Costumes</td>
<td>ரைசுப்பு ராஜபுர் பல்லல் பொனையாம் (Rajput Queen’s Like on Sarees)</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>6.</td>
<td>Traditional Arts</td>
<td>பாரதானத்தெரு பெருந்தில் பல்லாத் பொனையாம் (Various Dance Forms in Bharatanatyam Dance)</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>7.</td>
<td>Traditional Arts</td>
<td>தென்றாரு (A dance form called, thattadaru)</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>8.</td>
<td>Traditional Arts</td>
<td>கொலாட்டாம் (Kolaattam - Dance with two sticks)</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>9.</td>
<td>Traditional Arts</td>
<td>கராட்டாம் (Karaattam - Dance with a pot on the Head)</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>10.</td>
<td>Traditional Arts</td>
<td>குசுப்பட்டி (Kuchuppudi - A dance of Andhrapradesh, India)</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>11.</td>
<td>Traditional Arts</td>
<td>108 பொருவியான பல்லாத் பொனையாம் (Nadarajaa’s dance talents)</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>12.</td>
<td>Ancient Tamils' Lifestyles</td>
<td>சீயாக்கை தனியியவுக்கும் உயர்வு பொனையாம் (Seeyakkaai -Shampoo vegetable’s uniqueness and coolness)</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>13.</td>
<td>Ancient Tamils' Lifestyles</td>
<td>சாடனம் தனியியவுக்கும் உயர்வு பொனையாம் (Sandal wood’s uniqueness and transitional speciality)</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>14.</td>
<td>Ancient Tamils' Lifestyles</td>
<td>தமிழ் செயற்கை விளை (Special features of Tamil Fengsui)</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>15.</td>
<td>Ancient Tamils' Lifestyles</td>
<td>பொருந்தம் பொருந்தம் விளை (Special meaning of the housewarming function/celebration)</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>16.</td>
<td>Ancient Tamils' Lifestyles</td>
<td>க்ரெல்வில் கூத்துப் விளை (Specialities of the cradle cloth)</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>17.</td>
<td>Ancient Tamils' Lifestyles</td>
<td>நாட்டுமருமையைப் பார்க்கவும் சுவரிகாந்திகள் பொனையாம் (Use of Natural Herbs in Facial Care)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>18.</td>
<td>Contemporary Art</td>
<td>தமிழ் சினிமா நடனத்தொழியில் எட்கர் (Tamil Film Industry’s achievement)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Diploma in Education II trainees on their understanding of content in this project:

N=12  (3 absent)

<table>
<thead>
<tr>
<th></th>
<th>Category</th>
<th>Information</th>
<th>I knew already</th>
<th>Now only I know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Food</td>
<td>உப்புக் குறிப்பிட்டியதை முன்னாளில் பயன்படுத்தியது (Use of Orange Sugar)</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Food</td>
<td>லட்டு (Laddu)</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Food</td>
<td>பாலவரைகள் பாலம்பிள் (Banana Nuts Thosai)</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Ancient Tamils' Lifestyles</td>
<td>பட்டர பணல் கீழிச்சி பண்பாடு (Use of the Fan made up of Palm Leaves)</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Mythological Stories &amp; Values</td>
<td>கிருஷ்ணாவின் நரகா பெருமான் (A video clip on Krishna and Narakasura)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Traditional Arts</td>
<td>தங்கம் சிறு தமிழ் பண்பாடு (Arts Festival at the Tanjore Big Temple)</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>Traditional Arts</td>
<td>காரக்காட்டல் பண்பாடு (Varieties in karagattam Dance)</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>Traditional Arts</td>
<td>பிளிப்பான் விஜயாவின் சூழல் (Guinness Record of Vijaya Mohan’s kolam)</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>Overseas Tamils and Traditions</td>
<td>உண்மையான ஒரு வேலைச்சுற்றுக்கு நிற்கிய (A specially organized event titled, Namaste France)</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>Overseas Tamils and Traditions</td>
<td>பிளிப்பான் தமிழ் வழக்கில் (A specially organised event titled, Bonjour India)</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>Overseas Tamils and Traditions</td>
<td>பொங்கல் பரகம் பரகல் (Pongal festival celebrations at USA, London, Sri Lanka and Australia)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>Overseas Tamils and Traditions</td>
<td>உண்மையான ஒரு வேலைச்சுற்றுக்கு நிற்கிய (Australian Tamil Associations and Activities)</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>
Two groups of trainees gave their comments on their projects at the mid of the semester in October 2010 and they are given below:

**Diploma in Education Year 1 Trainees’ comments:**

**Advantages**

- Although it is difficult to find answers for all the questions raised by our friends, there is a kind of happiness at the end as I know and I learnt new information.
- Easy to learn a variety of information at the same time. Know more information about our culture which I did not know before.
- It is easy to find the answers for the question. At the same time, can learn a number of rare details about our culture.
- Facebook is an excellent training ground to search, collate and store information for our communication of ideas.

**Challenges**

- There is a shortage of time. Because of this, it is difficult to read everybody’s comments, postings and questions. At the same time, it is difficult to ask questions from everybody.
- I find it very difficult to type in Tamil. Hence it takes a lot time to upload the information.
- Time is a concern.

**Diploma in Education Year II trainees’ comments:**

**Advantages**

- FB has enabled me to see a new way in teaching and learning Mother Tongue.
- Has raised awareness and interest to learn,
- I have enjoyed FB interaction.
- FB has bridged people from different countries and also helps to inculcate our Indian traditions and cultures.
- The FB project has enabled me to take ownership of my learning.
- The friendly interactions between friends from various countries enabled me to share resources and information effectively.
- FB project has incorporated effective role play.
- ICT incorporated project which has enabled people of many countries to join in one network.
- This has enabled us to comment on each other’s way of celebrating certain occasions and learn more about the cultures in other countries
- Having a Facebook account in Tamil has been quite fun.
- It is definitely an enriching experience as it provides opportunity to explore the laments of Facebook in Tamil.

**Challenges**

- The challenge however is that sometimes it is difficult to track previous comments made by other FB friends.
• If primary school students are to be engaged in this process, time would be a factor which they have to consider.
• We need more time to read, do research in order to post, comment and reflect effectively.
• Spend time commenting effectively, takes time and little research.
• Strong passion, commitment and love for the language are important. If a student just comes in to comment for the sake of doing, it defeats the whole process.

Evaluation

For the Dip Ed I and II trainees, this project is part of their pedagogical module. Hence we had weightage of 45% of marks for the year 1 trainees in their teaching of Tamil language I module’s main project which comprises 70% of marks for the whole semester. For the Dip Ed II trainees, this project focuses nearly 50% of their major project which forms 70% of my part of the module marks.

Generally, we could witness that the year 1 trainees were very enthusiastic and involved in this project with a number of comments and quality comments than the year 2. Although their groups are different, their topics are more or less same on culture. At the same time, technical expertise and experience based analysis; the year 1 trainees were well versed than the year 2. It also because of their IT orientation in the previous two Tamil projects in their first year modules. For year 1, this is their first IT based Tamil project for their very first Tamil pedagogical module.

Evaluation on Facebook for the Diploma 1 and 2 trainees:

<table>
<thead>
<tr>
<th>No of trainees Year.1</th>
<th>Marks (%)</th>
<th>No of trainees Year.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>91-100%</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>81-90</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>71-80</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>61-70</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>51-60</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>41-50</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>40 and below</td>
<td>1</td>
</tr>
</tbody>
</table>

There was a qualitative feedback collection done with both groups at different dates. That provided more insights on this projects and our planning of future projects.

Generally the pair of trainer and instructional designer team came up well and it is a milestone in the teaching of Tamil language and learning. We have learnt more and we will use our experiences in
planning our future modules in order to provide confidence to our younger teacher trainees who generally have learnt Tamil as a second language. They can be a role model to their trainees as well. Here some of our future plans:

- To share the development in this field with the Tamil communities in Singapore, India and Diaspora
- Publish a book with all the inputs
- Provide guidance to the trainees in their Teaching Apprenticeship and Teaching Practice and eventually in their future schools
- Able to do research on Singapore Tamil Trainees’ writing, questioning and answering techniques.

In this January 2011 semester, we have embarked on a project to enable the lower primary students from primary 1 and 2 classes to speak up and use spoken Tamil in their oral presentations and conversations. It went very well and we received positive feedback from the trainees.

For the Master of Education course participants, we have introduced radio Jockey (RJ) style of oral features and presentations on teenage related hot topics. It also went well with the course’s maturated and senior teachers who are in service. We will share our experiences and lessons on them at different platforms.

Acknowledgements

I wish to record my sincere acknowledgements to:

- Ms Shamini Thilarajah, Instructional Designer, Centre for e-Learning for her greater support in the whole process of this Facebook project
- Ms Pratima Malar, Senior Instructional Designer, Centre for e-Learning
- Dr Ashley Tan, Head, Centre for e-Learning for inviting us to share about this project with the NIE staff and highlight this project twice at the NIE seminars and giving constant encouragement to me
- My Tamil Trainees from Dip Ed I and II classes for their full support for our new initiatives in teaching and learning of Tamil language

References

- Doris de Almeida Soares, 2008. Understanding class blogs as a tool for language development. Language Teaching Research October 2008 vol. 12 no. 4 517-533


MOE. 2008. 2008 Syllabus Tamil Language Primary, Singapore: Ministry of Education


Annexe:

DIP ED I & II (Primary) Trainees’ FB Profile Name Lists

<table>
<thead>
<tr>
<th>No</th>
<th>Name(English)</th>
<th>FB Profile Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mdm XXXX</td>
<td>மநிஷாரா时时彩 (Manisharai)</td>
</tr>
<tr>
<td>2.</td>
<td>Mr XXXX</td>
<td>வேவோசாமினா (Velu Saminathan)</td>
</tr>
<tr>
<td>3.</td>
<td>Miss XXXX</td>
<td>மசைதேமைசுவியாதியி (Music expert Sarvivaadhini)</td>
</tr>
<tr>
<td>4.</td>
<td>Miss XXXX</td>
<td>சரோஜதேவி (Sarojadevi)</td>
</tr>
<tr>
<td>5.</td>
<td>Miss XXXX</td>
<td>நீலம்பாரி சாரவானா (Neelambari Saravanan)</td>
</tr>
<tr>
<td>6.</td>
<td>Miss XXXX</td>
<td>அணலெத்சுமி முத்து (Annaletchumi Muthu)</td>
</tr>
<tr>
<td>7.</td>
<td>Mdm XXXX</td>
<td>சுருளியோகாமல் மாதவா (Sundry Shop Madhavan)</td>
</tr>
<tr>
<td>8.</td>
<td>Miss XXXX</td>
<td>அஞ்சல் ராஜராம் (Anjali Raguram)</td>
</tr>
<tr>
<td>9.</td>
<td>Miss XXXX</td>
<td>கவிதா நாய் (Kavitha Nair)</td>
</tr>
<tr>
<td>10.</td>
<td>Miss XXXX</td>
<td>தில்லானா மகானா (Thillaana Mogana)</td>
</tr>
<tr>
<td>11.</td>
<td>Miss XXXX</td>
<td>வலயாபதி அராண்மா (Valayaapathy Annamma)</td>
</tr>
<tr>
<td>12.</td>
<td>Miss XXXX</td>
<td>அராதீஸ்வரி ஆராதி (Aaratheeswari Aarathi)</td>
</tr>
<tr>
<td>13.</td>
<td>MR XXXX</td>
<td>தஞ்சைகாத்து ராஜா (Thannikaattu Raja)</td>
</tr>
<tr>
<td>No</td>
<td>Name in English</td>
<td>FB Profile Name</td>
</tr>
<tr>
<td>----</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>1</td>
<td>MISS XXXX</td>
<td>Vaishnavi Raguraam</td>
</tr>
<tr>
<td>2</td>
<td>MISS XXXX</td>
<td>Karagaaattaakkaaran Maanguiyilu</td>
</tr>
<tr>
<td>3</td>
<td>MISS XXXX</td>
<td>Singakkutti Omanakutti</td>
</tr>
<tr>
<td>4</td>
<td>MISS XXXX</td>
<td>Malligai Mullai</td>
</tr>
<tr>
<td>5</td>
<td>MR XXXX</td>
<td>Avuttaa Raakki</td>
</tr>
<tr>
<td>6</td>
<td>MISS XXXX</td>
<td>Shreyaa Segaran</td>
</tr>
<tr>
<td>7</td>
<td>MISS XXXX</td>
<td>Muthazhagu Singavelan</td>
</tr>
<tr>
<td>8</td>
<td>MDM XXXX</td>
<td>Kundalagesi Sona</td>
</tr>
<tr>
<td>9</td>
<td>MISS XXXX</td>
<td>Oviya Sundari</td>
</tr>
<tr>
<td>10</td>
<td>MR XXXX</td>
<td>Singam Leo</td>
</tr>
<tr>
<td>11</td>
<td>MISS XXXX</td>
<td>Samisalokangas</td>
</tr>
<tr>
<td>12</td>
<td>MISS XXXX</td>
<td>Kayalvizhi Paalkova</td>
</tr>
<tr>
<td>13</td>
<td>MISS XXXX</td>
<td>Vishaalini Vishwanathan</td>
</tr>
<tr>
<td>14</td>
<td>MRS XXXX</td>
<td>Priyadarshini</td>
</tr>
</tbody>
</table>
Virtual Environment as a Collaborative Platform
to Enhance Pupils Information literacy skills

Sivagouri Kaliamoorthy
Beacon Primary School, Singapore
sivagouri_kaliamoorthy@moe.edu.sg

Abstract

The emergence of informative environment technology has made the ideology of learning, “anywhere, anytime”, a reality. The 21st century learners are equipped with readily available information at a mere click of a button. The Internet has erased international boundaries allowing our young charges the potential to develop as global citizens. This also demands that pupils develop a deeper understanding of the complex arrays of issues that involve them, now further complicated by the lack of traditional boundaries. Research studies have given good insights on the role of information literacy on the effectiveness of learning. However, very little studies demonstrate an effective implementation of programmes in virtual learning environments. The paper highlights how Beacon Primary School, one of the futuristic schools in Singapore, has implemented its Tamil Language programmes in a virtual learning environment thus providing a collaborative platform for pupils to meet and discuss issues. All our students, studying at the Primary four (P4) level (ten year old students), own their own personal learning device (notebook computer) which they bring to school daily. P4 Tamil curriculum and lesson packages are designed to infuse Information Communications Technology (ICT) meaningfully and make virtual learning a reality. Information literacy had been weaved into the P4 Tamil language curriculum with an online Web 2.0 software, wikispace, PBworkspace, as the platform for collaborative virtual learning environment. This paper presents how the virtual environment acts as a collaborative platform to enhance the pupil’s information literacy skills.

Keywords: Virtual Environment, Information Literacy Skills

1. Introduction & Purpose

Pupils are surrounded by a wealth of knowledge. Today, at the click of a button, students have access to events occurring anywhere on the globe within seconds of it happening. Given this scenario, it is critical that our pupils are equipped with the skill to connect, construct and relate the information presented. The virtual environment provides the space for collaboration amongst pupils. The virtual environment eases and enriches the process out of which meaning is derived from the multitude of information presented. The virtual environment also presents a knowledge-based forum for pupils to build on each others’ contribution.

Today’s educational system has to respond to two seemingly contradictory demands: On one hand, it has to effectively transmit constantly evolving knowledge and know-how to a knowledge-driven civilization. On the other hand, it has to enable learners with the right skills to select pertinent information out of the explosion of available information. It also has to ensure that the personal and
social development of the young learner is catered for. Therefore ‘education must ... simultaneously provide maps of a complex world in constant turmoil and the compass that will enable people to find their way in it’ Delors, J.,(1996) This translates to a shift in focus for the amount and level of content taught in schools. It also calls for greater emphasis on equipping our pupils with relevant skills to pick out relevant information. This forms the basis of the nation-wide initiative of ‘Teaching Less, Learning More’\(^6\). In today’s context, the ability to access, evaluate, organize and use information in order to learn, problem-solve, make decisions in a formal and informal learning contexts are an integral part of their learning. A key characteristic of the lifelong learner is strongly connected with critical and reflective thinking.

Information communication technological tools are constructive tools that provide a collaborative platform for pupils to come on board and build on each other’s knowledge. “Constructive tools are general-purpose tools that can be used for manipulating information, constructing one’s own knowledge or visualizing one’s understanding” Lim., C.P., & Tay, L.Y.,(2003). Jonassen, D. H., Carr, C. S., & Lajoie, S. P. (2000) purport the following constructivist approach- “ICT as mind tools for constructing evaluating, analysing, connecting, elaborating, synthesizing, imagining, designing, problem-solving, and decision-making.” The term “constructive” stems from the fact that these tools enable students to produce a certain tangible product for a given instructional purpose. This paper takes a reflective, narrative approach in documenting our attempts to integrate the virtual environment as a collaborative platform in enhancing pupils’ information literacy skills.

2. My Reflections

One of the key themes in the P4 curriculum revolves around the topic of ‘My Country’. The broad objectives include exposing students to the various issues that surround the country. The lesson design is tailored to educate on the various national issues, including the importance of tourism and consequently make logical connections to the implications and impact it poses to Singapore’s economic growth. The lesson was planned and carried out via the virtual learning platform as a collaborative platform for pupils to virtually meet discuss and develop their knowledge on the issue.

The discussion began from an article on Tourism from the Singapore local Tamil newspaper, Tamil Murasu. The teacher posted questions adopting the Blooms Taxonomy to scaffold pupils skills up to the different stages. Relevant links for extended learning was also provided. These links however, was in the English language. Pupils were instructed to explore these links independently and gather pertinent information. They were subsequently asked to present them coherently in the Tamil language.

Pupils were taken through three main stages:

1) Connect – refers to the understanding of the article/ information presented.

---

\(^6\) ‘Teach Less; Learn More’ (TLLM) is a call for schools and teachers to focus more on the active learning of students and the construction of their own knowledge.
2) Construct – refers to the pupils’ ability to comprehend the information, build on possible relationships and extend their knowledge and understanding from the information presented collectively in the platform.

3) Relate – relates to the presentation of collective information, analysis, synthesis, evaluation and creation of new perspectives from the issues presented. The following section details the activities conducted as part of each of the stages.

1) Connect:
   • Pupils were asked to highlight the keywords and use the mind mapping technique to identify all the important points in the article.
   • Each pupil is to contribute one finding from the article via online postings.
   • Pupils also verified their friends’ understanding of the article and their related thoughts.
   • If there was a misunderstanding of aspects in the article, the responsibility lay on fellow mates in the team to post a more accurate interpretation of the information.
   • The teacher acts as a facilitator to ensure that pupils connect with their ideas.

2) Construct
   • Pupils paraphrase, translate or give a short summary to express their comprehension of the article and the related issues.
   • In response to the questions raised, other members in the class contribute and build on one another’s ideas via the platform.
   • The pupils’ understanding of the content matter becomes apparent when they are able to identify relationships amongst ideas posted.
   • Pupils also tap on prior knowledge to build on these ideas.

3) Relate
   • Pupils are challenged with questions that require them to analyse available information and find logical patterns.
   • Pupils then evaluate the information and relate it to the current situation and seek new perspectives and understanding.

Pupils were observed to be very engaged and used the language appropriately. However, there were instances where pupils used English language to express their ideas, instead of Tamil. Although pupils were strongly encouraged to use Tamil language, weaker pupils who needed to resort to code-switching to express their thoughts, were not discouraged. The other pupils in the subsequent postings helped to translate these ideas. This created a win-win situation for pupils to tap on and maximise each others’ strength and to learn collaboratively.

As part of the school ICT program, pupils were introduced to search engines and were guided in searching for the relevant information. Pupils were also taught principles of cyber-wellness and exercised civic respect in contributing ideas and in providing feedback and comments in the online platform. The contributions of students to the discussed topic and the postings of links leading to
other related information was motivating. Even students who were less proficient in the language, displayed interest in contributing to the discussion. Their posting displayed the collective understanding of the various points contributed in the platform. As all pupils had to work with their own personal learning, the learning was seamless.

The second extensive discussion took place after Japan’s natural disaster. Pupils were exposed to this information during the morning assembly programme. As an extension an article from the newspaper was selected for online discussions. There was an intense discussion amongst pupils including the implications to the society and country. Pupils related the probable consequences. They were able to relate chain actions that would take place because of this disaster. Pupils used the Internet search engines to look up for latest update on the disaster such as on the British Broadcasting Corporation (BBC) news website. It was gratifying to note that the students took it upon themselves to update one another on the latest developments. In addition, they discussed and evaluated the situation and thought about the loss of those affected and the possible implications on their lives. It was heart warming to note pupils expressed concern and empathy for those affected.

3. Discussion & Conclusion

Technology is used as a constructive tool to facilitate pupils’ learning and making sense of their learning via a collaborative platform. Pupils’ engagement was evident throughout the discussion. They were critical about their contributions and took great responsibility in actively using the net to search for information to enhance their learning. The project had benefited even pupils, less proficient in the Tamil language, who was observed to be actively contributing ideas. There was sincere commitment on the part of the students. They also showed initiative in providing additional links and support for others to make sense of the issue. This helped to bring out the best in each pupil. Pupils in addition, expressed positive feedback. Every pupil contributed and has equal share in collaboratively constructing the knowledge, thus the ownership was very strong amongst them. This was a demonstration that young age is not a barrier in understanding world issues if it is tailored to meet the needs of the young learners. What really matters is whether pupils are equipped with skill to understand the implication and impact of the issue discussed.

In terms of skills, all pupils were able to sieve out and decipher the main points from the information presented and build on this information. Through this communication, it was observed that pupils had tapped on prior knowledge and experience in developing their alternative perspectives. Pupils learned to use the information and ideas presented in a graphical organising format to organise ideas. Pupils exhibited strong bonding and collaboration during the various collaboration sessions. The usage of technology was pervasive and as Breivik., P. (2000) puts it “Information literacy (is not)... teaching a set of skills but rather a process that should transform both learning and the culture of communities for the better.”

This paper is my attempt to share possible strategies in integrating information literacy into our daily lessons. It is through such sharing and exchanges where ideas could build upon ideas to further push the boundaries of our pursuit for pedagogical break-through in this fast changing world.

References


CONFERENCE PAPERS

TAMIL INTERNET 2011
இலங்கழிய கணினி வழி தமிழ் கற்கைத் தரும் கற்பிகள்

தமிழாசிரி பட்டமனை விளக்கம்

பிள்ளையார், புதுக்கோட்டை அருங்களாலகிய நூறாண், குறிப்பிட்டு

குறுக்குக்

ஆறியதற்கு அந்தந்த மான்றகாரரும் இருந்துகொண்டுள்ளச் செயற்பாடுகளுக்கு நன்கு பார்ப்பதற்கு தாத்துக்கு மிகவும் பிரித்துவித்துக்கு வேண்டும். அவ்வாறு பார்ப்பது மிகவும் பிரித்துவிக்கும் செயற்பாட்டுகள் குறிப்பிட்டுக்கு வேண்டும். இதன் காரணமாக இக்குறுக்கு, இக்கைறுக்கு தரும் நூறாண் கல்வி மற்றும் செயற்பாடுகள் குறிப்பிட்டுக்கு வேண்டும். அவ்வாறு வேண்டும்: 

- தமிழாசிரி பட்டமனை விளக்கம்
- புதுக்கோட்டை அருங்களாலகிய நூறாண்
- குறிப்பிட்டு

இதை நூறாண் வழியாக தமிழாசிரி பட்டமனை விளக்கம்

தமிழாசிரி பட்டமனை விளக்கம்

இலங்கழிய கணினி வழி தமிழ் கற்கைத் தரும் கற்பிகள்

தமிழாசிரி பட்டமனை விளக்கம்

அவ்வாறு வேண்டும்: 

- தமிழாசிரி பட்டமனை விளக்கம்
- புதுக்கோட்டை அருங்களாலகிய நூறாண்
- குறிப்பிட்டு

1 "காண்போர் என்று புதியான கற்கை விளக்கு... மட்டும் நூறாண் செய்த மீது மேம்பானாக"
2. "கொலைச் செய்ய வணிக வாய்ப்புகளைப்... குழாக மனிதக் காடம்" என்று ஓரையும் கூற்றுறுப்பிட்டன்

3. பிமைப்பு பட்டியல் - பிமைப்பு கருத்தறிகைப்படி இந்துதீவிகளால் மேலும் குறிப்பிட்டது வாய்ந்தெடுக்கும் குறிக்குண்டு, குறிப்பிட்டு வந்துவிட்டது.


தீவிகள் பாடல்

தீவிகள் பாடல் - தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீவிகள் பாடல்

தீevi..
காவிைதன். அவள் குலநிலைகள் காணி நிறை தோன்றிய வலுவராகக் கலைக்காரர் மென்னை கலைப்பாடுகள் நிறைய கலைப்பாடுகளை நாய்ப்படுத்தியன. இவ்வாறு செய்து பெறுவதுக் கலைக்காரர் வலுவராகக் கொள்ளும் விளக்கம் உண்டாதது.

முனைப்பாடுகள்

"மலரூவு செடி பாது தொலை
தையாரியும் போன்ற மார்த"

"விளையென பானேன் குழுவாராம் போட்டை"
(Education without Examination)  
Through e-learning

(புத்தாணத்துவம் அல்லது இளங்கற்களின் கல்வி)  
(Education without Examination)  
Through e-learning

இன்னொருத்தில் அடங்கியுள்ள பல்வேறு அறிவியல் கல்விகள், வருடா வருடாகத்துக்குள் பல்வேறு முறைகளைக் காணலாம். இவற்றுடன் குறிப்பிட்டு - கற்றுக்கொண்ட அடங்குதல் அறிவியல், செயலாட்டு, கலைமுறைகள், பெறுவது வாய் அடிப்படை அடிப்படையையும். நேருடைய அறிவியல் விளக்கங்களை விளக்கங்கள் பல்வேறு முறைகளைக் காணலாம்.

வாழ்விப் பல்வேறு முறைகளில், குறிப்பிட்டு அறிவியல் விளக்கங்களை காணலாம். ‘விளக்கங்கள்’ வாழ்விப் பல்வேறு முறைகளில் விளக்கங்கள். அறிவியல் விளக்கங்கள் வாழ்விப் பல்வேறு முறைகளில் விளக்கங்கள். 

இம்முறைகளின் பல்வேறு முறைகளில், குறிப்பிட்டு - கற்றுக்கொண்ட அடங்குதல் அறிவியல், செயலாட்டு, கலைமுறை. இவை கூடுதல் அறிவியல் விளக்கங்களை விளக்கங்களை விளக்கங்களையும் காணலாம். 

சின்னத்து வாழ்விப் பல்வேறு முறைகளில், குறிப்பிட்டு - கற்றுக்கொண்ட அடங்குதல் அறிவியல் விளக்கங்களை விளக்கங்களை விளக்கங்களையும் காணலாம்.
எனேவானி ஒᾞாறிᾺபி᾵ட, மேக்கவிகẓேவிைடா᾿ மகாணிவாகி அடேர்வையான அம்யூங்கை சமைந்து காணத்திறேன்; அரிதை எனேவளைய மன்பாடு காண்டு ஒன்றின் விளக்கத்திலும் கூறிய காலை உணவரியிலும்; கருவண்டிய குறுக்கு அமன்றாக உணவரியிலும். இங்குத் தவறைச் செய்து அதே உணவரியிலும் செய்யாளை கூறியது. உணவரியின் மின் தோற்றத் தீர்மானம் என்னுடைய வருடத்தில் உணவரியிலும்.

முடியும் குவித்து கல்வி செய்ய வேண்டும் என்று மேலும் செய்யத்திைடும். இந்தக் காலையிலும் மேலும் விளக்கம் செய்யாளை கூறியது.

மேலும் எனேவானி ஒᾞாறிவுத்து எனேவானி ஒᾞாறிவுத்து என்று மேலும் செய்யாளை கூறியது.

நிைனᾫていき திறᾂட அிவளவாகக் கூறிய மேலும் செய்யாளை கூறியது.

இᾺபᾊᾺப᾵டக᾿வியிᾞᾸதாய AXIS (Knowledge facilitator) எனேவானி ஒᾞாறா᾿ அதையான கூறியாளை கூறியது.

1. காண்டு குறிப்பிட்டாம்
2. காண்டு குறிப்பிட்டாம்
3. காண்டு குறிப்பிட்டாம்
4. காண்டு குறிப்பிட்டாம்
5. காண்டு குறிப்பிட்டாம்

ஆண்டத் துறதியை ஆண்டத் துறதியை என்று கூறும். குறிப்பிட்டாம் குறிப்பிட்டாம் என்று கூறும். அதையான கூறும் கூறும் கூறும். குறிப்பிட்டாம் கூறும் கூறும். குறிப்பிட்டாம் கூறும் கூறும்.

( BUILD IN QUALITY IN EDUCATION)
அங்கிய மாணவர்கள் கூட் ஆதிகாரியக் குழாய் எனும் பங்கு பெறுவதற்கு. அதனால் ஏற்று காண்பதற்கு மக்களுக்கு முகம்பம். பெரும்பாலும் பயிரிக்காகவுடைய ஆதிகாரியக் குழாய்கள், குழாய்களால், மேலும் பல்வேறு பங்கு பெறும் வகையான இடங்கள் தோன்றும்.

முதலியம் கடலியுடன் ஆதிகாரியக் குழாய்கள் - மாணவர்கள் நீக்கு குழாய்கள்.

புது முறையில்,

மாணவர்கள் விளக்கம் செய்யும்படி ஆதிகாரியக் குழாய்கள் நீக்கு குழாய்கள்.

வரும் மாணவர்கள்கள் விளக்கம் நிற்போருக்கு ஆதிகாரியக் குழாய்கள் நீக்கு குழாய்கள்.

முதலியம் முன்னேட்டு விளக்கம் ஆதிகாரியக் குழாய்கள் நீக்கு குழாய்கள்.

300 பாத்திரக் குழாயும் மாணவர்கள் மேலும் நீக்கு குழாயும் மாணவர்கள். 300 பாத்திரக் குழாய்கள் தோன்றும் புது முறையில்.

குறுக்கு விளக்கம் குழாயும்

குறுக்கு விளக்கம் ஆதிகாரியக் குழாய்கள்

உருவகையாக அமர்ந்த குழாயும் மாணவர்கள் நீக்கு குழாய்கள்

இலங்கை முன்னேட்டு விளக்கம்

இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள். இலங்கை முன்னேட்டு விளக்கம் குழாய்கள் நீக்கு குழாய்கள்.
மாணவி சொல்லுயிர் பாடல்:

தங்க அந்தோ மூலம் மூலமண்டன காண் தான் மறு இல்லையும் மூலமண்டனும். அது குறிப்பிட்டு என்ன போன்ற வருமானவை? மாணவி சொல்லுயிர் பாடல்?

மறு என்ற அந்தோ மூலமண்டன காண்டோ? அம்மாலோச்சொல் நேரமுற்று பாடல்? என்ன பாடல் நேரமுற்று?

தும் பார்க்கலோ என்று மூலமண்டன் காண்டோ! அதைக் கூறும் பயண நேரம்மூலமண்டன் என்ன பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?

மறு என்ற அந்தோ மூலமண்டன காண்டோ! அம்மாலோச்சொல் நேரமுற்று பாடல் நேரமுற்று பாடல் நேரமுற்று?
Tamil Video Retrieval Based on Categorization in Cloud

V.Akila, Dr.T.Mala
Department of Information Science and Technology,
College of Engineering, Guindy,
Anna University, Chennai
veeakila@gmail.com, malanehru@annauniv.edu

Abstract

Tamil Video retrieval based on categorization in cloud has become a challenging and important issue. Video contains several types of visual information which are difficult to extract in common information retrieval process. Tamil Video retrieval for query clip is a high computation task because of the computation complexity and large amount of data. With cloud computing infrastructure, video retrieval process has some scope and is flexible for deployment. The proposed method categorize the Tamil video into subcategories, splits the video into a sequence of shots and extracts a small number of representative frames from each shot and subsequently calculates frame descriptors depending on the edge and color features. The color histogram is computed for all the key frames based on hue, saturation and intensity values. Edge features are extracted using canny edge detector algorithm. The features extracted are stored in feature library in cloud. The features are tagged with Tamil text in cloud in order to satisfy Tamil query clip. Also, Videos are retrieved based on the Tamil audio information. The EUCALYPTUS cloud computing environment is setup within academic settings and the similarity matching of the Tamil video query is performed. The similar videos are displayed based on the similarity value and the performance is evaluated. Eucalyptus cloud platform is setup in Linux OS and the Tamil video retrieval process is deployed within the cloud. The efficiency of cloud computing technology improves the Tamil video retrieval process and increases the performance.

Keywords—video retrieval, categorization, cloud computing, Tamil query, Eucalyptus

1. Introduction

The need for intelligent processing and analysis of multimedia information has been increasing on a regular basis.

Researchers have found numerous technologies for intelligent video management which includes the shot transition detection, key frame extraction, video retrieval and more. Content based retrieval is considered to be the most difficult and significant issue of practical value amongst all the others. It assists the users in the retrieval of favored video segments from a vast video database efficiently based on the video contents. This paper aims at presenting the process of Tamil video retrieval in cloud environment. Video contains both visual and audio information. Audio contains natural language information which can be used to retrieve similar video content. The Tamil text processing is performed for user Tamil query.

The video retrieval system can be divided into two principal constituents: a module for the extraction of representative characteristics from video segments and defining a retrieval process to find similar
video clips from video database. A large number of approaches use a wide variety of features to symbolize a video sequence of which color histogram; shape information and text analysis are a renowned few. Application that requires a large number of computational resources might have to contact several different resource providers in order to satisfy its requirements. Cloud computing systems provide a wide variety of interfaces ranging from the ability to dynamically provision entire virtual machines. The feature database is stored in the cloud and the users query is compared. As based on cloud computing infrastructure, video retrieval process can be easily extended.

The rest of this paper is organized as follows: Section 2 deals with literature survey in the domain related to the project. It gives the different techniques adopted in the domain. Section 3 deals with system architecture. It includes detailed design of various phases involved in the project. It describes the internal working of the system. Section 4 deals with simulation and results of video retrieval process in cloud for Tamil videos. Section 5 deals with performance evaluation and result analysis. Section 6 focuses on conclusion and future enhancement.

2. Related Works

Nurmi describes the basic principles of the EUCALYP TUS design, that allow cloud to be portable, modular and simple to use on infrastructure commonly found within academic settings [3]. EUCALYPTUS is an open source software framework for cloud computing that implements what is commonly referred to as Infrastructure as a Service. It allow users the ability to run and control entire virtual machine instances deployed across a variety physical resources.

Takagi explains a method for video categorization based on the camera motion[5].Camera motion parameters in the video sequence contain very significant information for categorization of video, because in most of the video, camera motions are closely related to the actions taken. Camera motion parameters can be extracted from video sequence by analyzing motion information. Camera motion parameter has many advantages for categorization of video. Camera motion parameters like pan, fix are obtained using motion vector. Motion vectors are classified into 8 directions and histogram is calculated in each category. By analyzing characteristics of this histogram, camera motion parameters are extracted for each video [2].

The video shot segmentation system uses mathematical characterization of cuts and dissolves in the video [1].Different kinds of transitions may occur. An abrupt transition is found in a couple of frames, when stopping and restarting the video camera. A gradual transition is obtained based on effects, such as fade in i.e. a gradual increase (decrease) in brightness or dissolves i.e. a gradual superimposition of two consecutive shots. Abrupt transitions are obtained for two uncorrelated successive frames. In gradual transitions, the difference between consecutive frames is reduced. Considerable work has been reported on the detection of abrupt transitions.

A method for key frame extraction [6] which dynamically decides the number of key frames depending on the complexity of video shots and requires less computation. Priya and Shanmugam describe a method for feature extraction which provides the steps for extracting low level features [4]. The spatial distribution of edges is captured by the edge histogram with the help of sobel operators. Color histogram is the most extensively used method because of its usage in various fields. The color histogram value are recognized using hsv color space. Texture analysis algorithms are used in random
field models to multi resolution filtering techniques such as the wavelet transform. Several factors influence the utilization of Gabor filters for extracting textured image features. The feature library stores the extracted features.

3. System Overview

The architecture of our proposed system is shown in Fig 1. In the offline process, set of videos are given as input and features are extracted from the video. In the online process, the features are extracted from the query clip and matched against the feature vectors stored.

![System Architecture Diagram]

Fig 1  System Architecture

A. Video Categorization

The first process to be carried out is video categorization which is shown in Fig 2. The content based video categorizing method uses camera motion parameters. This parameter helps to categorize the sports videos for identifying different sports types. Camera motion parameters are changing the state among 2 types (Fix and Pan) along with time scale in video sequence. Here, motion vector are classified and histogram is calculated. By analyzing the characteristics of this histogram, camera motion parameters are extracted for each MPEG video.
In a video, panning is the sweeping movement of a camera across a scene and Fix means the static position of the camera. For this parameter, camera motion extraction ratio is calculated.

\[
    \text{camera motion extraction ratio} \, w[x] = \left( \frac{\text{Num.appear}}{\text{Num.total}} \right) \times 100\%
\]

where,
Num.appear -> number of times of an appearance for camera work x.
Num.total -> total number of frames in the given video.

B. Shot Segmentation

To segment the shots, the video has to be split into video shots prior of conducting any video object analysis. Scene change detection, either abrupt scene changes or transitional (e.g. dissolve, fade in/out, wipe) is employed to achieve the video shot separation.

C. Key frame Extraction

A key frame is a frame that represents the content of a shot. This content is the most representative one. In the large amount of video data, first reduce each video to a set of shots and find the representative frames. Each shot obtained by video segmentation algorithm contains a set of frames.
These segments are represented by two dimensional representative images called key frames that greatly reduce amount of data that is searched. Key frames from each shot are obtained by comparing the color information between adjacent frames. A frame will be chosen as key frame if the value exceeds certain threshold.

D. Feature Extraction

Feature extraction is an area of image processing which involves using algorithms to detect and isolate various desired portions of a digitized image or video stream. Different kinds of video features, including edge and color for each key frame is being extracted. To minimize the dimensionality of the data, feature extraction is employed which extracts discriminative features of data.

Fig 4 shows the process of color histogram creation. Color histogram is the most extensively used method because of its robustness to changes due to scaling, orientation, perspective, and occlusion of images, which are recognized by using the HSV color space.

Edges in the key frames are detected based on the canny edge detector. The Canny operator works in a multi-stage process. First of all the image is smoothed by Gaussian convolution. Then a simple 2-D first derivative operator is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image.

E. Similarity matching

The query video is categorized and key frames are extracted. The color and edge features extracted are matched against the features in the repository. The color features are matched based on the naive similarity algorithm and edge features are matched based on region based histogram.

The algorithm first calculates the color histogram for the query clip and compares with the video set. Each key frame feature vector of query clip is matched with all the feature vectors in the repository and most similar match is retrieved. The histogram values contain mean, entropy and standard deviation of color. From the mostly matched key frames the edge histogram is calculated and matched against query clip. The edge histogram contains region information. The key frames which give the most similarity values are selected and the corresponding videos are retrieved as the similar videos for user query clip.

F. Audio Processing

The next way of Tamil video retrieval focuses on audio processing. The audio track is extracted from the Tamil video as the first step. The audio files are segmented in order to remove the silence and noise. The audio files of each video are processed and the words are extracted and stored as .wav files.
These .wav files are called as features of the audio content.

The user gives the query Tamil video clip as input. This input file contains both audio and video information. The audio data will be segmented to remove silence and extract key words. These key word files are pattern matched against all the .wav files in the feature set. The most matched patterns are found and the corresponding videos are extracted.

The pattern matching of wav files are performed and the results which exceed certain threshold are taken as the result.

**F. Text Processing**

The next way of video retrieval is based on Tamil text. The wav files of audio input are chosen and are tagged with Tamil text. The user input of Tamil text is transliterated and is searched against the feature set. The matched results corresponding video are retrieved and given as result to user. Transliteration is the practice of converting a text from one language into another language phonetically. Transliteration is different than translation. The Table 1 shows some transliterated English word for tamil word.

<table>
<thead>
<tr>
<th>Tamil word</th>
<th>Transliterated English word</th>
</tr>
</thead>
<tbody>
<tr>
<td>காணம்</td>
<td>Kadinam</td>
</tr>
<tr>
<td>புக்கல்</td>
<td>Pookkal</td>
</tr>
<tr>
<td>குழந்தை</td>
<td>Kuzhandhai</td>
</tr>
<tr>
<td>பாப்பா</td>
<td>Paappa</td>
</tr>
<tr>
<td>மழை</td>
<td>mazhai</td>
</tr>
</tbody>
</table>

*Table 1: Transliteration of Tamil to English*

**F. Cloud setup**

Eucalyptus is an open source cloud computing system.

The eucalyptus open source cloud environment is setup in Linux cluster. The eucalyptus software is installed. The cloud controller, cluster controller, walrus and storage controller are installed.

The cloud controller is the entry point into the cloud for users and administrators. It asks node managers for information about resources, makes scheduling decisions and implements them after requesting to cluster controller.

The cluster controller executes on a cluster front end machine, or any machine that can communicate to both the nodes running Node controllers and to the machine running cloud controller.
controllers gather information about and schedules virtual machine execution on specific node controller and also manages virtual instance network.

The Node controller is executed on every node that is used for hosting virtual machine instances. They control the execution, deployment and termination of virtual machine instances on the host where it runs.

The storage controller is capable of interfacing with various storage systems. It is a block device and it is attached to an instance file system. Walrus allows user to persistent data, organized as buckets and objects. It provides a unique mechanism for storing and accessing virtual machine images and user data.

The Tamil video retrieval process is developed as an application and this application is bundled to the virtual machine instance. The application bundled virtual machine image is uploaded and registered to the eucalyptus cloud. The instances are communicated and the application is run over the cloud. The query video clip is given as input in the cloud front end. The videos are categorized, the key frames are extracted and the similarity search is performed in separate parallel instances. The retrieved video result are given as output to the user.

4. Simulation and Results

The video retrieval process includes video categorization, key frame extraction, feature extraction and similarity matching. The process is carried out in Java media framework and Java advanced imaging.
Fig 6 shows the key frame extracted for a given Tamil video and Fig 7 shows the categorization and similar video result.

5. Performance Evaluation

The user gives the query video name as input and based on the commands the videos will be categorized, extracts key frames and features. The similar video will be retrieved if they give search command.

The video retrieval process is performed in cloud and the performance is evaluated while running in two instances.
Table 2: Relation between execution time in one instance and two instance

<table>
<thead>
<tr>
<th>No. of videos in dataset</th>
<th>Execution time in 2 instances</th>
<th>Execution time in 1 instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5206.2</td>
<td>18369.69</td>
</tr>
<tr>
<td>15</td>
<td>5522.63</td>
<td>18924.41</td>
</tr>
<tr>
<td>20</td>
<td>6202.2</td>
<td>21731.45</td>
</tr>
<tr>
<td>25</td>
<td>7291.7</td>
<td>25319.52</td>
</tr>
<tr>
<td>30</td>
<td>8575.6</td>
<td>28904.35</td>
</tr>
</tbody>
</table>

The application is run in EUCALYPTUS private cloud and the execution time is calculated while running in single instance and two instances. The execution time is much less when we run in two instances. This shows that the video retrieval process shows better performance in cloud.

![Fig 8 Performance graph in cloud environment](image)

The performance of video retrieval process is checked by precision recall graph.

Recall = \( \frac{DC}{DB} \) and Precision = \( \frac{DC}{DT} \)

Where DC is the number of similar clips which are detected correctly, DB is the number of similar clips in the database and DT is the total number of detected clips.

Table 3: Precision and recall for query video clips

<table>
<thead>
<tr>
<th>Query video</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Q2</td>
<td>0.35</td>
<td>0.78</td>
</tr>
<tr>
<td>Q3</td>
<td>0.39</td>
<td>0.69</td>
</tr>
<tr>
<td>Q4</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Q5</td>
<td>0.8</td>
<td>0.2</td>
</tr>
</tbody>
</table>
The precision and recall for various query video clips are computed. The efficiency of the video retrieval process is improved as the retrieval process includes categorization process.

![Performance graph for video retrieval](image)

Fig 8 Performance graph for video retrieval

The performance of Tamil video retrieval shows that the most similar videos are retrieved. Also the application in cloud environment shows that the cloud computing provides better performance through execution time and resource sharing.

6. Conclusion and Future work

The proposed video retrieval categorizes the video into different category based on camera motion parameters. It facilitates the segmentation of the elementary shots in the video sequence proficiently. Then the key frames are extracted from the video shots. Subsequently, the extraction of features like edge and color histogram of the video sequence is performed and the feature library is employed for storage purposes.

Then Video retrieval system based on query video clip is incorporated within the cloud. Cloud computing, due to its high performance and flexibility, is under high attention around the industry and research and reduces the computation complexity of Video retrieval process based on visual, audio and text input.

References


Animated Story Visualizer for Tamil Text

M. Janani, Dr. T. Mala
Department of Information Science and Technology
Anna University, Chennai, India
janani_mrl@yahoo.com, malanehru@annauniv.edu

Abstract

Natural language is a straightforward and efficient medium for describing visual facts and mental images. The System uses a novel approach to generate animation from Tamil texts such as stories. Tamil text is pre-processed and the necessary features like named entities, environmental constraints, temporal and emotion constraints for the given stories are extracted and placed in the database. The system automatically generates a query based on the users input and compare it with features stored in the database. Finally animation is dynamically generated using an external motion synthesis system. Using this system, even greenhorn users can generate animation quickly and easily by giving the Tamil text.

Keywords — Computer animation, Natural Language Processing, Pre-processing, Feature Extraction, Motion synthesis

I. INTRODUCTION

These days, animations are widely used in many applications, such as cartoons, web graphics, games, and so on. Computer animation is one of the best methods for depicting the dynamic content. A medium is necessary for the animation to be created in a convenient and natural manner. It should be possible to describe the scenes directly from natural language.NLP is an easy and effective way to analyze, understand and generate languages that humans use naturally.

The aim of this work is to generate an animation from Tamil texts such as movie scripts or stories. Training input text is given to the pre-processing module. Here tokenization is performed and the tokens are given to the morphological analyser which is used to convert the tokens into a POS tags. Information related to named entities, temporal constraints, emotion and environment inference features are extracted. A query is generated automatically from the input text which contains information for the search process and compares it with the information already stored in the database. Finally motion synthesis generates an animation. The interactions between characters are handled by this module based on the information provided in the database.

II. RELATED WORK

Generating animation from natural language texts has been a challenge. WordsEye developed by Coyne and Sproat [1] converts natural language texts to a scene. WordsEye focuses on generating a still image, when a character motion is specified in a given text, the system simply prefer to pose for
the action generated from the database. The Carsim system [2] describes a new version of text-to-scene converter that handles texts describing car accidents using computer program and it is visualized in the 3D environment. Storytelling System [3] illustrates a system called Interactive e-Hon, which provides storytelling in the form of animation and conversation translated from original text. A Constraint based scene conversion system [4] describes a Text2Scene conversion method which automatically converts text into 3D scenes. A large database of 3D models is used by this method to depict entities and actions.

III. SYSTEM OVERVIEW

In this section, overview of our system (Figure 1) is given, where the major components are identified. When the Tamil text is given to the system, Tamil text is preprocessed and the information are extracted and stored in the database along with the objects created. When an input text is given to the system it automatically generates the query from the input text and compares it with the information stored in the database. An animation is then generated using an external motion synthesis system.

![Figure 1. SYSTEM OVERVIEW](image)

**PRE-PROCESSING**

When the Tamil text is given to the system, natural language processes (Tokenization and morphological analysis) are applied first.

**Tokenization**

The first step in NLP is to identify tokens, which decomposes the delimiters like punctuation and whitespaces. Here Tamil text is given as the input to the tokenizer which breaks the text into meaningful tokens. The tokens generated by the tokenizer are passed to the analysis engine.

**Morphological Analysis**

RCILTS [5] developed a tool called Atcharam, an analyser which performs Morphological Analysis for Tamil text. The Morphological analyser takes a derived word as input and separates it into root word and associated morphemes. It is the basic tool used in spell checker, grammar checker, parser and machine translation systems. It has two major modules noun analyzer and verb analyzer.
Tamil Noun and Verb classification example

By this method morphemes are generated and given to the learning process where the necessary informations are extracted.

INFORMATION EXTRACTION

OBJECT IDENTIFIERS

Object Identifiers recognize named entities in text by Named Entity Recognition (NER). “Rule based approach” is used to extract named entities from the given text. Initially, root words say Noun, verb, adjective, pronoun, adverb from text file are extracted. Rules are created based on prefix and postfix of noun, i.e. noun that occurs between verb and noun, noun that occurs between noun and noun, noun that occurs between noun and verb and so on. If any of the above rules satisfies the input text named entities are extracted. Here is an example,

Input: ஒᾞ ஐᾑளᾛ้าதி᾿ எᾠάௐvrolet் "Given input text is pre-processed and the root words are extracted.

emsp;எᾠάௐvetica< Noun>

emsp;எᾠάௐvetica< Noun>

emsp;ெேௐ.setTextColor< Verb>

Now the rules are applied to this extracted root words. Here எᾠάௐvetica comes between noun and verb which satisfies the rule is extracted.

TEMPORAL AND EMOTION EXTRACTION

Temporal reasoning in NLP involves extraction, representation and reasoning with time and events in the natural language text. Here to extract temporal constraints, “manually created dictionary” is used. The root words are compared with the manually created dictionary and temporal constraints are extracted if the input text satisfies the inferences present in the dictionary. Similarly different emotion present in the text is also extracted using manually created dictionary.

Figure 2 shows the different emotional constraints to be depicted.

<table>
<thead>
<tr>
<th>HAPPY</th>
<th>இம்பம்,மைத்தில்,தொச்சம,சுப்பேந்தை,முட்டாீர்,கல்லூரி,வில்லாீயம்,தீம்பை</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANGER</td>
<td>அக்குதார்,சின்,ேகாப்,ெவ asshole,எாிைச்,சீயற் தாப்</td>
</tr>
<tr>
<td>SURPRISE</td>
<td>அதிசய்,அ்யுசாிய்,பிரமி்ழ்,மைல்,விய்௔ீ்</td>
</tr>
<tr>
<td>FEAR</td>
<td>அோச்,பய்,ெபா௔ீம்,விதி்ழ்,கவைல்,அாீ,கலல்க்க் கைலா</td>
</tr>
<tr>
<td>SADNESS</td>
<td>இசாக்,அைக்,ேசா்ைத்,ஐயர்,வ்ூைத்,ஐீப்,ாீீ,அலீ,கன்,அ்யுத்,அீத்,பய்,ாீட்,மைல்,விய்௔ீ்</td>
</tr>
</tbody>
</table>

Figure 2 Emotion Constraints

120
Finally, environmental constraints that specify location and actions are extracted.

**ANIMATION GENERATION**

Movie clips for the extracted information are created using Adobe Flash professional and stored onto the database.

**CONSTRAINT MATCHING**

When an input text is given, information constraint should match with the information present in the database to generate animation. String matching algorithm is used to compare the information extracted and information stored in the database. Let $P[1...M]$ and $T[1...N]$ be the character array for the given string. Pattern $P$ is said to occur with shift $s$ in text $T$. To find all valid shifts or possible values of $s$ so that $P[1...m] = T[s+1..s+m]$ ; There are $n-m+1$ possible values of $s$.

Procedure String Matcher($T,P$)
1. $n ←$ length $[T]$;
2. $m ←$ length $[P]$;
3. for $s ← 0$ to $n-m$
4. do if $P[1..m] = T[s+1..s+m]$
5. then shift $s$ is valid

Find first match of a pattern of length $M$ in a text stream of length $N$.

The extraction of Pattern $\text{attribute value}$ is done by,

$\text{attribute value} \ M = 4$

$\text{attribute value}$

$\text{attribute value}$

$\text{attribute value}$

$\text{attribute value}$

$\text{attribute value}$

$\text{attribute value}$

By this method exact string is matched from the database for the given information.

**STORY ASSEMBLER**

Storyboards are the only way to convey rich information, viewing a particular order of events in a most appealing way. Basically the system searches for noun and verb from the given input text then automatically assemble and analyse the subsequence like background, named entities, temporal, emotion and action movie clip from a database that matches the constraints.

**MOTION SYNTHESIS**

Animation is generated by motion synthesis by efficiently connecting the movie clips that are assembled by the story assembler from the database. The character and objects interactions are handled by this module based on the information that the movie clips have. Timeline specifies what
kind of action occurs at particular time. Once the timeline has been set, animation is generated for the
given Tamil text. Figure 3 shows the animation generated for the given Tamil Text

Figure 3 Animation is generated from Tamil text

PERFORMANCE ANALYSIS

The performance analysis is used to monitor the functioning, efficiency, accuracy and other such
aspects of a system. For the analysis performed for the learning process, the overall accuracy obtained
is 83%. Figure 4 shows the Performance analysis for Animation generated from the Tamil text.

<table>
<thead>
<tr>
<th>Viewer</th>
<th>No of Stories</th>
<th>Exactly generated</th>
<th>Approx generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>viewer1</td>
<td>10</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>viewer2</td>
<td>10</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>viewer3</td>
<td>10</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>viewer4</td>
<td>10</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>viewer5</td>
<td>10</td>
<td>14</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 1 Test case for Animation Generation

The overall accuracy obtained for the generation of animation is 80%. The performance can be further
improved by generating rules and optimizing the learning process.

CONCLUSION AND FUTURE WORK

The system provides automated generation of animation from Tamil text which provides a new
approach for users to create animation quickly. The proposed method takes Tamil text as the input and
it is pre-processed and features like named entities, temporal constraints, emotion and environmental
constraints are extracted and animation is generated dynamically by motion synthesis. Even non-
professional people can rely on this system and they can generate animation quickly and easily by
giving the Tamil text. Future work can be extended by generating animation via automatic speech
recognition rather than text.
REFERENCES

Popularity Based Scoring Model for Tamil Word Games

chezhiyank@gmail.com, tv_g@hotmail.com, rp@annauniv.edu, madhankarky@gmail.com
Tamil Computing Lab (TaCoLa),
College of Engineering Guindy, Anna University, Chennai.

Abstract

In this paper we propose a Popularity based Scoring model for Tamil word games. Games are one of the effective means to teach a language. There exist very few online word games for Tamil. Scoring is one of the key aspects of a game that nurtures interest in the player apart from the interface and logic. The Popularity Based Scoring Model proposed in this paper, uses a word statistics crawler to periodically collect the statistics of word usage in popular blogs, news articles and social nets. The popularity of every word is thus modeled in comparison with every other word in the language. The model was successfully implemented in a simple unscramble game titled ‘Miruginajambo’. Over three lac root words from Agaraadhi Project were crawled for statistics and 20000 words were obtained for the game based on threshold levels for increasing levels of complexity in the game. The paper concludes providing the results and analysis of implementing the model and also discusses various word games where this model can be used.

1. Introduction

Word based games can serve as an effective tool to teach vocabulary in any language. The complexity levels, the score achieved motivates a player to play more and there by learn more words. One key challenge in designing such games is the scoring model. A scoring model for a game means a lot more than just a value associated with the game level or complexity. An effective scoring model has to motivate a player to play more and there by retain the player’s interest to come back again.

Tamil language has very few online games available online. These games are mostly flashcard-based games. With new words being introduced in various domains such as medicine, computer science and other disciplines, such games can be the most effective way to teach words. The main reason for popularity of word games in English is their scoring models apart from the user interface they build around their games.

In this paper we propose a popularity-based scoring model for word-based games in Tamil. Providing this scoring model we test the scoring model over two games namely ‘miruginajambo’, a unscramble game and ‘thookkuthookki’, a Tamil equivalent of hangman. Popularity Based Scoring Model generates score based on the combination of the word’s popularity, length of the word and Tamil alphabet popularity.
This paper is organized into four major sections. The following section gives the background about scoring models and other relevant literature. The third section gives the popularity based scoring model and the components of the scoring model. The final section concludes discussing the results.

2. Background

A scoring model calculates scores based on performance of any system on various domains like Credit Scoring Model in banking application, Fuzzy Logic Approaches in game etc. In the new generation mobile multiplayer games, scoring was generated by using Fuzzy Logic Approach [1]. Automatic Target-scoring System of Shooting Game Based on Computer Vision [2], Online Score System Using Hierarchical Colored Petri Nets [3] to evaluate the outer and inner performances of the system, such as scan, score, and resource utilization. The Credit scoring models [4] are developed to classify the loan applicants as accepted or rejected. The decision is based on the information of each applicant such as age, income and debit ratio. First time we proposed The Popularity Based Scoring Model proposed for Word games, uses a word statistics crawler to periodically collect the statistics of word usage in popular blogs, news articles and social nets. This word Popularity was implemented in the Agaraadhi Online Dictionary [5]. The Word from agaraadhi is given to web and found the frequency distribution of the word across the popular blogs, news articles, social nets etc. The Scoring Model uses the Frequency Analysis of Tamil Alphabet [6]. The Frequency Analysis is done by splitting the Tamil word into alphabet, splited alphabet are added to their corresponding counter, frequency of each alphabet was identified individually.

3. A Game Framework based on Popularity-Based Scoring Model

We propose a simple game framework for an unscramble game in this paper, depicted in figure1. The framework can be basically divided into two major divisions, online and offline, in terms of the time of processing. This section describes the various scoring generator in detail.
3.1 Offline Processing

The Offline Process takes a Set of Tamil words as input, Word Popularity Scoring, and Tamil Alphabet Frequency Analysis are the key tasks here.

3.1.1 Bag of Tamil Words

Tamil words are obtained for analysis from the Agaraadhi project comprising over 3 lac words on various domains such as General, Engineering Technology, Literature, Medicine and Computer Science.

3.1.2 Word Popularity Scoring

Word Popularity shows the word usage in the web. The words from agaraadhi are crawled over popular news sites; blog articles and social networking sites periodically and the frequency distribution of the word across sites are identified and recorded. This overall information is then used to compute the popularity score for each word.

3.1.3 Tamil Alphabet Frequency Analysis

The words obtained from the set of words are split into alphabet that constitute the word and the split alphabets are added to their corresponding counter, frequency of each alphabet is identified individually. Alphabet Frequency Analysis generates score based on the frequency value of the alphabets. This Analysis result will be used later to find the popularity of a letter and thus the complexity of a word. So, a low frequency alphabet contained word gets a higher score.

3.2 Online Processing

The Online Process Comprises of Random word Picker, Word Jumbler and Popularity Based Score Generator

3.2.1 Random word Picker

Random word Picker fetches a random word based on the domain specified by the user and the Word Popularity. It will be possible now to decrease the word popularity score on every higher level to increase the complexity of the game.

3.2.2 Word Scrambler

Word Scrambler module scrambles the Tamil alphabets in a word such that all alphabets are not placed in their actual correct spots. The jumbler is randomized such that the next time the same word will be jumbled in another combination.

3.2.3 Popularity Based Score Generator

Score Generator generates the score based on the proposed Popularity Based Scoring Model using the parameters such as word popularity, level of the game, Low frequency scored occurrences of Tamil Alphabets in a word, total number of swaps to complete a level and time taken to complete a level. Let $w$ be the word, $l_w$ be the length, Let $P_w$ be the popularity score of the word, Let $P_a$ be the average alphabet popularity frequency of the alphabets in $w$, Let $t$ denote the time taken to solve the word in
seconds, Let s denotes the number of swaps needed to solve the word, Let $\alpha$ be the score for a perfect answer.

The score for a solving,

$$\text{Score } S(w) = \alpha \times (1-P_w) \times (1-P_s) \times (1-\frac{\hat{w}}{p}) \times (1-\frac{\hat{w}}{l})$$

4. Results and Discussion

The framework depicted in figure 1 was implemented with a simple web interface. The snapshots of the working game with the popularity based scoring model are given in figure 2. The same game was developed with and without the popularity based scoring model with the earlier version giving a score just based on the level and number of letters swapped. The version with popularity based scoring model was found to receive more users playing for a longer time and repeatedly as they find their current score rapidly increase if they identify a less popular word.

5. Conclusion

In this paper we proposed a Popularity-Based Scoring Model for computer based word games in Tamil. The popularity of words was identified by their usage over the internet by news, blog and micro blog writers. This information is then converted to scoring every word in a dictionary. This score is used in the Scoring model to compute the score for the user at different levels. The scoring model is compared to a traditional level based scoring model. Analyzing user behavior over the two models we conclude that the popularity based scoring model creates more interest in user to play the game for long time and repeatedly compared to the traditional level based scoring.

Fig 2: Snapshots of the working game with the popularity based scoring model
References


CONFERENCE PAPERS

TAMIL INTERNET 2011
Abstract

Automatic classification of web document increases in the regional languages, because of amount of information available in the regional languages (like Tamil, Telugu, Hindi) is huge in the internet in the form of e-Book, news, articles and other type of formats. It is difficult to categorize those documents based on the subject of interest. Tamil is a Rich Dravidian language, it have a millions of documents in the Web Repository, due to growth of digital documents, categorization needed to classify document. Too much classification techniques are present for the English documents classification like SVM, K-NN, Decision trees, Neural Network technique, but classification in regional languages like Tamil, it’s new and emerging. So that our proposed work involves first, genetic algorithm will be employed to reduce dimension of document. Second, Multilingual Cross- domain classification, involves the predefined labels in the English language will be used to classify the Tamil Corpus, because pre-defined labels in the source domain is expensive to create, so that look for other domain of same interest to classify the documents. Third Back Propagation Technique applied to classify Corpus.

Key Terms: Classification, Multilingual, Cross-Domain, Dimension Reduction

Introduction

Today most of the documents exist in the electronic repository like e-books, journals, news articles and other sources of information in form of English only. This electronic document exits in other regional languages also (like Tamil). To classify those Region documents lot of research going on.

Tamil\(^{[1]}\) is an oldest regional language present in the world. Around billion of people speaking Tamil and lot of documents present in the Tamil language. Natural Language processing of Tamil is difficult, because of little bit research is taken place. To analyze their keywords, linguistics plays an important role. Lot of research already taken place to classify English documents based on supervised and unsupervised learning. I.e. two types learning is their (i) supervised leaning means of classification documents based on the pre-defined label categorization. It first train the training document based upon the pre-defined labels, and test the test documents and classify based upon the training set. (ii) Unsupervised learning is a clustering.

Many machine learning technique available like SVM, KNN Classifier, Neural Network, Bayesian Classifier based on mathematical approaches. For Pre-label is expensive, to avoid that other domain
label is used for classification purposes, but in English document collection lot of Cross-Domain\textsuperscript{[10]} Labels are available, in order to classify the documents in other domain but in Classification based on rare. So that use labels in the domain of English Document, to the corresponding labels in Tamil documents, it reduce the classification effort, and also produce better results.

Before using those approaches dimension reduction plays an important role to minimize the no of keywords present in the document. Our proposed approach use the genetic algorithm for reduction of no of key attributes present in the documents.

Dimension reduction carried out based on feature selection and feature reduction methods.

Feature selection\textsuperscript{[2]} means that, it selects the keywords based on attributes, which contribute reduction of no of words in the documents. Selection plays an important role here. Two types present (i) filter method Separating the feature selection from the classifier learning, and relay on general characteristics of data, no bias over any learning algorithm, generally it fast. (ii) Wrapper model, relaying on predefined classification algorithm, and computationally expensive.

Genetic algorithm\textsuperscript{[3]} will be used as a dimension reduction technique, it takes the set of keywords as a population of terms, and neural network will be employed as a classifier, which train and classify the training documents and classify testing documents after that training phase.

Tamil corpus will be generated automatically, by using a web crawler. Crawler return the set of document pages (Tamil) particularly news articles. These collectively articles used to form the corpus. Further classification will be done using those Tamil news articles (Corpus).

This paper section 2 describes the web crawler section 3 describes the Tamil corpus, section 4. describes the dimension reduction using the genetic algorithm, section 5, describes the classification using neural network.

2. Web Crawler

Crawler is a software program, which can fetch the WebPages based on the seed URL given to the Crawler. Here seed URL will be “Tamil news article” site URL. This crawler crawls only site given to the input to the crawler, it doesn’t navigate to other site. It uses a muti threaded downloader to down load the web pages , based on that seed URL given to the crawler, this crawler, crawls the pages only within that link. Suppose it should news.goole.com means, it crawl the link fully, retrieve the document within that.

![Figure 1. Architecture of Web Crawler](image-url)
3. Tamil Corpus

Many research going on to build a corpus in Tamil. Central institute of Indian language (CIIL)\(^4\), Mysore actively involved in building the corpus in different regional languages. Here, corpus build by using a web crawler, it crawl a web pages and stored it in a local database. After that it will be edited in order to make and formed as a corpus. Tamil has 12 vowels and 18 consonants. This are combined with together 217 composite characters and 1 special characters counting to the total of 247 characters. To build a corpus for that rich type of grammar is too difficult. So that crawler will used to retrieve web content, and it edited to form a corpus.

4. Dimension Reduction using genetic algorithm

Generally dimension reduction used to reduce the no of words in a corpus. Because corpus have a huge collection of words, but few collections of words in a corpus makes the document meaningful, So that to identify those words, dimension reduction plays a vital role. Genetic algorithm used as an optimization technique. Here it plays as a dimension reduction, it choose a set of attributes like content name, sub-content title, and other. Genetic algorithm uses a input as a set of population of attributes, instead of choosing a single attributes, it reduce the no of words from a thousand to hundred, each attribute like a gene, group of attribute forms a chromosome, uses a various operation like,

1. Crossover: single or multi-point
2. Mutation
3. Reproduction

4.1 Algorithm of GA for dimension reduction

Step1: form the set of attribute as a chromosome.
Step2: generate the fitness function for each gene in the population.
Step3: apply the genetic operator, and evaluate the fitness once again.
Step4: stop, if attain the terminating condition, else generate new population and go to step2.

5. Classification of Web Document

After the Identification of set of key attributes, need to classify the training documents using a neural network. Neural network\(^9\) have a set of input nodes, and hidden nodes, and a corresponding output nodes. Training documents taken as a input to the system, that should be trained and classified accordingly based on the attributes generated by the genetic algorithm, theoretically says that, dimension reduction after that classification improve the result future.

Figure 2. Architecture of Neural Network.
Back propagation technique employed in the classification of web documents. Feedback given to the neural network with every set of documents trained. After that documents tested against the network, whether it should be classified correctly. Theoretical performance is better than other classification technique.

**Conclusion and Future Work**

Automatic classification of Tamil web content increase the need for separate classification approaches, for that genetic algorithm employed as a dimension reduction technique, and classified accordingly based on the selected attributes, it improve the precision and recall after the dimension reduction.

Future improvement in the neural network, will be use of winnow/preceptor technique with no hidden layer improve classification technique.

**References**

- Philomina Simon, S. Siva Sathya, Genetic Algorithm for Information Retrieval
- M. Ganesan, Tamil Corpus Generation and Text Analysis
- M. Selvam, and A. M. Natarajan, Language model adaptation in Tamil language using cross-lingual latent semantic analysis with document aligned corpora, CURRENT SCIENCE, VOL. 98, NO. 7, 10 APRIL 2010
- Thair Nu Phyu, Survey of Classification Techniques in Data Mining, Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I,IMECS 2009, March 18 - 20, 2009, Hong Kong
On Emotion Detection from Tamil Text

Giruba Beulah S E, and Madhan Karky V
Tamil Computing Lab (TaCoLa),
College of Engineering, Anna University, Chennai.
(ljcsegb@gmail.com) (madhankarky@gmail.com)

Abstract

Emotion detection from text is pragmatically complicated than such recognition from audio and video, as text has no audio or visual cues. Techniques of emotion identification from text are, by and large, linguistic based, machine learning based or a combination of both. This paper intends to perceive emotions from Tamil news text, in the perspective of a positive profile, through a neural network. The chief inputs to the neural network are outputs from a domain classifier, two schmaltzy analyzers and three affect taggers. To aid precise recognition, tense affect, inanimate/animate case affect and sub-emotion affect dole out as the supplementary inputs. The emotion thus recognized by the generalized neural net is displayed via a two dimensional animated face generator. The performance and evaluation of the neural network are then reported.

Index Terms—Emotion detection, Machine learning, Neural network, Tamil news text.

I. Introduction

Emotion detection from text attracts substantial attention these days as this if realized, could result in the realization of a lot of fascinating applications like emotive android assistants, blog emotion animators etc. However, emotion detection from text is not trouble-free, as one cannot, with a single glance get a hold of the emotions of the people about whom the text is based. Further, what appears to be sad for a person may perhaps appear fear for someone. Emotion detection from text is thus influenced by the empathy of the readers. Also, as there may be no background information, to shore up the emotion of a person in text, it is better if emotion detection is based on some model empathy.

The aim of this paper is to detect emotion from Tamil news text by a self learning neural network which takes in linguistic and part of speech emotive features. The emotion is identified by assigning weights for features based on their affective influence.

II. Background

Tamil is a Dravidian language as old as five thousand years and enjoys classical language of the world status along with Hebrew, Greek, Latin, Chinese and Sanskrit. Unlike Sanskrit, Latin and Greek, which are very rarely in use, it is a living language and has fathered many Dravidian languages like Malayalam, Telugu etc. It is morphologically very rich and has a partial free word order. It groups noun and verb modifiers, (adjectives and adverbs) under a single category, Urichols. Noun participles are equally affective as the verb participles. Thus, apart from parts of speech, morphological entities like cases and participles are also affective.

Among the methods of emotion detection, [1] observes that the Support Vector Machines using
manual lexicons and Bag Of Words approach perform better. The Support Vector Regression Correlation Ensemble is an album of classifiers, each trained using a feature subset tailored to find a single affect class. However, as support Vector machines suffer from high algorithmic complexity and requirement of quadratic programming and do not efficiently model non-linear problems, Neural networks has been opted as they provide greater accuracy than Support Vector Machines.

Nevertheless, Neural networks sport disadvantages like local minima and over fitting. The former can be handled by adding momentum and the latter is usually solved by early stopping. Since the neural net could memorize all training examples if over the need hidden neurons are present, three promising prototypes are constructed, trained and tested to find the optimal one. To this optimal neural net, the affective feature tags from Tamil news text are given, to recognize the inherent emotion, based on their respective affective strength.

Kao, Leo, Yahng, Hsieh and Soo use a combinatory approach of dependency trees, emotion model ontology and Case based reasoning [2] where cases are manually annotated. Such an approach may work for languages of few cases. However, not all cases in a language may be affect sensitive. Cases in Tamil are eight and only two of these are affect sensitive, namely the instrumental and the accusative case. Thus manually annotating cases in Tamil would normally fail as cases aid in increasing or decreasing the affect of the verb nearby.

The separate sub-class networks [10] for Emotion recognition with context independence in speech seem to be promising but for the low precision of 50 even when the learning is supervised using a simple backpropogation network. However, this project is similar to the above in one aspect; in assuming model empathy to support context independence, thereby avoiding the bias on a particular individual involved in the text.

Sugimoto and Masahide [9] split the text into discourse units and then into sentences identifying the emotion of discourse and sentences separately. The language of the text considered is Chinese which is monosyllabic partially with nouns, verbs and adjectives being largely disyllabic. Tamil is partial free word order and does not have such phonological restrictions. Hence, application of such an approach to Tamil may not be fitting intuitively.

Soo Seoul, Joo Kim and Woo Kim propose to use keyword based model for emotion recognition when keywords are present and to use Knowledge based ANN when text lacked emotional keywords [6]. The KBANN uses horn clauses and example data. However, the accuracy of emotion recognition using KBANN is less ranging from 45%-63% compared to the recognition range of 90% where emotional keyword affect analysis is incorporated.

Most of the research in emotion recognition is directed toward audio and video from which one can infer so many cues even without understanding the narration. For text, the features of the language of text has a major emphasis on emotion recognition. Tamil is a morphologically rich language and hence its affect sensitive features would drastically vary with the usually chosen languages for emotion recognition like Chinese which is character oriented and English which is least partially inflected.
III. An architecture for Emotion Recognition

This section throws light on each module and its functions. Tamil Morphological analyzer, a product of the Tamil Computing lab of Anna University is used as the tool to retrieve part of speech like nouns, verbs, modifiers and cases. The 2D animated face generator is another tool to display the found emotion via a two dimensional face.

Following is the architecture diagram of the Neural net framework.

![architecture_diagram](image)

Fig 1: The Overall Architecture

**A) The Domain Classifier**

The Domain Classifier takes in documents which are already classified under five domains namely, politics, cinema, sports, business and health. Training process involves retrieving nouns and verbs using the Tamil Morphological Analyzer, calculating file constituent terms’ domain frequencies and inserting them into the respective domain hash tables with terms as keys and term frequencies as values. Training reports a document’s domain, based on the highest test domain frequency counter.

The algorithm is as follows

i) **Preprocessing:**
Let $F_D$ represent the set of files in a domain and $f$ be a file such that $f \in F_D$. Let $T$ be the bag of words in a domain without stop words and $t \in T$.

1. ∀$f \in F_D$, remove stop words and tokenize.
2. Get the nouns and verbs using the Tamil Morphological analyzer.
3. ∀$t \in T$, $F_t = \sum f$, where $f_i$ is the frequency of $t$ in $f$.

ii) Training:
Let $H_D$ represent a domain hash table

∀$t \in T$, insert $(t, H_D)$ where $t$ being the key and the value $v$ being

$$v = I_{occ} + \sum f_{dup(o)}$$

where $I_{occ}$ is the first occurrence of the term in the domain and $f_{dup(o)}$ is the frequency of its duplicate.

iii) Testing:
Initialise domain frequency counters $P_c$, $C_c$, $B_c$, $S_c$ and $H_c$. Given a document, remove the stopwords and tokenise.

Let Tok be the bag of nouns and verbs in the document, retrieved using the Tamil Morphological Analyser.

Convert Tok to a set by removing the duplicates.

∀$t \in$ Tok, get the frequencies of $t$, from all the domain hash tables. Increment the domain frequency counter of the domain that yields the highest frequency for $t$.

Report the domain of the test document as the domain that has the highest counter value.

B) The Negation Scorer
The Negation classifier gets the documents, analyses whether positive words occur in the neighborhood of negative words or whether likes come close to dislikes and vice versa and assigns a score.

i) Training:
Let $F$ denote set of all files in the corpus and let $f \in F$. Let $T$ be the bag of words in a file $f$ without stop words and $t \in T$.

a) ∀$f \in F$, remove stop words and tokenize.

b) ∀$f \in F$, let Neg denote the negation score of $f$ primarily initialised to 1.

c) ∀$t \in T$, let $i$ be the current position. Check whether $t$ is a positive/negative word or like/dislike word.

if $t$ is positive/like, and a negative/dislike word occurs in a window of three places to the left or right, calculate Neg as

$\text{Neg} = \text{Neg} - 0.01$

if $t$ is negative/dislike, and a negative/like word occurs in a window of three places to the left or right, calculate Neg as

$\text{Neg} = \text{Neg} - 0.1$

d) If $\text{Neg} < -0.5$, report the document as negative.
C) The Flow Scorer

The Flow scorer assigns a score to each document based on the pleasantness of the words it has. The score is set in view of the phonetic classification in Tamil alongside with place and manner of articulation. Let $F$ denote set of all files in the corpus and let $f \in F$. Let $T$ be the bag of words in a file $f$ without stop words and $t \in T$. Let $t_g$ be the English equivalent of a Tamil word $t$.

a) $\forall f \in F$, remove stop words and tokenize.

b) $\forall t \in f$, convert $t$ to $t_g$

c) $\forall t_g \in f$, compute the flow score as the sum of the Maaththirai counts diminishing when Kurukkams appear.

Table 1: Kurukkams and Maaththirai

<table>
<thead>
<tr>
<th>Rule</th>
<th>Context</th>
<th>Final Maaththirai</th>
</tr>
</thead>
<tbody>
<tr>
<td>KuttriyaLukaram</td>
<td>One among these $\mathbb{g} \not\in \mathbb{g}_1 \cup \mathbb{g}_2$ \ at \ the \ end \ of \ the \ word</td>
<td>1,(Decrease is 0.5)</td>
</tr>
<tr>
<td>Aukaarakkurukkam</td>
<td>$\mathbb{g}_2$ in the beginning</td>
<td>1.5,(Decrease is 0.5)</td>
</tr>
<tr>
<td>Aikaarakkurukkam</td>
<td>$\mathbb{g}_2$ in the beginning and middle</td>
<td>1.5,(Decrease is 0.5)</td>
</tr>
<tr>
<td></td>
<td>$\mathbb{g}_2$ in the end</td>
<td>1,(Decrease is 1)</td>
</tr>
<tr>
<td>Maharakkurukkam</td>
<td>$\mathbb{g}_1$ before $\mathbb{g}$</td>
<td>0.25,(Decrease is 1/4)</td>
</tr>
</tbody>
</table>

Table 2. Phonemes under Manner of articulation

<table>
<thead>
<tr>
<th>Category</th>
<th>Manner of Articulation</th>
<th>Phoneme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater Rough</td>
<td>Retroflex, Trill</td>
<td>$\leftarrow \mathbb{g}$</td>
</tr>
<tr>
<td>Rough</td>
<td>Tap, Dental, Bilabial</td>
<td>$\mathbb{g} \not\in \mathbb{g}_1 \cup \mathbb{g}_2$</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Semivowels, Approximants</td>
<td>$\mathbb{u} \mathbb{a} \mathbb{v} \mathbb{y} \mathbb{v}$</td>
</tr>
<tr>
<td>Soft</td>
<td>Nasal</td>
<td>$\mathbb{u} \mathbb{a} \mathbb{v} \mathbb{y} \mathbb{v}$</td>
</tr>
</tbody>
</table>

Let $t$ be the Tamil word, $t_g$ be the Grapheme form and $P_t$ be the bag of phonemes in $t$ with $p \in P_t$. Let $GRscore$ be the score of the greater rough category, $Rscore$ be the score of the rough category, $Iscore$ be the intermediate score and $Sscore$ be the score of the soft category.

i) Calculate $GRscore$, $Rscore$, $Iscore$, $Sscore$ as

$$GRscore = \sum f(p)_{GR},$$
$$Rscore = \sum f(p)_R,$$
$$Iscore = \sum f(p)_I,$$
$$Sscore = \sum f(p)_S,$$

where $f(p)_{GR}$ is the frequency of a greater rough category phoneme, $f(p)_R$ is the frequency of a rough category phoneme, $f(p)_I$ is the frequency of an intermediate category phoneme and $f(p)_S$ is the frequency of a soft category phoneme.
Calculate the FinalRoughScore and FinalSoftScore as

\[ \text{FinalRoughScore} = \text{GRscore} + \text{Rscore} . \]

\[ \text{FinalSoftScore} = \text{Iscore} + \text{Sscore} . \]

iii) If FinalRoughScore > FinalSoftScore \( T \rightarrow \text{Pleasant} \)

iv) Else if FinalSoftScore > FinalRoughScore \( T \rightarrow \text{Unpleasant} \)

v) Else \( T \rightarrow \text{Neutral} \)

D) The Taggers

At the start four taggers were constructed specifically, the noun tagger, verb tagger, Urichol tagger and case tagger. But for the instrumental and accusative cases, the left behind cases are not affective. For this reason, the constructed case tagger is unused as an input to the neural network. Nonetheless, the affect sensitive cases are incorporated in the affect specificity improver, namely the animate/inanimate affect handler.

Each tagger, picks up the respective part of speech in the document, analyses the major affect and tags the document with that affect. As a consequence, noun tagger reports the affect of nouns, verb tagger reports the affect of verbs and Urichol tagger reports the affect of Urichols. The generalized algorithm is as follows.

Let \( F \) denote set of all files in the corpus and let \( f \in F \). Let \( T \) be the bag of words in a file \( f \) without stop words and \( t \in T \). Let \( N \) denote nouns, \( V \) denote the verbs and \( U \) denote the Urichols in a file. Let \( n \in N \), \( v \in V \), \( u \in U \).

\[ a) \ \forall f \in F, \text{remove stop words and tokenize.} \]
\[ b) \ \forall f \in F, \text{get } N \text{ for Noun Tagger, } V \text{ for Verb tagger and } U \text{ for Urichol tagger, using the Tamil Morphological analyzer.} \]
\[ c) \ \forall n \in N, \text{use the noun affect lexicon to determine the noun affects. Initialize respective noun affect counters for the basic six emotions and increment them depending on the affect of each } n. \]
\[ d) \ \forall v \in V, \text{use the Verb affect lexicon to finalize the verb affects. Initialize respective affect verb counters for the basic six emotions and increment them depending on the affect of each } v. \]
\[ e) \ \forall u \in U, \text{find the Urichol affect using the Urichol affect lexicon. Initialise respective affect counters for the basic six emotions and increment them depending on the affect of each } u. \]
\[ f) \ \text{For each tagger, report the final affect of a file } f, \text{as the affect of the respective affect counter that has the maximum value.} \]

E) The Tense Affect Handler

\[ a) \ \forall f \in F, \text{get the verbs under each of the three tenses. Let } \text{Pr represent the present tense, } \text{Pa denote the past tense and } F \text{ denote the future tense.} \]
\[ b) \ \text{Analyze the affect of each tense category verbs.} \]
\[ c) \ \text{Prioritize Present, Future and then Past tenses.} \]
\[ d) \ \text{Report the high frequent affect of the Present tense. If present tense verbs are absent, report the maximum frequent affect of future tense, else report the high frequent affect category of the past tense.} \]

F) The Inanimate/Animate Handler

\[ a) \ \forall f \in F, \text{get the verbs using the Morphological analyzer.} \]
b) Restrict a window size of one to the left to find whether affect sensitive cases are present.

c) Analyze the affect of the verbs and categorize them under mild and dangerous.

d) Analyze the cases and see if they add to the affect of the verbs or nullify it.

e) If affect intensity is increased report danger, else report the affect implied.

G) The Contrary Affect Handler

   a) \( \forall f \in F \), check whether there are multiple entities in a sentence.
   b) Using the Profile monitor, check whether the entities are in like list or dislike list.
   c) Analyze the nouns next to the entities to see whether they are favorable to the preferred entities or not.
   d) Report the affect as the affect of the nouns with reference to the preferred entity.

H) The Sub-emotion Spotter

   a) \( \forall f \in F \), Get all nouns, verbs and urichols using the Tamil morphological analyzer.
   b) Use the constructed sub-emotive affect lexicon to identify the high frequent sub emotion.
   c) Report the maximum occurring affect based on the sub-affect.

I) The Neural Network

Initially a supervised Backpropogation network was constructed with the architecture 6:3:1. However, since emotion recognition is to do with emotional intelligence, unsupervised learning was preferred and Hebbian learning was incorporated. The forget factor is fixed as 0.02 and the learning factor as 0.1. Following is the pseudo code.

   a) \( \forall f \in F \), get the outputs from Domain Classifier, Negation Scorer, Flow Scorer and the three affect taggers. (The Improver modules fail to aid emotion recognition by getting eclipsed toward certain domains and are hence overlooked.)
   b) Initialize the network weights and threshold values which are set in the range of 0 to 1.
   c) Start the learning for each file using Hebb’s rule.
   d) Report the affect as the affect of the output activation that approximately equals an affect value.
   e) Stop the learning when steady state is embarked or sufficient number of epochs has been elapsed.

IV. Results and Evaluation

Both Batch and Sequential learning are supported. The precision of the Neural network is 60%. Other datasets used were lyrics and stories. Stories usually have sub plots and hence overall affect usually gets eclipsed towards neutral. Hence, news (400 files) and lyrics (230) were used for training. Validation set included 50 news files with equal contribution from five domains and 10 lyrics.

For lyrics and stories, the neural net has a better precision of 70%. The precision gets increased when epochs increase for all the datasets. Simple supervised Back propogation network was implemented and used as the Baseline whose precision was 30% more than the constructed even with minimum epochs.
Following is a graph that depicts how precision of emotion found varies with respect to the number of epochs, in the case of a single news text. Values nearer to zero indicate the drastic variation of the reported affect from the actual one (ie joy instead of sad) and values near 100 indicate the closeness towards the actual affect.

![Precision of a document vs Epochs](image)

Fig 2: Epochs vs Precision of affect of a document.

The neural network suffers from ambiguity problems in the case of closely related categories like love-joy and sad-fear irrespective of dataset. The supervised Back propogation network is used as the baseline which is 30% more in precision than the Unsupervised Hebbian Learning.

Besides, the Unsupervised Hebbian learning Neural net has exponential complexity and consumes two thirds of the physical memory. The affect reporting of a single file amounts to three minutes where approximately one and a half minute is taken for indexing and loading of the domain hash tables by Domain Classifier. The CPU usage during the learning varies roughly from 11% to 72%.

V. Conclusion

Thus, an unsupervised Hebbian learning neural network is constructed which fetches its major inputs from a domain classifier, two sentimental scorers and three part of speech taggers to figure out the affect in the presented text. As is the case with almost all natural language processing applications, ambiguities do exist in emotion recognition; here among closely related affective categories like love-joy and sad-fear. However, a competitive strategy in unsupervised learning can be opted to resolve this issue, as such a learning is concerned with demarcating one from the rest. Identification of other affect sensitive features could further aid in precise emotion detection.

VI. References


Abstract

We present a fractal coding method to recognize online handwritten Tamil characters and propose a novel technique to increase the efficiency in terms of time while coding and decoding. This technique exploits the redundancy in data, thereby achieving better compression and usage of lesser memory. It also reduces the encoding time and causes little distortion during reconstruction. Experiments have been conducted to use these fractal codes to classify the online handwritten Tamil characters from the IWFHR 2006 competition dataset. In one approach, we use coding and decoding process. A recognition accuracy of 90% has been achieved by using DTW for distortion evaluation during classification and encoding processes as compared to 78% using nearest neighbor classifier. In other experiments, we use the fractal code, fractal dimensions and features derived from fractal codes as features in separate classifiers. Whereas the fractal code was successful as a feature, the other two features are not able to capture the wide within-class variations.

Introduction

Fractal codes are the compressed representation of patterns, based on iterative contractive transformations in metric spaces proposed by Barnsley. A simplified version of the fractal block coding technique for digital images has been used to encode the 1-D ordered online handwritten character patterns. A novel partitioning algorithm has been proposed to reduce the computation complexity of encoding and decoding, with a minor fall in recognition accuracy.

Building fractal codes for handwritten characters

We need to find the collection of affine transforms of the online handwritten character. The raw online handwritten character is first preprocessed using three steps: (i) smoothing (ii) re-sampling the variable number of points in each character to 60 points. (iii) normalizing the x and y coordinates between 0 and 1. The handwritten character locus is divided into non-overlapping range segments. Each range segment has a fixed number of points (R) in it. Last point of each range becomes the first point of the next range, except in the case of the last range.

Creating a pool of domain segments

The domain pool is formed for each character locus. Domain pool is the collection of all possible domain segments. The number of points in each domain segment is chosen to be double that in each range segment, D = 2R. Domain pool can be obtained by sliding the window containing D points at a time. The window is first located at the beginning of the stroke. The window is moved along the
stroke by δ points, in such a way that it does not cross the end point of the stroke. The step δ has been chosen as R/2 in our experiments.

**Constructing transformed Domain pool**

Transformed domain pool is constructed by multiplying each domain segment with the eight isometrics that involve reflection and rotation about different axes. To begin with, each domain is translated to its centroid and scaled down by the contractivity factor (s=0.5). The following transformations are then applied to each of the candidate domain segment.

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix},
\begin{bmatrix}
1 & 0 \\
0 & -1
\end{bmatrix},
\begin{bmatrix}
-1 & 0 \\
0 & 1
\end{bmatrix},
\begin{bmatrix}
0 & 1 \\
-1 & 0
\end{bmatrix},
\begin{bmatrix}
0 & -1 \\
1 & 0
\end{bmatrix},
\begin{bmatrix}
-1 & 0 \\
0 & -1
\end{bmatrix},
\begin{bmatrix}
1 & 0 \\
0 & -1
\end{bmatrix},
\begin{bmatrix}
-1 & 0 \\
0 & 1
\end{bmatrix}
\]

The above transformations produce a whole family of geometrically related domain segments. In domain pool, matching blocks will be looked for encoding the online handwritten character.

**Searching for the most similar domain segment for each range:** Each affine transformed domain segment is re-sampled into R points and then its centroid is translated to that of the concerned range segment. Distance between them is found. Similarly, distances w.r.t to the all the domain segments is calculated. The most similar domain segment corresponding to each range segment is identified and fractal code is stored corresponding to the particular range. The fractal codes are similarly obtained for all the online handwritten characters.

Fractal codes corresponding to each range segment consists of (1) the range segment index, (2) the range segment centroid, (3) index of the most similar domain segment and (4) the index of transformation used out of the 8 transformations.

**Issue related to constructing fractal codes**

The whole character is divided into range segments of equal number of points. Smaller the number of points in each range segment, the more minutely we can capture the complexity in any region of the character. The number of range segments per character is thus inversely proportional to the number of points in each range. Again, the encoding speed is inversely proportional to the number of range segments per character. It has been noted that there are certain region in a character where the curliness is minimal so in those area the range segment size could be increased still encoding the region precisely.

**Steps to encode a handwritten character where the number of points in each range is variable:**

![Fig 1. Tamil handwritten character ‘aa’](image)
Cumulative angle ‘θc’ is calculated starting from the first point and traversing the character stroke till it crosses an empirically set threshold of θT. Smaller the threshold, finer is the encoding. Figure 1 shows a sample of the handwritten character /aa/ in Tamil. Figure 2 shows the effect of the choice of the angle threshold on the reconstruction error.

Fig 2. Illustration of the distortion in reconstruction with different angle thresholds ϕT. Here reconstruction is performed from the fractal codes of the character /aa/ shown in Fig 1. For encoding, different values of range sizes are used, namely, 4, 8, 12, 16 and 20.

Algorithm for partitioning

1. Domain pool of different sizes (namely 8, 16, 24, 32, 40) are constructed corresponding to the range sizes of 4, 8, 12, 16 and 20. By size, we mean the number of points in each domain.
2. Start from the first point and move along the character from one point to the next and calculate the cumulative change in angle θc.
3. The No. of points (K) till the point penultimate to the one, where θc crosses the threshold ϕT is noted.
4. The range size closest to and less than K is chosen. The most suitable domain is chosen from the corresponding domain pool and the fractal codes are stored.
5. The last point of the present range is then considered as the first point of the new range and the process repeats starting from step 2.

Along with the fractal codes, the size of the range chosen is also stored. If the end point is reached with θc < ϕT, then step 4 is followed, where K includes the last point also since θC < ϕT. If at the end, few points are left which is less than the smallest range, they are discarded else step 4 is repeated.

Algorithm for reconstruction: Banach’s contractive mapping theorem states: “If a contractive mapping ‘W’ (which are the fractal codes here) is defined, then iterative application of the mapping on any sequence of the same space will lead to a Cauchy’s sequence which will converge to a fixed, unique point.

CASE I: Range having fixed number of points

A random initial pattern having the same number of points is taken or generated. A domain pool of size double that of the range is created in a manner similar to the encoding process. First fractal code is
taken corresponding to the first range of the pattern, and operations are performed on the corresponding domain indicated by the domain index in the code of first range. The indicated domain segment’s origin is shifted to its origin and then it is scaled down by the contractivity factor (0.5 here). Then the affine transformation as indicated in the code is applied on the scaled domain segment. Finally the transformed domain’s centroid is shifted to the range segment centroid as present in the fractal code. The above steps are repeated to decode all the range segments. Then the whole decoded locus is smoothened. The above steps are repeated till the termination condition is satisfied to finally converge to a fixed and unique pattern. Termination condition could be (i) an empirically set fixed number of iterations, sufficient for convergence or (ii) minimal or no distortion between two consecutive patterns produced by 2 consecutive iterations.

**CASE II: Range having variable number of points in each range**

In this case, multiple domain pools are created out of the random pattern taken for the reconstruction. Using the extra information given in the fractal code pertaining to the range size to be chosen so that domain segment is picked up from the appropriate domain pool. The rest of the steps are same as the case for the reconstruction with fixed range size. Using above two methods, fractal codes of any given pattern can be created and the same pattern could be decoded using any random pattern after applying this reconstruction algorithm iteratively for few times.

**1. Classification by using fractal codes in construction and reconstruction:**

The above fractal encoding and decoding method has been used for classification of characters. The assumption behind this classification is that if a sample of a class (say /a/) is encoded and fractal codes are obtained. The following process of reconstruction if started in 2 ways i.e. firstly by applying the reconstruction algorithm iteratively on a random pattern of any different class (anything other than ‘A’) and secondly doing the same on any random pattern of the same class (i.e. ‘A’) then the distortion between the initial pattern and the pattern obtained after first iteration of reconstruction is relatively much more in the first case than in the second. The reason behind this is that reconstruction process leads to convergence to the pattern whose code is used for reconstruction. And since the class of the initial pattern in the second case and the fractal code is same, the distortion in the second case is smaller than the first case.
Fig 3. In the above image, reconstruction process is shown which starts from a random straight line and finally converges to a pattern which is very close to the original pattern after 8 iterations. The original pattern (in Fig 1) was encoded using different range sizes of 4, 8, 12, 16 and 20 with the threshold angle of 30 degrees.

**Character classification using fractal codes:**

![Image of reconstructed patterns]

Fig. 4. The above image shows the distortion created, when an iteration of reconstruction was performed. This shows that the distortion is huge if the starting pattern (above left) is very different from the original pattern, whose fractal codes are used for reconstruction (in this case Fig 1).

![Image of reconstructed patterns]

Fig. 5. The above image shows the distortion created, when an iteration of reconstruction was performed. This shows that the distortion is much less if the starting pattern (above left) is not very different from the original pattern, whose fractal codes are used for reconstruction (in this case Fig 1).

**Classification Algorithm**

In the present research, fractal codes of ‘N’ samples of entire 156 classes are computed and stored. To classify a test sample, following steps are taken.

a) An iteration of the reconstruction algorithm is applied on the test sample using the fractal code of each sample of each class.

b) The distortion ‘D’ is calculated between the initial pattern and the pattern obtained after one step of reconstruction. Thus the distortion matrix of size 156*N is obtained.

c) The row number of the minimum value of the distortion is found from the distortion matrix and assigned to the test sample.

**Distortion evaluation:**

The distortion between two patterns can be evaluated by finding the distance between them. The distance between 2 patterns is measured by Nearest Neighbor (NN) method. The issue with NN is that the matching is done point by point which increases the distance unusually and thus decreases the classification accuracy (evident from the result table). This drawback in distance evaluation of NN
is addressed by DTW pattern matching which is more intuitive. This intuitive matching takes place because DTW matches similar subsections of the patterns thus producing a reasonable distance between patterns. This method hence produces a remarkable increase in accuracy as shown Table 1.

**Classification using the fractal codes as a features in the Nearest Neighbor (NN)**

In this method, the fractal codes are used as features and fed into a NN classifier. An accuracy of approximately 65% is obtained.

**Classification using fractal dimension or features derived from the fractal codes**

Fractal dimension is a unique identity of any pattern or object. However, because of the mere nature of handwritten character recognition (i.e. large variation within every class), it fails completely to classify any random sample. The features derived from the fractal codes like MMVA and DRCLM, though successful in problems like face recognition and signature verification, fail in recognizing handwritten characters.

**Results**

Table 1: Results show how the DTW comparison during reconstruction impacts the recognition accuracy

<table>
<thead>
<tr>
<th>Fixed Range Size</th>
<th>DTW used?</th>
<th>No. of Training samples</th>
<th>No. of Testing samples</th>
<th>Accuracy (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>No</td>
<td>20</td>
<td>50</td>
<td>78.9</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>20</td>
<td>50</td>
<td>90.4</td>
</tr>
</tbody>
</table>

Table 2: Results show the efficacy of the Partitioning algorithm in improving the efficiency of the recognition system (in terms of time) with marginal drop in accuracy. Variable range sizes used (4, 8, 16, 24 and 32). No. of training and testing samples used are 5 and 30, respectively, No. of classes used is 156.

<table>
<thead>
<tr>
<th>Threshold angle (degree)</th>
<th>DTW used for distortion evaluation?</th>
<th>Encoding time per sample (sec)</th>
<th>Accuracy (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Yes</td>
<td>31</td>
<td>90.44</td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>22</td>
<td>86.43</td>
</tr>
<tr>
<td>30</td>
<td>Yes</td>
<td>12</td>
<td>85.04</td>
</tr>
<tr>
<td>50</td>
<td>Yes</td>
<td>10</td>
<td>83.55</td>
</tr>
<tr>
<td>70</td>
<td>Yes</td>
<td>8</td>
<td>81.60</td>
</tr>
<tr>
<td>90</td>
<td>Yes</td>
<td>6</td>
<td>80.87</td>
</tr>
</tbody>
</table>

**Acknowledgment:** The authors thank Technology Development for Indian Languages (TDIL), DIT, Government of India for funding this research, as part of the research consortium on Online handwriting recognition of Indian languages.

**References**

Neuroscience inspired segmentation of handwritten words

A G Ramakrishnan and Suresh Sundaram
MILE Lab, Dept of Electrical Engineering, Indian Institute of Science, Bangalore.

The challenge of segmenting online handwritten Tamil words has hardly been investigated. In this paper, we report a neuroscience-inspired, lexicon-free approach to segment Tamil words into its constituent symbols (recognizable entities). Based on a simple dominant overlap criterion, the word is grossly segmented into candidate symbols (stroke groups). However, this segmentation is not fully reliable because of varying writing styles resulting in varying levels of overlap. Taking cues from vertebrate visual perception, we utilize both feature based attention and feedback from the classifier to detect possible wrong segmentations. This attention-feedback segmentation (AFS) strategy splits or merges the stroke groups to correct the segmentation errors and forms valid symbols. This maiden attempt on segmentation is tested on 10000 handwritten words collected from hundreds of writers. The efficacy of AFS in segmentation and improving the recognition performance of the handwriting system is amply demonstrated. Our results show a segmentation accuracy of over 99% at symbol level.

Need for segmenting handwritten words

Since attempts to segment cursively handwritten English words have largely failed, researchers working on Indic scripts too feel that it is not advisable to try to segment individual characters from handwritten documents. However, we firmly believe that it is not only possible, but also something that ought to be done, if one is interested in recognizing words such as proper names appearing in the name and address fields of handwritten forms. Thus, this opens up the possibility of developing a recognizer that can handle unrestricted vocabulary, including any unusual word of foreign origin, such as names of people or places from other countries. Thus, we believe that our work is the first of its kind in proposing an approach for handwriting recognition that does not limit the writer from writing any text of any origin.

Motivation for Attention-Feedback Segmentation

Traditional pattern recognition [1, 3-5, 8, 10, 13-15] primarily follows a feedforward architecture, whereas the same in mammalian brain involves complex feedback structures. Studies on visual perception in primates demonstrate the effect of attention on the response of the visual neurons [2]. Feature based attention biases the neuronal responses as though the attended stimulus was presented alone. Also, shifting spatial attention from outside to the inside of the receptive field increases the neuronal responses. Motivated by these observations, we incorporate local feature based attention to correct and improve segmentation [9]. Further, studies on visual pathways show extensive feedback from the cortex to the lateral geniculate nucleus (LGN), which have both inhibitory and facilitatory effects on the responses of LGN relay cells. In our work, we use feedback based on features as well as from the classifier posterior probabilities to rectify any incorrect segmentation by regrouping the strokes. Thus, we call our approach as ‘attention-feedback’ strategy for segmentation.
Further, studies on scene perception by humans [6] indicate that visual processing follows a top-down approach. The global cues characterizing the visual object, that appear within the visual span, are perceived before the local features. The human perceptual system treats every scene as if it were in the process of being focussed or zoomed in on, whereas initially, it is relatively less distinct. Moreover, the human perceptual processor has the capability to select parts of the input stimulus that are worth to be paid attention to. Motivated with these observations from the field of neuroscience, we present a segmentation strategy that first works on the global feature of overlap to output candidate Tamil stroke groups for the given input strokes. By analyzing local features characteristic to the given input pattern, we reevaluate the segmentation and modify the segmentation when found necessary. The localized features are derived by zooming on paying attention to specific parts of the online trace. Essentially, we adopt a multi-pass system, wherein fine grained processing is guided by the prior cursory (global) processing.

Data used for the study

The 155 distinct Tamil symbols (comprising 11 vowels, 23 base consonants, 23 pure consonants, 92 CV combinations and 6 additional symbols) are presented in Appendix A. The publicly available corpus of isolated Tamil symbols (IWFHR database) is used for learning various statistics about Tamil symbols. The primary focus of this work is to address the challenges of segmentation. Towards this purpose, Tamil words are collected using a custom application running on a tablet PC and saved using a XML standard [7]. High school students from across 6 educational institutions in Tamil Nadu contributed in building the word database of 100,000 words, referred to as the ‘MILE Word Database’ in this work [12]. Out of these, 10,000 words are used for this study. The words have been divided into 40 sets, each comprising 250 words. Owing to the comparable resolution of our input device to that used in the IWFHR dataset, statistical analysis performed on the symbols in the IWFHR database are applicable to the Tamil symbols in the MILE word database.

Dominant Overlap Criterion Segmentation

An online word can be represented as a sequence of n strokes \( W = \{s_1, s_2, \ldots, s_n\} \). In the case of multi-stroke Tamil symbols, strokes of the same symbol may significantly overlap in the horizontal direction. The word is first grossly segmented based on a bounding box overlap criterion, generating a set of stroke groups. In this ‘Dominant Overlap Criterion Segmentation’ (DOCS), the heavily overlapped strokes are merged. A stroke group is defined as a set of consecutive strokes merged by the DOCS step, which is possibly a valid Tamil symbol.

For the \( k \)-th stroke group \( S_k \) under consideration, its successive stroke is taken and checked for possible overlap. Significant overlap necessitates the successive stroke to be merged with the stroke group \( S_k \). Otherwise, the successive stroke is considered to begin a new stroke group \( S_{k+1} \). The algorithm proceeds till all the strokes of the word are exhausted.

Neuroscience-inspired segmentation

The stroke groups obtained from the above dominant overlap criterion segmentation are preprocessed by smoothing, normalization and resampling into standard number of equi-arc length spaced points.
The x and y coordinates of these processed stroke groups and their first and second derivatives are used as features for recognition using a support vector machine (SVM) classifier that outputs class labels and their posterior probabilities. Obviously, DOCS being simple, does not always result in correct segmentation. Sometimes it results in over segmentation of a single multi-stroke character into two stroke groups; other times, two distinct characters get combined into a single stroke group, due to the way they are written.

Attention Features

Figure 1 shows the complete block schematic of the proposed segmentation scheme. An over-segmented symbol is usually small and hence results in low aspect ratio as well as has very few dominant points (points where the curvature is high). By paying attention to these features extracted from the stroke groups output by DOCS block, one can suspect wrong segmentation. Further, the symbols that result from over- or under-segmentation are classes that the classifier has not come across. Thus, these symbols usually result in a low confidence level of the classifier. Thus, the posterior probability of the classifier, when fed back to the input stages, can be used to invoke the computation of the attention features. The feedback, together with the attention features suggest possible resegmentation of the input strokes, resulting in new possible stroke groups. These modified stroke groups based on merger or splitting of original stroke groups, are once again recognized by the classifier after preprocessing and extraction of recognition features. An improved posterior probability of the new stroke group confirms right segmentation. Thus, the refinement in segmentation is caused based on memory, attention and feedback mechanisms prevalent in human perception. We call this as “attention-feedback segmentation (AFS)”.

Commonly found segmentation issues

The two Tamil characters that ought to have a minimum of three strokes are the long /i/ (nedil) and the aydam. Since in both of these cases, in general there is no overlap between the final dot and the rest of the character, they always are over segmented into two or more stroke groups.

Pure consonants (ney ezhuthu), when they are written with the dot (pulli) beyond the base consonant, result in over segmentation too.

Characters such as /ka/, /nga/ and /ra/, which start with an initial vertical segment, are written by many with multiple strokes, with the first stroke being a simple down-going vertical line. These characters have a potential to be over segmented, if the following part of the character does not clearly overlap with the vertical line.

All CV combinations of /i/ and /I/ and the CV combinations of /u/ and /U/ with borrowed consonants such as /ja/ and /sha/ also have a tendency to be over-segmented, if the vowel matra is written with no horizontal overlap with the consonant.

Under segmentation occurs if the ending part (usually bottom extensions of /ta/ or /Ra/) of the following character goes far left below the previous character, causing significant horizontal overlap between them. At other times, people write two successive characters so closely, that there is significant overlap between them.
Naturally, in all the above cases, the simple segmentation (DOCS) is likely to result in wrong segmentation leading to erroneous recognition results.

Segmentation results on the MILE Tamil Word Database

The proposed techniques are tested on the subset of 10,000 words. However, to start with, we evaluate the performance on a set of 250 words (denoted as DB1), that has a significant number of errors resulting from the DOCS. Of the 103 errors, 89 (or 86%) correspond to the merging of valid symbols, and the rest, to broken symbols. The AFS module aids in properly detecting and correcting 91 (or 90%) of these errors. In addition, the methods proposed effectively merge 11 (or 78%) of the over-segmented stroke groups to valid symbols. The improvement in character segmentation rate in turn reduces the number of wrongly segmented words. It is observed that only 7 of the total 250 words remain wrongly segmented after the AFS scheme, against 67 words after the DOCS scheme. On evaluating the performance across the database of 10000 words, we obtain a 86% reduction in character segmentation errors.

Recognition results on the MILE Database

We report experimental results demonstrating the impact of the proposed AFS strategy on the recognition of symbols in the MILE word database. Since a significant percentage of DOCS errors are corrected by AFS, a drastic improvement of 16% (from 70.5 % to 87.1 %) in symbol recognition is observed. In computing the symbol recognition rate, apart from the substitution errors, we take into account the insertion and deletion errors, caused by over-segmentation and under-segmentation, respectively. The edit distance is used for matching the recognized symbols with the ground truth data. Moreover, 11.6 % of the words, (29 additional words) wrongly recognized after DOCS, have
been corrected by the proposed technique. Across the 10000 words in the MILE Word database, an improvement of 4% (from 83 to 87%) in symbol recognition rate has been obtained.

Conclusion

In this paper, we present a maiden attempt based on significant feedback from the classifier to the input blocks such as feature extraction and segmentation, as well as the use of memory (prior knowledge) to result in a very effective segmentation of online handwritten Tamil words. This approach being general, can be extended to any other Dravidian script, as well as any other script where cursive writing is not practiced. To our knowledge, there is no reported systematic research work on segmenting the individual characters or recognizable standard symbols from online handwritten words for any Indic language that does not have the shiro rekha (head line). Thus, we are unable to compare the performance of our work with any other technique. However, the results are promising and have also led to improved recognition of the handwritten words [16], thus confirming the possibility of proper segmentation of online Tamil words. We intend to extend this work very soon to online Kannada handwritten words.

Acknowledgment

We thank Ms. Nethra Nayak, Mr. Rituraj Kunwar, Ms. Archana C P, Mr. Shashi Kiran, Ms. Chandrakala, Mrs. Shanthi Devaraj, Ms. Saranya and Ms. Southeriya for their efforts in data collection and annotation, which made these experiments possible. We thank Dr. Arun Sripathi of the Centre for Neuroscience, IISc for the very useful discussions we had on visual perception. Special thanks to Technology Development for Indian Languages (TDIL), Department of Information Technology (DIT), Government of India for funding this research, as part of a research consortium on online handwriting recognition in several Indic scripts. We thank Prof. Deivasundram (University of Madras), AVM Matriculation Higher Secondary School Virugambakkam, Chennai, Govt. Boys Higher Secondary School, Sulur, Presidency College, Triplicane, Chennai and IIT Madras for contributing to the data set. We also thank Dr. Anoop Namboodiri for giving us the word level annotation tool. We thank CDAC, Pune for being a partner in finalizing the XML standard for handwritten data collection in Indic languages.

References

- Suresh Sundaram and A G Ramakrishnan, “Attention-feedback based robust segmentation of online Tamil words”, under review, Pattern Recognition, 2011.
Improving Tamil-English Cross-Language Information Retrieval by Transliteration Generation and Mining

A Kumaran, K Saravanan & Ragavendra Udupa
Multilingual Systems Research
Microsoft Research India
Bangalore, India.
{kumaranav, v2sarak, raghavu}@microsoft.com

Abstract

While state of the art Cross-Language Information Retrieval (CLIR) systems are reasonably accurate and largely robust, they typically make mistakes in handling proper or common nouns. Such terms suffer from compounding of errors during the query translation phase, and during the document retrieval phase. In this paper, we propose two techniques, specifically, transliteration generation and mining, to effectively handle such query terms that may occur in their transliterated form in the target corpus. Transliteration generation approach generates the possible transliteration equivalents for the out of vocabulary (OOV) terms during the query translation phase. The mining approach mines potential transliteration equivalents for the OOV terms, from the first-pass retrieval from the target corpus, for a final retrieval. An implementation of such an integrated system achieved the peak retrieval performance of a MAP of 0.5133 in the monolingual English-English task, and 0.4145 in the Tamil-English task. The Tamil-English cross-language retrieval performance improved from 75% to 81% of the English-English monolingual retrieval performance, underscoring the effectiveness of the integrated CLIR system in enhancing the performance of the CLIR system.

1. Introduction

With the exponential growth of non-English population in the Internet over the last two decades, Cross-Language Information Retrieval (CLIR) has gained importance as a research discipline and as an end-user technology. While the core CLIR system is fairly robust and accurate with sufficient training data, it’s handling of proper or common nouns (or, more generally, those query terms that could occur in their transliterated form in the target corpus in cross language environments) is far from desirable in most implementations. In essence, the name translations are not typically part of translation lexicons used for query translations, and hence do not get translated properly in the target language. Note that, from the CLIR point of view, any un-translated word is an out-of-vocabulary word, which typically include words that are literally transliterated into local language words (such as, computer, corporation, etc.), specifically in those countries where English is spoken as a second language. This phenomenon is called as code-mixing. Such words are not available in the translation lexicon, but are typically part of the local language corpora. Also, given that a name may be spelled differently in the target corpus – particularly for those names that are not native to the target language – the retrieval performance suffers further, as the errors in query translation and spelling variations compound. Given that proper and common nouns form a significant portion of query terms, it is
critical that such query terms are handled effectively in a CLIR system. This is precisely the research theme that we explore in this paper.

Evaluation of Tamil-English cross-language information retrieval systems started first in the Forum for Information Retrieval Evaluation (FIRE) [1], modeled after the highly successful CLEF [2] and NTCIR [3] campaigns. In 2010, FIRE organized several ad hoc monolingual and cross-language retrieval tracks, and we participated in the English monolingual and cross-language Hindi-English and Tamil-English ad hoc retrieval tracks. This paper presents the details of our participation, specifically, in the Tamil-English cross-language tasks and present the performance of our official runs.

2. Cross-language Retrieval System

In this section, we outline the various components of our CLIR system integrated with the two techniques for handling OOV words.

Our monolingual retrieval system is based on the well-known Language Modeling framework to information retrieval. In this framework, the queries as well the documents are viewed as probability distributions. The similarity of a query with a document is measured in terms of the likelihood of the query under the document language model. We refer interested readers to [6, 7] for the retrieval model and the details of this framework. In our CLIR model, the query in a source language is translated into the target language – English – using a probabilistic translation lexicon, learnt from a given parallel corpora of about 50,000 parallel sentences between English and Tamil. Such learnt translation dictionary included ~107 K Tamil words and ~45 K English words. From this dictionary, we used only top 4 translations for every source word, an empirically determined limit to avoid generation of noisy terms in the query translations.

Like any cross-language system that makes use of a translation lexicon, we too faced the problem of out of vocabulary (OOV) query terms. To handle these OOV terms, we used two different techniques, (i) generation of transliteration equivalents, and (ii) mining of transliteration equivalents:

- In Transliteration Generation, the transliterations of the OOV terms in the target language are generated using an automatic Machine Transliteration system, and used for augmenting the query in the target language.
- In Transliteration Mining, the transliteration equivalents of the OOV terms are mined from the top-retrieved documents from the first pass, which are subsequently used in the query for a final retrieval [8].

2.4 Generating Transliterations

We adopted a conditional random fields based approach using purely orthographic features, as a systematic comparison of the various transliteration systems in the NEWS-2009 workshop [9] showed conclusively that orthography based discriminative models performed the best among all competing systems and approaches. In addition, since the Indian languages share many characteristics among them, such as distinct orthographic representation for different variations – aspirated or un-aspirated, voiced or voiceless, etc. – of many consonants, we introduced a word origin detection module (trained with about 3000 hand-classified training set) to identify specifically Indian origin names. All other
names are transliterated through an engine that is trained on non-Indian origin names. Manual verification showed that this method about 97% accurate. We used CRF++, an open source implementation of CRF model, trained on about 15,000 parallel names between English and Tamil. The transliteration engine was trained on a rich feature set (aligned characters in each direction within a distance of 2 and source and target bigrams and trigrams) generated from this character-aligned data.

### 2.5 Mining Transliteration Equivalents

The mining algorithm issues the translated query minus OOV terms to the target language information retrieval system and mines transliterations of the OOV terms from the top results of the first-pass retrieval. Hence, in the first pass, each query-result pair is viewed as a “comparable” document pair, assuming that the retrieval brought in a reasonably good quality results set based on the translated query without the OOV terms. The mining algorithm hypothesizes a match between an OOV query term and a document term in the “comparable” document pair and employs a transliteration similarity model to decide whether the document term is a transliteration of the query term Transliterations mined in this manner are then used to retranslate the query and issued again, for the final retrieval. The details of transliteration similarity model may be found in [6, 7], and the details of our training are given in [8].

### 3. Experimental Setup & Results

In this section, we specify all the data used in our experiments, both that were released for the CLIR experiments for FIRE task, and that used for training our CLIR system.

#### 3.1 FIRE Data

The English document collection provided by FIRE was used in all our runs. The English document collection consists of ~124,000 news articles from “The Telegraph India” from 2004-07. All the English documents were stemmed. Totally 50 topics were provided in each of the languages, each topic having a title (T), description (D) and narrative (N), successively expanding the scope of the query. Table 1 shows a typical topic in Tamil, and the TDN components of the topic, for which relevant English documents are to be retrieved from the aforementioned English news corpus. It should be noted that FIRE has also released a set of 50 English (i.e., target language) topics, equivalent to each of the source language topics.

#### Table 1. A Typical FIRE Topic in Tamil.

<table>
<thead>
<tr>
<th>Type</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>தாங்கா முகளி பாராகன்கலா காட்சிக்குரிய முகுந்தம் இடம்பெற்று.</td>
</tr>
<tr>
<td>Description</td>
<td>கையா அரும்பு பராக்கத்தின் தாங்கா முகளிப்பாராக்கலா காட்சியான முகுந்தம் முகுந்தம் இடம்பெற்று.</td>
</tr>
</tbody>
</table>
| Narration  | கையா அரும்பு பராக்கத்தின் தாங்கா முகளிப்பாராக்கலா காட்சியான முகுந்தம் முகுந்தம் இடம்பெற்று. முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் இடம்பெற்று. முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் இடம்பெற்று. முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் இடம்பெற்று. முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் இடம்பெற்று. முகுந்தம் முகுந்தம் முகுந்தம் முகுந்தம் இடம்பெற்று.
**Table 2.** A Typical FIRE Topic in English.

<table>
<thead>
<tr>
<th>Type</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Links between Gutkha manufacturers and the underworld.</td>
</tr>
<tr>
<td>Description</td>
<td>Links between the Goa and Manikchand Gutkha manufacturing companies and Dawood Ibrahim.</td>
</tr>
<tr>
<td>Narration</td>
<td>A relevant document should contain information about the links between the owners of the Manikchand Gutkha and Goa Gutkha companies and Dawood Ibrahim, the gangster. Information about links between Dawood Ibrahim and other companies is not relevant.</td>
</tr>
</tbody>
</table>

### 3.2 Metrics & Performance Data

The standard measures for evaluating our tasks were used, specifically, Mean Average Precision (MAP) and Precision at top-10 (P@10). As shown in Table 1, each of the 50 topics in Tamil has a title (T), description (D) and narrative (N), successively expanding the scope of the query. We ran our experiments taking progressively each of (title), (title and description), and (title, description and narrative), calibrating the cross-language retrieval performance at each stage, to explore whether expanding the query adds useful information for retrieval or just noise. Table 3 shows the notation used in our description of various configurations to interpret the results presented in Tables 4 and 5.

**Table 3.** Notations used

<table>
<thead>
<tr>
<th>T/TD/TDN</th>
<th>Title/ Title and Description / Title, Description and Narration</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>Transliteration Mining</td>
</tr>
<tr>
<td>GD</td>
<td>Transliteration Generation</td>
</tr>
</tbody>
</table>

Tables 4 and 5 show the results of our monolingual as well as cross-language official runs submitted to FIRE 2010 shared task. The format of the run ids in the results table is ‘Source-Target-Data-Technique’, where ‘Data’ indicates the data used for topic, and is one of {T, TD, TDN} and ‘Technique’ indicates the technique and from the set {M, GD, GT, M+GD, M+GT}. The ‘+’ refers to the combination of more than one approach. The symbols double star (**) and single star (*) indicate statistically significant differences with 95% and 90% confidence respectively according to the paired t-test over the baseline. The best results achieved are highlighted in bold.

### 4.4 Monolingual English Retrieval

We submitted 3 official runs for the English monolingual track, as shown in the Table 4. For these runs, the English topics provided by the FIRE 2010 organizers were used.
Table 4. English Monolingual Retrieval Performance

<table>
<thead>
<tr>
<th>Run</th>
<th>MAP</th>
<th>P@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>English-English-T</td>
<td>0.3653</td>
<td>0.344</td>
</tr>
<tr>
<td>English-English-TD</td>
<td>0.4571</td>
<td>0.406</td>
</tr>
<tr>
<td>English-English-TDN</td>
<td>0.5133</td>
<td>0.462</td>
</tr>
</tbody>
</table>

With the full topic (TDN), our system achieved a peak MAP score of 0.5133. Generally this performance is thought to be the upper bound for cross-language performance, presented in Table 5.

4.6 Tamil-English Cross-Language Retrieval

We submitted totally 12 official Tamil-English cross-language runs, as shown in Table 5. As discussed for the Hindi-English runs, the first run under each of the ‘T’, ‘TD’ and ‘TDN’ sections in Table 5 present the results of the runs without handling the OOV terms, and hence provide a baseline for measuring the incremental performance due to transliteration generation or mining, provided subsequently.

Table 5. Tamil-English Cross-Language Retrieval Performance

<table>
<thead>
<tr>
<th>Run</th>
<th>MAP</th>
<th>P@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tamil-English-T</td>
<td>0.2710</td>
<td>0.258</td>
</tr>
<tr>
<td>Tamil-English-T[G_D]</td>
<td>0.2891*</td>
<td>0.268</td>
</tr>
<tr>
<td>Tamil-English-T[M]</td>
<td>0.2815**</td>
<td>0.258</td>
</tr>
<tr>
<td>Tamil-English-T[M+G_D]</td>
<td>0.2816*</td>
<td>0.268</td>
</tr>
<tr>
<td>Tamil-English-TD</td>
<td>0.3439</td>
<td>0.346</td>
</tr>
<tr>
<td>Tamil-English-TD[G_D]</td>
<td>0.3548*</td>
<td>0.35</td>
</tr>
<tr>
<td>Tamil-English-TD[M]</td>
<td>0.3621**</td>
<td>0.346</td>
</tr>
<tr>
<td>Tamil-English-TD[M+G_D]</td>
<td>0.3617**</td>
<td>0.362</td>
</tr>
<tr>
<td>Tamil-English-TDN</td>
<td>0.3912</td>
<td>0.368</td>
</tr>
<tr>
<td>Tamil-English-TDN[G_D]</td>
<td>0.4068**</td>
<td>0.378</td>
</tr>
<tr>
<td>Tamil-English-TDN[M]</td>
<td>0.4145**</td>
<td>0.368</td>
</tr>
<tr>
<td>Tamil-English-TDN[M+G_D]</td>
<td>0.4139**</td>
<td>0.394</td>
</tr>
</tbody>
</table>

From the results presented in Table 5, we observe that the usage of all of the components of the topic, namely T, D and N, produced the best retrieval performance. The basic Tamil-English cross-language run ‘Tamil-English-TDN’ (without transliteration generation or mining), achieved the MAP score
0.3912, and our best cross-language run ‘Tamil-English-TDN[M]’ with mining achieved a MAP score of 0.4145. We observe, in general, similar trends in the other runs that use only the title, or title and description sections of the topics. While the cross-language performance of Tamil-English achieves ~81% of our monolingual English retrieval performance, we observe that this is not as high as the Hindi-English retrieval, perhaps due to the highly agglutinative nature of Tamil.

Given that the mining technique performed generally above the other techniques, we focus on addition of mining to the base CLIR, for subsequent analysis in the following sections.

4.7 Mining OOV terms and its effect on CLIR performance

In this section, we analyze the volume of the OOV terms in FIRE topics, and to what extent they are handled by our mining technique, which clearly emerged as the better technique for boosting the retrieval performance. Also, we show the effect of handling the OOVs on the cross-language retrieval performance, for both the Hindi-English and Tamil-English CLIR runs. Table 6 enumerates the number of unique OOV terms in Tamil FIRE 2010 topics.

Table 6. OOV terms in Tamil-English CLIR.

<table>
<thead>
<tr>
<th>Topic Config.</th>
<th>OOV terms</th>
<th>Transliteratable OOV terms</th>
<th>Transliteratable OOV terms handled</th>
<th>% of Transliteratable OOV terms handled</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>24</td>
<td>13</td>
<td>5</td>
<td>38.46</td>
</tr>
<tr>
<td>TD</td>
<td>58</td>
<td>29</td>
<td>15</td>
<td>51.72</td>
</tr>
<tr>
<td>TDN</td>
<td>129</td>
<td>47</td>
<td>24</td>
<td>51.06</td>
</tr>
</tbody>
</table>

As shown in the third line in Table 6 in the TDN configuration of the 50 Tamil topics of the FIRE2010 shared task, there were totally 129 unique OOV terms, out of which 47 were proper or common nouns, handling of which may help improving the cross-language retrieval performance. Our mining technique did find at least one transliteration equivalent for 24 of these OOV terms (that is, 51%). As shown in Table 5, handling these OOV’s resulted in nearly 6% improvement in the MAP score over the baseline where no OOV’s were handled.

Note that Tamil OOV’s pose specific challenges as outlined below: First, the transliteratable terms mentioned in the third column of the Table 6 excludes some terms whose equivalents are multiword expression in English; mining such multiword transliteration equivalents is beyond the scope of our work and hence they were not handled. Second, 26 out of the 47 terms that are transliteratable were inflected or agglutinated. While our mining algorithm could mine some of them, many of the terms were missed at our parameter settings for mining. By relaxing the mining parameters settings we could mine more such terms, but such relaxation introduced many more noisy terms, affecting the overall retrieval performance. We believe that the use of a good stemmer for inflectional languages like Tamil may help our mining algorithm and, compositionally, the cross-language retrieval performance.
4.8 Mining OOV terms and its effect on individual topic performance

In this section, we discuss the effect of our approaches on individual topics of the FIRE 2010 shared task, both for Hindi-English and Tamil-English tasks. Figure 1 shows the difference in the Average Precision – topic-wise – between the baseline CLIR system and that integrated with our mining technique. Individual figures provide the differences for each topic, in each of the three configurations T, TD and TDN, for the Tamil-English tasks. We see that many more topics benefited from the mining technique; for example, in the Tamil-English language pair, in TDN configurations, 11 topics were improved (with 3 of them with an improvement of ≥ 0.2 in MAP score) whereas only 4 topics were negatively impacted (all of them dropping < 0.2 in MAP score). Similar trends could be seen for all configurations, in the Tamil-English tasks.

Fig. 1. Differences in Average Precision between the baseline and CLIR with mining

From Table 6, we note that the retrieval performance in Tamil-English test collection mining brings maximum improvement of 6% over the baseline in TDN setup.

5. Conclusion

In this paper, we underscored the need for handling proper and common nouns for improving the retrieval performance of cross-language information retrieval systems. We proposed and outlined two techniques for handling out of vocabulary (OOV) words – using transliteration generation, and transliteration equivalents mining – to enhance a state of the art baseline CLIR system. We presented the performance of our system under various topic configurations, specifically for English monolingual task and Tamil-English cross-language tasks, on the standard FIRE 2010 dataset. We show that the performance of our baseline CLIR system is improved significantly by each of the two techniques for handling OOV terms, but consistently more so by the mining technique. Significantly, we also show empirically that the performance of the CLIR system enhanced with transliteration mining is close to that of monolingual performance, validating our techniques for handling OOV terms in the cross-language retrieval.

References

- NTCIR: http://research.nii.ac.jp/ntcir/


A Package for Learning Negations in Tamil

Dr. G. Singaravelu
Reader, UGC-Academic Staff College,&B.Ed Coordinator
Bharathiar University, Coimbatore-641 046.

Introduction

Tamil is an important language to learn different cultures of Tamilnadu and India. Teaching of Tamil is difficult to the teachers of Tamil due to the more letters in Tamil and learning Tamil grammar is difficult to the learners of primary and upper primary schools due to ineffective methods of teaching. Grammar is indispensable for learning any language. Maximum teaching methods of grammar is adopting formal grammar. Less concentration is transacted in the classroom of the Tamil language in functional grammar. Conventional methods of teaching of functional grammar are ineffective and it lead the learners towards aversion in learning grammar. Negations have unique place in communicative competency. Conventional methods discourage the students to learn negation effectively at school level. Students are able to use it inappropriately. This study investigates the effectiveness of learning package of Negations in Tamil among the learners of standard VI.

Need of the study

Conventional methods are unable to create the appropriate learning atmosphere for scoring more marks in Tamil grammar of the mother tongue of the learners and also for the learners of the second language as Tamil. Traditional methods did not help the learners to learn Negations in Tamil. It was a challenging task to the learners of standard VI. An innovative Learning package can be encouraged the young learners to learn more negations in limited time. Hence the researcher endeavoured to prepare a learning package for acquiring more negations in Tamil for the young learners.

Objectives

The researcher has framed the following objectives of the study:

1. To find out the problems of conventional methods in learning Negations in Tamil at Government school.
2. To find out the problems of conventional methods in learning Negations in Tamil at Aided school.
3. To find out the significant difference in achievement mean score between the pre test of control group and the post test of control group in Government school.
4. To find out the significant difference in achievement mean score between the pre test of control group and the post test of control group in Aided school.
5. To find out the significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Government school.
6. To find out the significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Aided school.

7. To find out the impact of innovative Learning package in Negations of Tamil at standard VI in Government school and Aided school.

Hypotheses

The research has framed the following hypotheses

1. Students of standard VI have problems of conventional methods in learning Negations in Tamil at Government school.

2. Students of standard VI have problems of conventional methods in learning Negations in Tamil at Aided school.

3. There is no significant difference in achievement mean score between the pre test of control group and the post test of control group in Government school.

4. There is no significant difference in achievement mean score between the pre test of control group and the post test of control group in Aided school.

5. There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Government school.

6. There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Aided school.

7. To find out the impact of innovative Learning package in Negations of Tamil at standard VI in Government school and Aided school.

Method of study

Methodology: Equivalent group Experimental method was adopted in the study.

Sample selected for the study

Sixty pupils of studying in standard VI from Government Higher Secondary school, Kalveeranpalayam, Coimbatore and another Sixty pupils of studying in standard VI from Maruthamalai Devasdanam Subramanian swamy Higher secondary school, Vadavalli, Coimbatore were selected as sample for the study. Sixty students were considered as Controlled group and another Sixty were considered as Experimental group.

Instrumentation

Researcher’s self-made achievement test was used as a tool for the study.

Reliability of the tool

Test-retest method was used for the study. The co-efficient correlation was found 0.85 in the tool through test-retest method.
ValiditY of the tool

Content validity was established for the test through expert suggestions. Hence reliability and validity were properly established for the study.

Statistical Technique

Percentage, mean, SD and t test were adopted in the study for analyzing the tabulated data.

Procedures of the study:

Phase 1: Assessing the problems of the students in acquiring competency in learning Tamil Negations for both schools of Govt and Aided in existing methods through administering pretest.

Phase 2: Pre-production stage.

Phase 3: Production stage.

Phase 4: Preparation of package

Phase 5: Execution of activities through using the learning package

Phase 6: Administering pretest and post test to the control group and tabulated the scores.

Phase 7: Administering pre test and post test to the Experimental group and tabulated the scores.

Phase 8: Finding the effectiveness of the Package for Negation.

Data collection:

The researcher administered a diagnostic test to identify the problems of the students in learning Tamil with permission of Principals of the schools. Pretest –Treatment-Posttest was used in the study.

Hypothesis testing

Hypothesis 1&2

1. Students of standard VI have problems of conventional methods in learning Negations in Tamil at Government school.

2. Students of standard VI have problems of conventional methods in learning Negations in Tamil at Aided school.

In the pre-test, students of Govt schools and Aided schools score 19%, 28% marks respectively in acquiring Negation in Tamil through conventional method and the Experimental group students score 49 %, 56% marks respectively. It shows the problems of acquisition of Negation in Tamil through conventional methods among the students.

Hypothesis 3:

There is no significant difference in achievement mean score between the pre test of control group and the post test of control group in Government school.
The calculated t value is (0.08) less than table value (1.96). 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 teachers in learning Tamil Negations in Govt school.

**Hypothesis 4:**

There is no significant difference in achievement mean score between the pre test of control group and the post test of control group in Private school.

<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>30</td>
<td>10.63</td>
<td>3.23</td>
<td>58</td>
<td>0.08</td>
<td>insignificant</td>
</tr>
<tr>
<td>Post test control group group</td>
<td>30</td>
<td>10.78</td>
<td>3.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The calculated t value is (1.85) less than table value (1.96). 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 teachers in learning Tamil Negations in private school.

**Hypothesis 5:**

There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Government school.

<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>Pre test Experimental group</td>
<td>30</td>
<td>10.62</td>
<td>3.23</td>
<td>58</td>
<td>7.14</td>
<td>significant</td>
</tr>
<tr>
<td>Post test Experimental group</td>
<td>30</td>
<td>16.56</td>
<td>3.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Achievement mean scores between pre test of Experimental and posttest of Experimental group.

The calculated t value is (7.14) greater than table value (1.96). Hence null hypothesis is rejected at 0.05 levels. Hence there is significant difference between the pretest of experimental group and post test of experimental group in achievement mean scores of the students in learning Negation in Tamil.

Hypothesis 6:

There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Aided school.

Achievement mean scores between pretest of experimental group and posttest of Experimental group.

The calculated ‘t’ value is (7.08) greater than table value (1.96). Hence null hypothesis is rejected at 0.05 levels. Hence there is significant difference in achievement mean score between the pre test of Experimental group and post test experimental group in achievement mean scores of the students in Tamil Negation.

Hypothesis 7.

Learning package is more effective than conventional learning in learning Negation in Tamil

The above two tables prove and confirm the Learning Package is more effective than traditional approaches in developing Negation in Tamil. Mean scores in pre-test of Experimental group is (10.62 and 13.70) greater than the mean score of post test of Experimental group by using Learning Package in acquiring Negation in Tamil (16.56 and 19.65).

Findings:

1. Students of standard VI have problems of conventional methods in learning Negations in Tamil at Government school.

2. Students of standard VI have problems of conventional methods in learning Negations in Tamil at Aided school.

3. There is no significant difference in achievement mean score between the pre test of control group and the post test of control group in Government school.
4. There is no significant difference in achievement mean score between the pre test of control group and the post test of control group in Aided school.

5. There is significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Government school.

6. There is significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group in Aided school.

7. Learning package in Negations of Tamil is more effective than conventional methods in learning Tamil Negation at standard VI in Government school and Aided school.

Educational Implications:

1. Learning package can be prepared for other subjects also.

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

3. It may be implemented in Higher education

Conclusion

The study reveals that the students have problems in learning Negation in Tamil by using traditional approaches. Learning Package is more effective in Learning Tamil Negation. Hence it will be more supportive to promote the learners in learning Tamil.

References


- INFIT (2009) Conference papers, Tamil internet 2009, University of Cologne: Germany

CONFERENCE PAPERS

TAMIL INTERNET 2011

பற்றிய நூற்றுற்று நூற்றன்

Penn

University of Pennsylvania

INFITT

International Forum for Information Technology in Tamil

முன்னெட்டு நூற்றுற்று

Tamil Internet 2011
Morphology based Factored Statistical Machine Translation (F-SMT) system from English to Tamil

Anand Kumar M1, Dhanalakshmi V1, Soman K P1, Rajendran S2

1Computational Engineering and Networking
Amrita Vishwa Vidyapeetham Coimbatore, India
{m_anandkumar,v_dhanalakshmi,kp_soman}@cb.amrita.edu
2Tamil University, Thanjavur, India

Abstract

This paper presents a novel preprocessing methodology in factorized Statistical Machine Translation system from English to Tamil language. SMT system considers the translation problem as a machine learning problem. Statistical machine translation system for morphologically rich languages is a challenging task. Moreover it is very complex for the different word order language pair. So a simple SMT alone would not give good result for English to Tamil, which differs in morphological structure and word order. A simple SMT system performs only at the lexical level mapping. Because of the highly rich morphological structure of Tamil language, a simple lexical mapping alone will suffer a lacuna in collecting all the morphological and syntactic information from the English language. The proposed SMT system is based on factored translation models. The factored SMT uses machine learning techniques to automatically learn translation patterns from factored corpora. Using the learned model FSMT predicts the output factors for the given input factors. Using the Tamil morphological generator the factored output is synthesized.

Introduction

Statistical approach to machine translation learns translation patterns directly from training sentences and generalized them to handle new sentences. When translating from simple morphological language to the rich morphological language, the SMT baseline system will not generate the word forms that are not present in the training corpora. For training the SMT system, both monolingual and bilingual sentence-aligned parallel corpora of significant size are essential. The corpus size decides the accuracy of machine translation. The limited availability of parallel corpora for Tamil language and high inflectional variation increases a data sparseness problem for phrase-based SMT. To reduce the data sparseness, the words are split into lemma and their inflected forms based on their part of speech. Factored translation models [Koehn and Hoang, 2007] allow the integration of the linguistic information into a phrase-based translation model. These linguistic features are treated as separate tokens during the factored training process.
\[
P(T | E) = \frac{P(T) \cdot P(E | T)}{P(E)}
\]

\[T^* = \text{argmax } P(T) \cdot P(E | T)\]

SMT works on the above equation. Where T represents Tamil language and E represents English language. We have to find the best Tamil translation sentence \(T^*\) using \(P(T)\) and \(P(E | T)\), Where \(P(T)\) is given by the Language model and \(P(E | T)\) is given by the translation model.

**Factored SMT for Tamil**

Tamil language is morphologically rich language with free word order of SOV pattern. English language is morphologically simple with the word order of SVO pattern. The baseline SMT would not perform well for the languages with different word order and disparate morphological structure. For resolving this, we go for factored SMT system (F-SMT). A factored model, which is a subtype of SMT [Koehn and Hoang, 2007], will allow multiple levels of representation of the word from the most specific level to more general levels of analysis such as lemma, part-of-speech and morphological features. A preprocessing module is externally attached to the SMT system for Factored SMT.

The preprocessing module for source language includes three stages, which are reordering, factorization and compounding. In reordering stage the source language sentence is syntactically reordered according to the Tamil language syntax using reordering rules. After reordering, the English words are factored into lemma and other morphological features. A compounding process for English language is then followed, in which the various function words are removed from the reordered sentence and attached as a morphological factor to the corresponding content word. This reduces the length of English sentence. Now the representation of the source syntax is closely related to the target language syntax. This decreases the complexity in alignment, which is also a key problem in SMT from English to Tamil language.

Parallel corpora and monolingual corpora are used to train the statistical translation models. Parallel corpora contains factored English sentences (using Stanford parser) along with its factored Tamil translated sentences (using Tamil POS Tagger [V Dhanalakshmi et.al, 2009] and Morphological analyzer [M Anand kumar et.al,2009]. Factorized monolingual corpus is used in the Language model.

The parsed source language is reordered according to the target language structure using the syntax based reordering system. A compounding process for English language is then followed, in which the various function words are removed from the reordered sentence and attached as a morphological factor to the corresponding content word. This reduces the length of English sentence. Now the representation of the source syntax is closely related to the target language syntax. This decreases the complexity in alignment, which is also a key problem in SMT from English to Tamil language.

The factored SMT system’s output is post processed, where the Tamil Morphological generator is pipelined to generate the target sentence. Figure.1 shows the architecture of the prototype factored SMT system from English to Tamil.
Morphological models for Tamil language

Morphological models for target language Tamil are used in preprocessing as well as post processing stage. In preprocessing, Tamil POS tagger and Morphological analyzer are used to factorize the Tamil parallel corpus and monolingual corpus. Morphological generator is used in the post processing stage to generate the Tamil words from Factored SMT output.

Tamil POS tagger

Parts of speech (POS) tagging means labeling grammatical classes i.e. assigning parts of speech tags to each and every word of the given input sentence. POS tagging for Tamil is done using SVM based machine learning tool [V Dhanalakshmi et.al, 2009], which make the task simple and efficient. The SVM Tool[] is used for training the tagged sentences and tagging the untagged sentences. In this method, one requires Part of speech tagged corpus to create a trained model.

Tamil Morphological Analyzer

The Tamil morphological analyzer is based on sequence labeling and training by kernel methods. It captures the non-linear relationships and various morphological features of natural language in a better and simpler way. In this machine learning approach two training models are created for morphological analyzer. These two models are represented as Model-I and Model-II. First model is trained using the sequence of input characters and their corresponding output labels. This trained model-I is used for finding the morpheme boundaries [M Anand kumar et.al, 2009].

Second model is trained using sequence of morphemes and their grammatical categories. This trained Model-II is used for assigning grammatical classes to each morpheme. The SVMTool is used for training the data. Generally SVMTool is developed for POS tagging but here this tool is used in morphological analysis.

Tamil Morphological Generator

The developed morphological generator receives an input in the form of lemma+word_class+Morphological Information, where lemma specifies the lemma of the word-form to be generated, word_class
specifies the grammatical category (POS category) and Morpho-lexical Information specifies the type of inflection. The morphological generator system needs to handle three major things; first one is the lemma part, then the word class and finally the morpho-lexical information. By the way the generator is implemented makes it distinct from other morphological generator[M Anand kumar et.al,2010].

The input which is in Unicode format is first Romanized and then the paradigm number is identified by end characters. For sake of easy computation we are using romanized form. A Perl program has been written for identifying paradigm number, which is referred as column index. The morpho-lexical information of the required word class is given by the user as input. From the morpho-lexicon information list the index number of the corresponding input is identified, this is referred as row index. A verb and noun suffix tables are used in this system. Using the word class specified by the user the system uses the corresponding suffix table. In this two-dimensional suffix table rows are morpho-lexical information index and columns are paradigm numbers.

Conclusion

In this paper, we have presented a morphology based Factored SMT for English to Tamil language. The morphology based Factored SMT improves the performance of translation system for morphologically rich language and also it drastically reduces the training corpus size. So this model is suitable for languages which have less parallel corpus. Tamil morphological models are used to create a factorized parallel corpus. Source language reordering module captures structural difference between source and target language and reorder it accordingly. Compounding module converts the source language structure to fit into the target language structure. Initial results obtained from the Factored SMT are encouraging.

References

Tamil Shallow Parser using Machine Learning Approach

Dhanalakshmi V¹, Anand Kumar M¹, Soman K P¹ and Rajendran S²

¹Computational Engineering and Networking
Amrita Vishwa Vidyapeetham Coimbatore, India
{m_anandkumar,v_dhanalakshmi, kp_soman} @cb.amrita.edu
²Tamil University, Thanjavur, India

Abstract

This paper presents the Shallow Parser for Tamil using machine learning approach. Tamil Shallow Parser is an important module in Machine Translation from Tamil to any other language. It is also a key component in all NLP applications. It is used to understand natural language by machine and also useful for second language learners. The Tamil Shallow Parser was developed using the new and state of the art machine learning approach. The POS Tagger, Chunker, Morphological Analyzer and Dependency Parser were built for implementing the Tamil Shallow Parser. The above modules gives an encouraging result.

Introduction

Partial or Shallow Parsing is the task of recovering a limited amount of syntactic information from a natural language sentence. A full parser often provides more information than needed and sometimes it may also give less information. For example, in Information Retrieval, it may be enough to find simple NPs (Noun Phrases) and VPs (Verb Phrases). In Information Extraction, Summary Generation, and Question Answering System, information about special syntactico-semantic relations such as subject, object, location, time, etc, are needed than elaborate configurational syntactic analyses. In full parsing, grammar and search strategies are used to assign a complete syntactic structure to sentences. The main problem here is to select the most possible syntactic analysis to be obtained from thousands of possible analyses a typical parser with a sophisticated grammar may return. This complexity of the task makes machine learning an attractive option in comparison to the handcrafted rules.

Methodology

Machine learning approach is applied here to develop the shallow parser for Tamil. Part of speech tagger for Tamil has been generated using Support Vector Machine approach [Dhanalakshmi V e.t.al., 2009]. A novel approach using machine learning has been built for developing morphological analyzer for Tamil [Anand kumar M e.t.al., 2009]. Tamil Chunker has been developed using CRF++ tool [Dhanalakshmi V e.t.al., 2009]. And finally, Tamil Dependency parser, which is used to find syntactico-semantic relations such as subject, object, location, time, etc, is built using MALT Parser [Dhanalakshmi V e.t.al., 2011].
General Framework and Modules

- The general block diagram for Tamil Shallow parser is given in Figure 1.

![Diagram of Tamil Shallow Parser](image)

**Figure 1. General Framework for Tamil Shallow Parser**

- **Tamil Part-of-Speech Tagger** [Dhanalakshmi et al., 2009]: The Part of Speech (POS) tagging is the process of labeling a part of speech or other lexical class marker (noun, verb, adjective, etc.) to each and every word in a sentence. POS tagger was developed for Tamil language using SVMTool [Jesús Giménez and Lluís Márquez, 2004].

- **Tamil Morphological Analyzer** [Anand Kumar et al., 2009]: Morphological Analysis is the process of breaking down morphologically complex words into their constituent morphemes. It is the primary step for word formation analysis of any language. Morphological Analyzer was developed using a novel machine learning approach and was implemented using SVMTool.

- **Tamil Chunker** [Dhanalakshmi et al., 2009]: Chunks are normally taken to be non recursive correlated group of words. Chunker divides a sentence into its major-non-overlapping phrases.
(noun phrase, verb phrase, etc.) and attaches a label to each chunk. Chunker for Tamil language was developed using CRF++ Tool[Sha F and Pereira F, 2003].

- **Tamil Dependency Parser for Relation finding** [Dhanalakshmi V et al., 2011]: Given the POS tag, Morphological information and chunks in a sentence, this decides which relations they have with the main verb (subject, object, location, etc.). Dependency parser was developed for Tamil language using Malt Parser tool [Joakim Nivre and Johan Hall, 2005].

**Dependency Parsing using Malt Parser**

MALT Parser Tool is used for dependency parsing, which uses supervised machine learning algorithm. Using this tool dependency relations and position of the head are obtained for Tamil sentence. There are 10 tuples used in the training data that can be user define. For Tamil dependency parsing, the following features are defined and others are set as NULL and are mentioned as ‘_’ in the training data format.

- **WordID**: Position of each word in the input sentence.
- **Words**: Each word in the input sentence.
- **CPos Tag and Pos Tag**: Defines the Parts Of Speech of each word.
- **Head**: The position of the parent of each word.
- **Lemma**: The lemma of the word.
- **Morph Features**: The Morphological features of the word.
- **Chunk**: The chunk information of the word.
- **Dependency Relation**: The terminology given for each parent - child relation.

**Sample Training Data**

```
1 இவ் _ <PRP> <PRP> 8 <N.SUB> _ _
2 அவைடகைள _ <NN> <NN> 3 <D.OBJ> _ _
3 வாகி _ <VNAV> <VNAV> 4 <ATT> _ _
4 ஸைமன்ற _ <VNAV> <VNAV> 6 <VNAV.MOD> _ _
5 தனொ _ <NN> <NN> 6 <NST.MOD> _ _
6 ஒம்பு _ <VNAV> <VNAV> 8 <V.COMP> _ _
7 அல்லா கி _ <PRP> <PRP> 8 <LOBJ> _ _
8 உன்னுறி பாகி காறி _ <VF> <VF> 0 <ROOT> _ _
9 . <DOT> <DOT> 8 <SYM> _ _
```

For Tamil language, a corpus of three thousand sentences is annotated with dependency relations and labels using the customized tag set (Table.1). The corpus is trained using the MALT Parser tool which generates a model. Using this model the new input sentences are tested.
Table 1 Shallow Dependency Tagset

<table>
<thead>
<tr>
<th>S.No</th>
<th>Tags</th>
<th>Description</th>
<th>S.No</th>
<th>Tags</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ROOT</td>
<td>Head word</td>
<td>5</td>
<td>NST-MOD</td>
<td>Spatial Time</td>
</tr>
<tr>
<td></td>
<td>N-SUB</td>
<td>Subject</td>
<td>6</td>
<td>SYM</td>
<td>Symbols</td>
</tr>
<tr>
<td>3</td>
<td>D-OBJ</td>
<td>Direct Object</td>
<td>7</td>
<td>X</td>
<td>Others</td>
</tr>
<tr>
<td>4</td>
<td>I-OBJ</td>
<td>Indirect Object</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Application of Shallow Parser**

Shallow parsers were used in Verbmobil project [Wahlster W, 2000], to add robustness to a large speech-to-speech translation system. Shallow parsers are also typically used to reduce the search space for full-blown, ‘deep’ parsers [Collins, 1999]. Yet another application of shallow parsing is question-answering on the World Wide Web, where there is a need to efficiently process large quantities of ill-formed documents [Buchholz and Daelemans, 2001] and more generally, all text mining applications, e.g. in biology [Sekimizu et al., 1998].

The developed Tamil Shallow Parser can be used to develop the following systems for Tamil language.

- Information extraction and retrieval system for Tamil.
- Simple Tamil Machine Translation system.
- Tamil Grammar checker.
- Automatic Tamil Sentence Structure Analyzer.
- Language based educational exercises for Tamil language learners.

**Conclusion**

Shallow Parsing has proved to be a useful technology for written and spoken language domains. Full parsing is expensive, and is not very robust. Partial parsing has proved to be much faster and more robust. Dependency parser is better suited than phrase structure parser for languages with free or flexible word order like Tamil. Fully functional Shallow Parser for Tamil gives reliable results. The Shallow Parser system developed for Tamil is an important tool for Machine Translation between Tamil and other languages.

**References**


CONFERENCE PAPERS

TAMIL INTERNET 2011
Ambiguities in Computer Assisted Tamil Language Processing

Sundar Baskar

Gondwana Research Institute, Gondwana University,
Kadmar, Kallakurichi, Tamil Nadu, SRM University, Coimbatore. Email: sundarbaskar@gmail.com

Abstract

Ambiguities in Tamil Language Processing are a major challenge in Natural Language Processing. Tamil Language Processing involves the processing of Tamil text using computer algorithms. There are several types of ambiguities in Tamil Language Processing, such as morphological ambiguity, syntactic ambiguity, and semantic ambiguity. These ambiguities arise due to the complex nature of the Tamil language, which has a rich and diverse grammar and vocabulary.

Morphological Ambiguity

Morphological ambiguity arises due to the fact that a single word can have multiple meanings depending on the context in which it is used. For example, the word "தமிழு" can mean "Tamil" or "language" depending on the context.

Syntactic Ambiguity

Syntactic ambiguity occurs when a sentence can be interpreted in multiple ways. For example, the sentence "தமிழ்க் க்குள் நன்னையுடன்" can be interpreted as either "I like my Tamil book" or "I like the Tamil book with you".

Semantic Ambiguity

Semantic ambiguity occurs when a word has multiple meanings that are not necessarily related to each other. For example, the word "பால்" can mean "fish" or "meat".

Conclusion

Ambiguities in Tamil Language Processing are a significant challenge in Natural Language Processing. Research in this area is ongoing, and there are several approaches being explored to overcome these ambiguities. Future work should focus on developing more robust and effective algorithms to handle these ambiguities.
விளையாட்டு பதவியல் குறிப்பிட்டிய தகவல் பதவியல் தகவல் பல்கலைக்கழகத்திற்கு ரீதியான பதவியல் பல்கலைக்கழகத்திற்கு பதவியல் (Parsing) பல்கலைக்கழகத்திற்கு பதவியல் பல்கலைக்கழகத்திற்கு (Sorting) பல்கலை குறிப்பிட்டிய தகவல் தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் (Word Sense Ambiguity).

தேவ்பக்தி பலிேடுகள் (Sandhi Checker), பலிேடுகள் பலிேடுகள் (Morphological Parser), பலிேடுகள் பலிேடுகள் (Syntactic Parser), பலிேடுகள் (Indexing) பலிேடுகள், பலிேடுகள் (Automatic Speech Recognizer-ASR), பலிேடுகள் பலிேடுகள் (Machine Translation) பலிேடுகள் பலிேடுகள் பலிேடுகள் கணினியியல் கணினியியல் கணினியியல் கணினியியல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல் தகவல் தகவல். 'Word Sense Disambiguation' WSD(‘Word Sense Disambiguation’)

அம்மைய மாகங்கள் செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல். செய்தல் பதவியல் பல்கலைக்கழகத்திற்கு தகவல் தகவல் தகவல்.
1. கலனினமுறை  வரும்பிள்ளை  பாடல்கொள்கிறான்

2. பதிவுகொள்கிறார் தெளிவான உயர்வு மாதம்
3. பெருக்கடையில் பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும்

கைழியும் பெருக்கடையில் பம்பையல் மட்டும் திட்டும். அமாவாசை விழாவின் போது ஆண்டுகாலம் ஒரு நேரப் பலியடித்து. இவ்வாறு பெருக்கடையில் பெருமையான கைழியில் வேட்டவர் திட்டும் லேனின் பெருமை மாறும்.

கைழியை பெருமலச் பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும் (Auxiliary Verb), வள்ளை (Affixes) பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும் லேனின் பெருமை மாறும் லேனின் பெருமை மாறும்.

பெருக்கடையில் ஒரு பெருமலச் பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும் ஆண்டுகாலம் மூன்று பெருமலச் பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும்.

அவர்கள் ஒரு பெருமலச் பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும் (Affixes) பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும் லேனின் பெருமை மாறும். பிள்ளையார் வேட்டவர் திட்டும் லேனின் பெருமை மாறும் லேனின் பெருமை மாறும்.
பார்டியல் ஆணையத்தில்

துச்சரம் தலை துணையாளர் தகத் மாதி மாடாக துத்தகத் திசையான. பாதியாளருக்கு துணையாளர் தருள்தான் காதிரியாக பெறப்படும் பொருள் ‘காதிரியாகப் பெறப்படும்’ சொல்லு. குறுகிய துணையாளர் பதிமானம் (Morphological Parsers) துணையாளர் பதிமானத்திற்கு. இதனை துணையாளர் பதிமானத்திற்கு காதிரியாக பெறுவது தன்னைச் சீர்க்கக் கூறுவது தொடரும்.
'அந்த இன்மையான மீத்தரை' தங்க நாட்டை, அந்த இன்மையான மீத்தரை(இருமுகர்) மீத்தரை தங்க குறிப்பிட்டுள்ள இருவருக்கு பாலத்துவதற்கு? அவ்விலை அந்த தொன்மைத் திண்மநிலை (இருமுகர்+தர்சரு+அல்) புது மீத்தரை தங்க நிலைத்தேக்கு? தங்க நிலை ஓர்மக்குறிக்கு. இத்தகைய திண்மநிலை நிலைத்தேக்கு அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக் குறிப்பிட்டுள்ள அலையக்

அண்டர்க்ரி மூன்றாம் பாதையான இன்மைவாய்வதாக வைக்கப்பட்டது. நிறுவனத்தை வைக்கப்பட்ட மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் பாதையான அண்டர்க்ரி மூன்றாம் 

அந்த தொன்மையான மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) மீத்தரை (இருமுகர்) 

கணினியில் இருந்த அகலமும் துறைப்பட்டு
வாிவᾊவΆ மிகᾫΆ ேதைவᾺபᾌΆ ஒᾹᾠ. தமிழ் இலᾰகியᾱகைள் பலறிய ᾲ ெசᾼᾜΆ ேநாᾰகி᾿ பலᾞΆ யᾹᾠ வᾞகிᾹறன᾽. கா᾵டாக, மᾐைர唳தி᾵டΆ தமிழ் இலᾰகியᾱகைள ராமᾹ வᾊவᾺெபய᾽ᾺபிᾤΆ தᾞகிறᾐ. ைலᾺராி ஆᾺ காᾰகிரᾆΆ அெமாிᾰக ஐேராᾺப ிய நாᾌகளிᾤ῀ள அவ᾽க唑ைடய ᾓலகᾱகorderby வைலயตำบதி᾿ தΆ ᾓ᾿ ப᾵ᾊயைல, வᾊᾺபா᾽ஃஆΆ வசதிᾜடᾹஇவιறி׳ ஆசாைகைளளᾀ பல ராமᾹ எᾨᾑᾐᾰகorderby பயᾹபᾌᾐகிᾹறன. தமிழ்ெமாழி உஇளன.

Digital Dictionaries of South India எｧற தளตำบதி׳ பல மாழி அகராதிகஇளாத உஇளன. இவைறி׳ ஆசாைகைளளᾀ பலட ராமᾹ எᾨᾐᾰகorderby பயᾹபᾌᾐகிᾹறன. தமிழ்ெமாழி எᾨᾐவதிா அதிகம் பழᾱகமிாலாத தமிழ்ெைலாகorderby மிாமயாக வா淝வᾊவᾺ அவίறிழகான சிறᾺจีܣகᾛᾌகளிᾹறிᾺ பயᾹபᾌᾐகிᾤᾐகிᾺற ந. கணினியிேல தமிழ்ெசᾼயᾺ பயி்ிஆ் இைரய ி᾿ திι部副ᾐ உைரய ி᾿ திι部副ᾐ ஆᾱகில உைரᾰᾁᾲ ஆᾱகிலᾙΆ கீfolios திைரயிா தமிழ்ெமᾹமமானᾐ இᾞ திைரகைள உᾞவாᾰகில் திைரயி� தமிழ்ெமᾹமణீல இᾞ திைரகைள உᾞவாᾰகில் திைரயி� தமிழ்ெமᾹமణீல இᾞ திைரகைள உᾞவாᾰகில் திைரயி� தமிழ்ெமᾹಮణீல இᾞ திைரகைள உᾞவாᾰகில் திைரயி� தமிழ்ெமᾹமణீல இᾞ திைரகைள உᾞவாᾰகில் திைரயி� தமிழ்ெமᾹமణீல இᾞ திைரகைள உᾞவா آلاف ஆᾱகில் உைரᾰᾁᾲ ஆᾱகிலᾙΆ கீfolios திைரயிா தமிழ்ெமᾹமణீல இᾞ திைரகைள உᾞவா آلاف ஆᾱகில் உைரᾰᾁᾲ ஆᾱகிலᾙΆ கீfolios திைரயிா தமிழ்ெமᾹமణீல இᾞ திைரகைள உᾞவா آلاف ஆᾱகில் உைரᾰᾁᾲ ஆᾱகிலᾙΆ கீfolios திைரயிா தமிழ்ெмᾹமణீல இᾞ திைரகைள உᾞவா آلاف ஆᾱகில் உைர predecess

1926-36 களி�ைனஃகளி் பைகள் அழகி எᾨᾐᾰகைளᾜΆ சில சிறᾺஃகளி் பயᾹபᾌ ᾷதின . இஸைதᾀ த淝ைவᾠ் 1912-இ் ஏெதᾢசி் நடᾸத கீைழᾇேதயᾇைாி் பᾹனா கீᾌᾰ கழக மாநா᾵ᾊᾹ பாிᾸᾐைரைய ஒ᾵ᾊயᾐ. Universal Digital Library எைறவிஹானி் பாலண் பாலன் எைறவிஹானி் கீ��ணலான நெடையும் பயாதிகள். ஆவாத் என்னகாந் ஆராயின் நேமுறவு உரை ஈந்தலளவு வழியாது குறிப்பிட்டலாம்.

• 1926-36 களியையலையம் மூலச்சாரைோந்தம் இனைகள்ே் கையுடைய பக்காத் பிறானா ஆண் அகுறையின் மூலத்தமிர் பாரல்லாமல் நிலவலாம். பிறிருந்தை பாரிச்சார்ந்ததை இலைன் பிறானா நேமுறவு உரை ஈந்தலளவு வழியாது குறிப்பிட்டலாம்.

• ITRANS- ஆவாத் என்னகாந் என்னகாந் பயாதிகள் மூலம்பையறக்கின். இலையர்கள் 1912-இ் ஏெத்தைய நிலவலாம் என்ன குறிப்பிட்டு பாரிச்சார்ந்ததை இலையர்கள் மூலம்பையறக்கின். ஆவாத் என்னகாந் என்னகாந் பயாதிகள் மூலம்பையறக்கின்.
• நூற்றாண்டிற்கும் முன்புக்கு பூர்வத்தாக, பார்த்து விளையாட்டு, இரண்டு மூன்று செழுத்துகள், கட்டுடன் தொடர்கொள்ளும் கட்டுப்பாடு தொடர்விளையாடும் இயல்பு வாய்ந்த ஆறு குறுகிகளில் சில முதல்நிலைகள், அவற்றின் யாராவரையாக இருக்கவிடுது.
<table>
<thead>
<tr>
<th>விளக்கம்</th>
<th>கருள் முறையும்</th>
<th>வகையெடுப்பு</th>
<th>பின்னர்</th>
<th>பயன்பாடு</th>
<th>புனித</th>
<th>புனித</th>
<th>ISO (1) 15919</th>
<th>ISO (2)</th>
<th>Penn. Uty</th>
<th>TL</th>
<th>LC</th>
<th>Mad urai</th>
<th>KOE LN</th>
<th>Google Indic</th>
<th>ITRAN S</th>
</tr>
</thead>
</table>
| குணா அந்தாய் பயன்பாடு தொடச் செய்ய்வு தொடச் செய்ய்வு பயன்பாடுப் பயன்பாடு பயன்பாடுப் பயன்பாடு
| குணா | au | au | au | au | au | au | au | au | au | au | au | au | au | au/A | au |

- குணா அந்தாய் பயன்பாடு தொடச் செய்ய்வு தொடச் செய்ய்வு பயன்பாடுப் பயன்பாடு பயன்பாடு

<table>
<thead>
<tr>
<th>விளக்கம்</th>
<th>கருள் முறையும்</th>
<th>வகையெடுப்பு</th>
<th>பின்னர்</th>
<th>பயன்பாடு</th>
<th>புனித</th>
<th>புனித</th>
<th>ISO (1) 15919</th>
<th>ISO (2)</th>
<th>Penn. Uty</th>
<th>TL</th>
<th>LC</th>
<th>Mad urai</th>
<th>KOE LN</th>
<th>google Indic</th>
<th>ITRAN S</th>
</tr>
</thead>
</table>
| குணா அந்தாய் பயன்பாடு தொடச் செய்ய்வு தொடச் செய்ய்வு பயன்பாடுப் பயன்பாடு பயன்பாடு
| குணா | au | au | au | au | au | au | au | au | au | au | au | au | au | au/A | au |

- குணா அந்தாய் பயன்பாடு தொடச் செய்ய்வு தொடச் செய்ய்வு பயன்பாடுப் பயன்பாடு பயன்பாடு

<table>
<thead>
<tr>
<th>விளக்கம்</th>
<th>கருள் முறையும்</th>
<th>வகையெடுப்பு</th>
<th>பின்னர்</th>
<th>பயன்பாடு</th>
<th>புனித</th>
<th>புனித</th>
<th>ISO (1) 15919</th>
<th>ISO (2)</th>
<th>Penn. Uty</th>
<th>TL</th>
<th>LC</th>
<th>Mad urai</th>
<th>KOE LN</th>
<th>google Indic</th>
<th>ITRAN S</th>
</tr>
</thead>
</table>
| குணா அந்தாய் பயன்பாடு தொடச் செய்ய்வு தொடச் செய்ய்வு பயன்பாடுப் பயன்பாடு பயன்பாடு
<p>| குணா | au | au | au | au | au | au | au | au | au | au | au | au | au | au/A | au |</p>
<table>
<thead>
<tr>
<th>மிகு விளக்கம்</th>
<th>விளக்கம்</th>
<th>பட்டையூடு</th>
<th>பட்டையூடு</th>
<th>ISO (1)</th>
<th>ISO (2)</th>
<th>Pen. Ut. y</th>
<th>T</th>
<th>L</th>
<th>C</th>
<th>Madurai</th>
<th>KOELN</th>
<th>google Indic</th>
<th>ITRA NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>க</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>k</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K/g</td>
<td>k, kh, g, gh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ச</td>
<td>s</td>
<td>ch, j</td>
<td>c, j</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>C</td>
<td>ch</td>
<td>c, ch</td>
<td></td>
</tr>
<tr>
<td>ஞ</td>
<td>t</td>
<td>t, d</td>
<td>t, th</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t, th</td>
<td>t</td>
<td>t</td>
<td>T</td>
<td>th</td>
<td>ta, th, d, dh</td>
<td></td>
</tr>
<tr>
<td>ப</td>
<td>p</td>
<td>p, b</td>
<td>p, b</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>p</td>
<td>P</td>
<td>p</td>
<td>p, ph, b, bh</td>
<td></td>
</tr>
</tbody>
</table>

- க, ச, ஞ, ப என்ற பட்டையூடுகளும் குறிப்பிட்டு விளக்கம் m, y, r, l, v

- KOELN கெறைசெய் ஒலிப்பு நிறுவனம்: M, Y, R, L, V என்ற பட்டையூடுகள்.
<table>
<thead>
<tr>
<th>மொழி விளக்கங்கள்</th>
<th>கணக்கோடரிக்கும் குறுக்குகள்</th>
<th>வருவதாயிருக்கும் குறுக்குகள்</th>
<th>சொல் பாடல்</th>
<th>பட்டியல்</th>
<th>ISO (1) 15919</th>
<th>ISO (2)</th>
<th>Penn. UTy</th>
<th>TL</th>
<th>LC</th>
<th>Madurai</th>
<th>KO</th>
<th>google</th>
<th>Indic</th>
<th>ITR</th>
<th>ANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ம</td>
<td>m</td>
<td>M</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>M</td>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
<tr>
<td>ப</td>
<td>y</td>
<td>Y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>Y</td>
<td>y</td>
<td>y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>இ</td>
<td>r</td>
<td>R</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>R</td>
<td>r</td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ட்</td>
<td>l</td>
<td>L</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>L</td>
<td>l</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>பம்</td>
<td>v</td>
<td>V</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>V</td>
<td>v</td>
<td>v</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>பொருள் எழுத்து</td>
<td>குறியீடு</td>
<td>பொருள் எழுத்து நுண்கூற்றம்</td>
<td>மொழியிட்டு வடிவம்</td>
<td>முதல் வடிவம்</td>
<td>பொருள் எழுத்து பொருள் எழுத்து</td>
<td>ISO (1)</td>
<td>ISO (2)</td>
<td>Penn. Uty</td>
<td>TL</td>
<td>LC</td>
<td>Madurai</td>
<td>KO EL N</td>
<td>google Indic</td>
<td>ITRA NS</td>
<td></td>
</tr>
<tr>
<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>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>ல</td>
<td>ட</td>
<td>ட, ட்</td>
<td>ட, ட்</td>
<td>ட</td>
<td>ட</td>
<td>#t</td>
<td>T; d</td>
<td>□</td>
<td>□</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T, Th, D, Dh</td>
<td>□ , □ h</td>
<td>d, dh</td>
</tr>
<tr>
<td>ங</td>
<td>ஐ</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>L</td>
<td>□</td>
<td>□</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L, □</td>
<td></td>
<td></td>
</tr>
<tr>
<td>மு</td>
<td>ஐ</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>_L</td>
<td>z; zh</td>
<td>□</td>
<td>□</td>
<td>z</td>
<td>Z</td>
<td>LZ/z</td>
<td>z</td>
<td></td>
</tr>
</tbody>
</table>

- ல - ட, ல
- ங - □, #1 ல
- மு - ஐ, □ □, _L zh, z, Z LZ
- ல, ல
<table>
<thead>
<tr>
<th>மொழி</th>
<th>பிரிவு</th>
<th>மாதிரி</th>
<th>கருவியான</th>
<th>இருபுரியக்</th>
<th>இருபுரிய</th>
<th>புரியக்</th>
<th>ISO (1) 1591</th>
<th>ISO (2)</th>
<th>Penn Uty</th>
<th>TL</th>
<th>LC</th>
<th>Madurai</th>
<th>KOE</th>
<th>google Indic</th>
<th>ITTRANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>மா</td>
<td>ng</td>
<td>*</td>
<td>Ń</td>
<td>Ń</td>
<td>*</td>
<td>^n</td>
<td>ng</td>
<td>*</td>
<td>*</td>
<td>ng</td>
<td>G</td>
<td>NG</td>
<td>~N, N^</td>
<td></td>
<td></td>
</tr>
<tr>
<td>கா</td>
<td>nj</td>
<td>Ń</td>
<td>Ń</td>
<td>Ń</td>
<td>Ń</td>
<td>*</td>
<td>n</td>
<td>nj</td>
<td>Ń</td>
<td>n</td>
<td>n</td>
<td>n^/j</td>
<td>NY</td>
<td>JN</td>
<td>Ń</td>
</tr>
</tbody>
</table>

- மா - ng, * , Ń, ^n, n, G, NG, ~N, N^ 
- கா - nj, Ń, n, n^/j, NY, JN

இயற்றின் பிற்பலன் ஒரு பக்கத்தில் பிற்பலன் மாநிலம் பற்றியவைகளைக் கொண்டுள்ளது. கொலையில் மாநிலம் பற்றியவைகளைக் கொண்டுள்ளது. தமிழ் எழுத்துக்கள் பற்றியவைகளைக் கொண்டுள்ளது. மாநிலம் பற்றியவைகள் பற்றியவைகளைக் கொண்டுள்ளது. மாநிலம் பற்றியவைகள் பற்றியவைகளைக் கொண்டுள்ளது. தமிழ் எழுத்துக்கள் பற்றியவைகளைக் கொண்டுள்ளது. மாநிலம் பற்றியவைகள் பற்றியவைகளைக் கொண்டுள்ளது. தமிழ் எழுத்துக்கள் பற்றியவைகளைக் கொண்டுள்ளது. மாநிலம் பற்றியவைகள் பற்றியவைகளைக் கொண்டுள்ளது. தமிழ் எழுத்துக்கள் பற்றியவைகளைக் கொண்டுள்ளது.
பாராம்பரியம் தமிழில் இந்தையியரை வரையறுக்கிறது உரையான பூர்வத்திய வரலாறு. வரலாறு
அடர்த்தியாகப் பாராம்பரியம் இளங்கல் காவல் காண்டு வரையறுக்கி வரையறுக்கிறது.

பாராளும்: pākā al -ISO (1) 15919 மேல் tami -ISO (1) 15919
p_a_rka#tal - ISO (2) tami 1 - ISO (2)
pāRkaDal, pARkaDal - Penn Uty tamiz/tamizh - PennUty
pā • ka • al - TL tami• - TL
pā • ka • al - LC tami• - LC
pARkaTal - Madurai tamiz - Madurai
pARkaTal - KOELN tamiz - KOELN
pARkaTal, paaRkaTal - Adami thamiz/tamizh - Adami

1. அறிஞாலன வல்லமை விளக்கங்கள்.
2. பொருள்பாடு தொடர்புப்படுத்தல் வாய்ந்த வேள் விளக்கங்கள் நூற்றணிமுறை நூற்றணி

3. வரலாறு சுற்றுக்கு பந்தனையில் உள்ளது, செய்டு செய்டு செய்டை அற்று பொருள்பாடு

4. வரலாறு பொருள்பாடுகள் வரலாறு அனுமானப்படுத்தல் பூர்வம் பூர்வம் அடர்த்தியாக
Electronic Tamil Dictionary

(மிா் அகராதி)
Agaraadhi: A Novel Online Dictionary Framework

Elanchezhiyan.K, Karthikeyan.S, T V Geetha,
Ranjani Parthasarathi & Madhan Karky

{chezhiyank@gmail.com, sethuramankarthikeyan@gmail.com, madhankarky@gmail.com}
Tamil Computing Lab (TaCoLa),
College of Engineering Guindy, Anna University, Chennai.

Abstract

This paper describes Agaraadhi, a dictionary framework for indexing and retrieving Tamil words, their meaning, analysis and related information. With a database of over 3 lac root words and their corresponding meaning in English and Tamil, this paper proposes a framework to encompass various features such as morphological analysis, morphological generation, word usage statistics, word pleasantness analysis, spell checking, similar word finder, word usage in literature, picture dictionary, number to text conversion, phonetic transliteration, live usage analysis from micro blogs and more. Describing various components of the framework the paper concludes with a discussion over dictionary statistics and possible features for future extension of the framework.

1. Introduction

Most of the Tamil dictionaries are synonym based and they do not give enough information such as morphological analysis of the word, possible case endings for requested word, pleasantness score, word usage in the web and social networks, equivalent words or meaning etc. To overcome these issues we propose Agaraadhi, a framework. Agaraadhi Framework consists of a Morphological analyser, Morphological generator, Word pleasantness and Word usage score finder as well as analysis of current usage in Social Networks, Picture dictionary, equivalent Tamil words, Generator (Word suggestions), Spell checker, Phonetic transliteration, Number to Text Converter, Rare-Word of the day and Social Network sharing.

Agaraadhi dictionary has more than 3 lac words in various domains such as General, Literature, Medical, Engineering, Computer Science, etc. The Agaraadhi framework dictionary is a Tamil English bilingual dictionary. The following sections describe the framework and list the benefits of such a framework over traditional online Tamil-English dictionaries. A few features proposed in this framework such as popularity score for a word, to best of our knowledge, are not present in any other world dictionaries.

2. Agaraadhi Framework

Agaraadhi Dictionary Framework was designed to provide additional information to the user regarding the word that they query about. Agaraadhi framework presented in figure 1 can be divided
into two major divisions, online and offline, in terms of the time of processing. This section describes the various components used in the Agaraadhi framework in detail.

2.1 Online Process

Any user query is sent sequentially to dictionary and literature, to retrieve corresponding data from those indices, fetching phonetic transliteration from transliteration modules, morphological information from morphological analysis and generator module and fetching live usage analysis from micro blogs. All those Information are sent to user interface pages, shown in fig 1.

![Fig 1: Agaraadhi Online Dictionary Framework](image)

2.2 Offline Process

Tamil words and their meanings are entered manually and stored in text files. Those words are sequentially sent to modules such as popularity score generator, pleasantness score generator, picture dictionary, phonetic transliteration module and the resulting information is abstracted as a word object. Tamil literature such as Bharathiyaar songs, Avvaiyar songs, Thirukkural and lyrics are crawled using a static web crawler and are indexed in hash table as key value pairs.

2.3 Features of Agaraadhi Framework

Agaraadhi dictionary framework consists of more than twenty features such as Morphological Analysis Morph Generation, pleasantness scoring, popularity scoring, spell error suggestions etc.
3.1 Morphological Analyser

Morphological analyser [1] chunks the query word and gives the morphological features of the query word such as root word, parts of speech, gender, tense and count. If the Query word is padithaan, Morphological Analyser gives as padi as root, word represents male gender and query word is past tense and so on.

3.2 Morphological Generator

A Tamil morphological generator[2] needs to tackle different syntactic categories such as nouns, verbs, post positions, adjectives, adverbs etc. separately, since the addition of morphological constituents to each of these syntactic categories depends on different types of information. The generator is used to generate possible morphological variations of the query word.

3.3 Spell Checker

Spell Checker is used to check the spelling of Tamil words and to provide alternative suggestions for the wrong words. It uses the Morphological Analyzer. The Morphological Analyzer is used to split the given Tamil word into the root word and a set of suffixes. If the word is fully split by the analyzer and its root word is also found in the Agaraadhi dictionary, the given word is termed as correct. Otherwise, the correction process is invoked to generate all the possible suggestions with minimum variations from the given word.

3.4 Word Suggestions

Word Suggestion gives the list of equivalent or related words for the given query word.

3.5 Word Pleasantness and Word Popularity Score

Word Pleasantness score generator provides how easy to pronounce the word.

Word Popularity shows the word usage in the web. The Word from agaraadhi is given to web and found the frequency distribution of the word across the popular blogs, news articles, social nets etc.

3.6 Word Usage in Literature

This feature finds the usage of words in popular literature such as Thirukural, Bharathiyar Padalgal, Avvai songs and Lyrics.

3.7 Number to Text Converter

It converts a number to Tamil word equivalent as well as in English text. For example in Tamil we represent oru Arpputham (அற்புதம்) for 100 million, Kumbam (கும்பம்) for 10 billion and finally up to Anniyan (அனியம்) for one zillion.

3.8 Phonetic Transliteration

The pronunciation of words in Tamil and English language, as distinct from their written form based on the phonology and it can also vary greatly among dialects of a language. Phonetic transliteration module splits the word into syllables and gives the transliteration for each syllable.
3.9 Picture Dictionary

Pictures, photos or line drawings to depict popular words have been included in the dictionary to enable efficient learning for children using this tool.

3.10 Social net Sharing and Twitter Update

The framework also provides features to format results to be shared effectively on social networks. An Agaraadhi Bot was designed to post updates and word of the day on Twitter automatically.

3.11 Word of the Day and Word Usage statistics

A rare word is randomly chosen and is displayed in the opening page to facilitate users to learn a new word every day.

Word Usage Statistics [3] shows the usage of the word in the social network over the past one week.

3.12 Tamil Word Games

Games play a vital role in learning. Currently Agaraadhi has two Tamil word games namely Miruginajambo and Thookku Thookki. Miruginajambo is an unscramble game and Thookku Thookki is a Hangman game in Tamil.

4. Conclusion and Future Work

This paper describes Agaraadhi, an Online Dictionary Framework. Agaraadhi online dictionary is a bilingual dictionary containing over 3 lac words on various domains like General, Medical, Engineering, Computer science, Literature etc. This Online Dictionary framework encompass various features such as morphological analysis, morphological generation, word usage statistics, word pleasantness analysis, spell checking, similar word finder, word usage in literature, picture dictionary, number to text conversion, phonetic transliteration, live usage analysis from micro blogs etc. Providing APIs for programmers and developing mobile apps for Agaraadhi framework will open a good platform for many researchers and developers working in Tamil Computing area.

References

தமிழ் கதை அகராதி

புதல்வி ச. தமிழ்கிலையால்,
இந்து மகா娠மபிறைப்
நேர்வியல் பள்ளிக்காட்சியால்,
ுண்மை - 632 014., ஒசூர் மாவட்டம். கதைக், இலக்கிய.

மின்சலர் : ktamilselvan@vit.ac.in

பரவலாறு:
இலங்கை விளையாட்டின் போதுகே, புதிய புதிய கலமறியியும் வகையான வேதியியல் விளைந்தது. இதன் கருரச்சடியாக புதிய புதிய வேதியியான வகையான வேதியியான வளர்ச்சியாக, Cell Phone-கள் மேம்பட்டுள்ளன, காரணம் காரணம் அறிவியான காரணம். இதன் படி, புதிய வேதியியான வருகையான பின்னர் நேர்வியல் அறிவியான நிறுவக் குடும்பளவை வருகையான காரணம் (வளர்ச்சி - வளர்ச்சி). வளர்ச்சிகளில் இலங்கையில் வேதியியான வருகையான பின்னர் நேர்வியல் அறிவியான நிறுவக் குடும்பளவை வருகையான காரணம்.
புது தமிழ் அகராதி

● தமிழ் ● English

மாநிலத்தில் உள்ள வட்டம் மின் மையம்

புது தமிழ் அகராதி (சூரணங்களில்) புதுப்புத்தற்கு
The first letter of Tamil alphabet

மீண்டும் - 2:
நிலை புது தமிழில் உள்ள வட்டம், "a" எனும் வட்டத்தில் உள்ள வட்டம் மின் மையம் அகராதி புதுப்புத்தற்கு மையம். புது தமிழில் உள்ள வட்டம் மின் மையம் அகராதி புதுப்புத்தற்கு மையம், நிலை புது தமிழில் மையம்.

புது தமிழ் அகராதி

● தமிழ் ● English

Enter your word  a

a
abacus
abbey
தமிழ் தமிழ் அகராதி

○ தமிழ் ○ English

Enter your word

டி ஒலியத் தமிழில் அகராதி (Prototype)

2

இந்த முக்கியான பல்லாம் (Prototype Approval)

3

வைரபடான இந்த பல்லாம் மாதிாிய (Prototype Approval)

4

வாிைச (Prototype Approval)

5

ப்ளைன் முழிய முழிய முழிய முழிய (QA Testing)

6

மாணவரின் முழிய முழிய பயண்முழிய (QA Testing)

7

இலைய வசதி இலைய கணிணியிைள் (Desktop Application)

8

மாணவரின் வசதி இலைய கணிணியிைள் (Desktop Application)

9

றாந்தை கணிணியிைள்.
<table>
<thead>
<tr>
<th>S.No.</th>
<th>Components</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Operating System</td>
<td>Windows 2000 Professional or XP or above</td>
</tr>
<tr>
<td>2</td>
<td>Application Server</td>
<td>ASP.NET 2.0, AJAX</td>
</tr>
<tr>
<td>3</td>
<td>Web Tools</td>
<td>HTML, Javascript</td>
</tr>
<tr>
<td>4</td>
<td>Database</td>
<td>MS SQL Server 2005</td>
</tr>
<tr>
<td>5</td>
<td>Browser</td>
<td>Internet Explorer, Mozilla Firefox, Netscape, Google chrome - with unicode support</td>
</tr>
</tbody>
</table>


நூற்றாண்டு வட்டியான கொள்கை ஆற்றியால் உள்ளது 100% சுற்றங்களும் தமிழ் எழுதிய வட்டியான நூற்றாண்டுகளுடன் உள்ளது. இது ஆற்றியால் உள்ளது கொள்கையின் வட்டியான நூற்றாண்டுகளுடன் உள்ளது. 100% Bug Free கொள்கையின் வட்டியான நூற்றாண்டுகளுடன் உள்ளது.
மாழியியி கலவில் அகராதியி பந்தமர்

இலங்கைச் சிவபை பிற்கருணை

சதுர்க்கருதியுடன் முடிரையின் நூற்றாண்டுகளும், விளையாடின

E-mail: s.ilangkumaran@gmail.com

205

1. இணை முறையில் அறிவியலும் செயற்கையும் பயணிகளையாக.

- பயணிகள் முறையில் செயற்கையும் பயணிகளையாக பயணிகளை பிரித்து கொள்ள, மேலும் முறையில் அறிவியலும் பயணிகளை முறையிலும் அறிவியலும் பயணிகளை பிரித்து கொள்ள. இதில் குறுக்கு செயற்கையும் அறிவியலும் பயணிகளை பிரித்து கொள்ள வேண்டும் அடிப்படையானது அறிவியலும் பயணிகளை பிரித்து கொள்ள வேண்டும் பயணிகளை வேண்டும் பயணிகளை வேண்டும்.
காலனித்துறையில் புதுப்பிக்கப்பட்டுள்ள புதுக்கோட்டேற்றம் கற்று ஓர் நிராகரிப்புப் படிகளில் ஒன்றாகும் நேரடி குறிப்பிட்டுள்ளது.

2. அவர்களுக்கு அருங்காட்சியடைந்த தரவுகளை.

- பின்னர் பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பரிமாறுவதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள். இந்த முனையில் இவ்வாறு உரியைகளை சேர்ச்சி செய்து பின்னர் பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பகிர்வதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள்.

2. அருங்காட்சிய பரிமாற்றக்கூடிய தரவுகளை காட்சிப்படைந்த சேர்ச்சிகளை.

- பின்னர் பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பரிமாறுவதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள். இந்த முனையில் இவ்வாறு உரியைகளை சேர்ச்சி செய்து பின்னர் பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பகிர்வதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள்.

3. முக்கியமான விஷயங்களை எடுத்துக்காட்டு.

- இந்த பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பகிர்வதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள். இந்த பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பகிர்வதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள். இந்த பகிர்வதற்கு அருங்காட்சியடைந்த தரவுகளை பகிர்வதற்காக காலனித்துறையில் ஓர் நிராகரிப்புப் படிகுகள் பதிவிட்டார்கள்.
4. என்று வெளிப்படுத்தி தமது கல்வியும் பணிகளும் வைக்கிறார்.

- தமது விளைவு பணிகளில் கல்வியும் பணிகளும் வைக்கிறார். இவ்வாறு விளைவு பணிகளில் கல்வி பெறுவதற்காக வெளிப்படுத்தி வைக்கிறார்.

5. பலதர் ஆராய்ச்சியின் பணிப்புகள்

- பலதர் ஆராய்ச்சியின் பணிப்புகள் தமது விளைவு பணிகளில் கிளையிலிருந்து விளைவு பணிகளில் கல்வியும் பணிகளும் வைக்கிறார். விளைவு பணிகளில் கல்வியும் பணிகளும் வைக்கிறார் என்று அறிய வேண்டும்.
Agaraadhi: A Novel Online Dictionary Framework

{chezhiyank@gmail.com, sethuramankarthikeyan@gmail.com, madhankarky@gmail.com}
Tamil Computing Lab (TaCoLa), College of Engineering Guindy, Anna University, Chennai.

Abstract
This paper describes Agaraadhi, a dictionary framework for indexing and retrieving Tamil words, their meaning, analysis and related information. With a database of over 3 lac root words and their corresponding meaning in English and Tamil, this paper proposes a framework to encompass various features such as morphological analysis, morphological generation, word usage statistics, word pleasantness analysis, spell checking, similar word finder, word usage in literature, picture dictionary, number to text conversion, phonetic transliteration, live usage analysis from micro blogs and more. Describing various components of the framework the paper concludes with a discussion over dictionary statistics and possible features for future extension of the framework.

1. Introduction
Most of the Tamil dictionaries are synonym based and they do not give enough information such as morphological analysis of the word, possible case endings for requested word, pleasantness score, word usage in the web and social networks, equivalent words or meaning etc. To overcome these issues we propose Agaraadhi, a framework. Agaraadhi Framework consists of a Morphological analyser, Morphological generator, Word pleasantness and Word usage score finder as well as analysis of current usage in Social Networks, Picture dictionary, equivalent Tamil words, Generator (Word suggestions), Spell checker, Phonetic transliteration, Number to Text Converter, Rare-Word of the day and Social Network sharing.

Agaraadhi dictionary has more than 3 lac words in various domains such as General, Literature, Medical, Engineering, Computer Science, etc. The Agaraadhi framework dictionary is a Tamil English bilingual dictionary. The following sections describe the framework and list the benefits of such a framework over traditional online Tamil-English dictionaries. A few features proposed in this framework such as popularity score for a word, to best of our knowledge, are not present in any other world dictionaries.

2. Agaraadhi Framework
Agaraadhi Dictionary Framework was designed to provide additional information to the user regarding the word that they query about. Agaraadhi framework presented in figure 1 can be divided
into two major divisions, online and offline, in terms of the time of processing. This section describes the various components used in the Agaraadhi framework in detail.

2.1 Online Process

Any user query is sent sequentially to dictionary and literature, to retrieve corresponding data from those indices, fetching phonetic transliteration from transliteration modules, morphological information from morphological analysis and generator module and fetching live usage analysis from micro blogs. All those information are sent to user interface pages, shown in fig 1.

2.2 Offline Process

Tamil words and their meanings are entered manually and stored in text files. Those words are sequentially sent to modules such as popularity score generator, pleasantness score generator, picture dictionary, phonetic transliteration module and the resulting information is abstracted as a word object. Tamil literature such as Bharathiyaar songs, Avvaiyar songs, Thirukkural and lyrics are crawled using a static web crawler and are indexed in hash table as key value pairs.

2.3 Features of Agaraadhi Framework

Agaraadhi dictionary framework consists of more than twenty features such as Morphological Analysis Morph Generation, pleasantness scoring, popularity scoring, spell error suggestions etc.
3.1 Morphological Analyser

Morphological analyser [1] chunks the query word and gives the morphological features of the query word such as root word, parts of speech, gender, tense and count. If the Query word is padithaan, Morphological Analyser gives as padi as root, word represents male gender and query word is past tense and so on.

3.2 Morphological Generator

A Tamil morphological generator[2] needs to tackle different syntactic categories such as nouns, verbs, post positions, adjectives, adverbs etc. separately, since the addition of morphological constituents to each of these syntactic categories depends on different types of information. The generator is used to generate possible morphological variations of the query word.

3.3 Spell Checker

Spell Checker is used to check the spelling of Tamil words and to provide alternative suggestions for the wrong words. It uses the Morphological Analyzer. The Morphological Analyzer is used to split the given Tamil word into the root word and a set of suffixes. If the word is fully split by the analyzer and its root word is also found in the Agaraadhi dictionary, the given word is termed as correct. Otherwise, the correction process is invoked to generate all the possible suggestions with minimum variations from the given word.

3.4 Word Suggestions

Word Suggestion gives the list of equivalent or related words for the given query word.

3.5 Word Pleasantness and Word Popularity Score

Word Pleasantness score generator provides how easy to pronounce the word.

Word Popularity shows the word usage in the web. The Word from agaraadhi is given to web and found the frequency distribution of the word across the popular blogs, news articles, social nets etc.

3.6 Word Usage in Literature

This feature finds the usage of words in popular literature such as Thirukural, Bharathiyar Padalgal, Avvai songs and Lyrics.

3.7 Number to Text Converter

It converts a number to Tamil word equivalent as well as in English text. For example in Tamil we represent oru Arpputham (அஇ合肥ம்) for 100 million, Kumbam (அ்கூம்) for 10 billion and finally up to Anniyan (அ்நியம்) for one zillion.

3.8 Phonetic Transliteration

The pronunciation of words in Tamil and English language, as distinct from their written form based on the phonology and it can also vary greatly among dialects of a language. Phonetic transliteration module splits the word into syllables and gives the transliteration for each syllable.
3.9 Picture Dictionary

Pictures, photos or line drawings to depict popular words have been included in the dictionary to enable efficient learning for children using this tool.

3.10 Social net Sharing and Twitter Update

The framework also provides features to format results to be shared effectively on social networks. An Agaraadhi Bot was designed to post updates and word of the day on Twitter automatically.

3.11 Word of the Day and Word Usage statistics

A rare word is randomly chosen and is displayed in the opening page to facilitate users to learn a new word every day.

Word Usage Statistics [3] shows the usage of the word in the social network over the past one week.

3.12 Tamil Word Games

Games play a vital role in learning. Currently Agaraadhi has two Tamil word games namely Miruginajambo and Thookku Thookki. Miruginajambo is an unscramble game and Thookku Thookki is a Hangman game in Tamil.

4. Conclusion and Future Work

This paper describes Agaraadhi, an Online Dictionary Framework. Agaraadhi online dictionary is a bilingual dictionary containing over 3 lac words on various domains like General, Medical, Engineering, Computer science, Literature etc. This Online Dictionary framework encompass various features such as morphological analysis, morphological generation, word usage statistics, word pleasantness analysis, spell checking, similar word finder, word usage in literature, picture dictionary, number to text conversion, phonetic transliteration, live usage analysis from micro blogs etc. Providing APIs for programmers and developing mobile apps for Agaraadhi framework will open a good platform for many researchers and developers working in Tamil Computing area.

References

தனியுரிய அகராதி

பலவச க. துருகியங்கனம்,

முன்னநாள் புரோஜகாரம்,

சுருக அரசியல் பைன்யாைர்களுடன், விளையாட்டு படத்திற்குக் கட்டடம், தினங்கள் - 632 014., விளங்கு வேலஸ்டாம். கனிமஞ்சல், துருகம்.

மின்சலர்: ktamiliselvan@vit.ac.in

பலவாயபாலாம்:

பலவசத்திய நூற்றாண்டு புதுமை, புதுமை பெறுமின்றியும் முழுமை என்றுடையதை எந்தோன்றுள்ளது. புதுமை காணவோராக புதுமை என்று கூறுவதால் முதல்லோருடைய பத்திரக்கல். Cell Phone-ல் விளைவு நடத்தி, தொடர்பு நடத்துவதால் அது என்னும் பத்திரக்கல். Cell Phone, பெருமை என்று பங்கு மாறுகூர்ந்து புதுமை என்று கூறுவதால் அது என்னும் பத்திரக்கல். செய்தியக்குரிய தொடர்பு நடத்தும் பயன்பாட்டைத் தொடங்க என்று கூறுவதால் க உடன் என்னும் பத்திரக்கல்.

இந்தக் காலநிலை அரசு செய்தியாளர் புதுமையாக என்று கூறுவதால், இது மாற்றியலாம் அவர்களின் குட்டி செய்திகளை விளையாட்டு என்று கூறுவதால். இந்தக் காலநிலை அரசியல் புதுமை - கிளைநா கட்டடம் என்று கூறுவதால். செய்தியக்குரிய வட்டம் பட்டியல் என்று. இந்தக் காலநிலை செய்தியாளர் புதுமையாக என்று கூறுவதால். இந்தக் காலநிலை அரசு செய்தியாளர் புதுமையாக என்று கூறுவதால். இந்தக் காலநிலை அரசு செய்தியாளர் புதுமையாக என்று கூறுவதால்.

மின்சல்-1:

புதுமை காணவோராக என்று கூறுவதால். பின்னர் மாற்றியலாம் அவர்களின் குட்டி என்று பத்திரக்கல்களை விளையாட்டு என்று கூறுவதால். அவாறு அது என்று மாற்றியலாம் என்று கூறுவதால். இந்த மாற்றனால், மாற்றியலாம் அளவிற்கு வரையாகேண்பாற்றும்.
தமிழ் அகராதி

○ தமிழ் ○ English

மாற்றத்தக்க குற்று மூலக்கூறு

Go

பிரிவு: தமிழ் அர்வத்தினல் (அர்வத்தினல்) (புதியமுறை)
The first letter of Tamil alphabet

முடி - 2:
தமிழ் அகராதி முதற் சொல், "a" சொல் முதலும் தொடர்ந்து விளக்கம், முதலில் மீது ஐந்து ஆண்டு அவ்வாறு தொடர்ந்து விளக்க வும். இவ்விரண்டு உரையானது எண் மீது அல்லது பிரிவு மீது எண் வேதாதாரம், நம்பிக் காட்சியில் விளக்கம்.

தமிழ் அகராதி

○ தமிழ் ○ English

Enter your word

Go

a

abacus
abbey
நூற்றாண்டு அகாதமி

 தி காற்று எ எண்ணிய

Enter your word

தமிழ் அறிவியல் (அறிவியல்) புத்தகம்

முதலில்: The first letter of the English alphabet

உட்புறம்: ER Diagram for Tamil Dictionary

<table>
<thead>
<tr>
<th>மாறிகள் வரிசை</th>
<th>பொருளியல்</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 பதிமம்பாட்டு மறியுந்து எண்ணியகுதி (Prototype)</td>
<td></td>
</tr>
<tr>
<td>2 பதிமம்பாட்டு மறியுந்து எண்ணியகுதி (Prototype Approval)</td>
<td></td>
</tr>
<tr>
<td>3 பதிமம்பாட்டு மறியுந்து எண்ணியகுதி (Prototype Approval)</td>
<td></td>
</tr>
<tr>
<td>4 பதிமம்பாட்டு மறியுந்து எண்ணியகுதி (Prototype Approval)</td>
<td></td>
</tr>
<tr>
<td>5 பதிமம்பாட்டு மறியுந்து எண்ணியகுதி (Prototype Approval)</td>
<td></td>
</tr>
<tr>
<td>6 பதிமம்பாட்டு மறியுந்து எண்ணியகுதிபேசும் (QA Testing)</td>
<td></td>
</tr>
<tr>
<td>7 பதிமம்பாட்டு மறியுந்து எண்ணியகுதிபேசும் (QA Testing)</td>
<td></td>
</tr>
<tr>
<td>8 பதிமம்பாட்டு மறியுந்து எண்ணியகுதிபேசும் (QA Testing) Desktop Application</td>
<td></td>
</tr>
<tr>
<td>9 பதிமம்பாட்டு மறியுந்து எண்ணியகுதிபேசும் (QA Testing) Desktop Application</td>
<td></td>
</tr>
<tr>
<td>S.No.</td>
<td>Components</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
<td>Operating System</td>
</tr>
<tr>
<td>2</td>
<td>Application Server</td>
</tr>
<tr>
<td>3</td>
<td>Web Tools</td>
</tr>
<tr>
<td>4</td>
<td>Database</td>
</tr>
<tr>
<td>5</td>
<td>Browser</td>
</tr>
</tbody>
</table>


MS SQL 2005 Server வாழை பரிசாய்படுத்துவது எண் கொள்ளல் கொண்டுவந்து.

இதில் எண் கொள்ளல் உள்ளது ஆறு மித்தமளிக்கும் வாழை வண்ணம் கொள்ளல் கொண்டுவந்து. 100% Bug Free வாழை பரிசாய்படுத்துவது எண் கொள்ளல் கொண்டுவந்து.
CONFERENCE PAPERS

Tamil Internet 2011
மாறுமிப்புக்கு கணங்கு அநுக்கினி முன்னர்

இந்துவர கி.வி. திருக்கோவில், வைணவார
E-mail: s.ilangkumaran@gmail.com

மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் பல்கலைக் கல்லூரியில் அற்குரையிடி. நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறுமிப்புப்படும் பல்கலைக் கல்லூரியில் அற்குரையிடின் நூற்றோட்டம் நூற்றுக்கணக்கான பக்தியுடன் மாறு

1. வழிபாட்டில் ஆற்றுக் கையில் மேற்குப் பத்திரிகையுள்ள எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினாலும் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினாலும் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினாலும் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினாலும் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசை�ிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்று�்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசை�ிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசை�ில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசை�ிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குற்றுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குறுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குறுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குறுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுудைய குறுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும் அடிகுடைய குறுக்கையுள்ள சாப்பிட்டினால் வரிசையில் வரிசையில் எந்த வரிசையிலும்
2. அகராதிகளில் தென்பெற்று விளைந்த மூட்டலை நாள்புழக்கான

- நம்பிக்கையில் இராணுவத் தன்னி கையற்ற பல்லுரைகளின் மூலம் அனுப்பியது. “இராணுவத் தன்னிகளின் கடவை நகரம் முழு விளைந்த நூற்றுப்புறங்களும், கேரளத்தில், அமெரிக்காவில் கேரளத் தனித்துவமாக புனிதவாக இருந்தது அரியணியாளரின் அரவணாளரின் காலத்தில் கேரளத்தின் தொடர்புகளை விளக்கப்பட்டார். அவர்கள் தன்னிகளின் கேரளத் தன்னிகளைப் பின்னர் கேரளத்தின் தொடர்புகளை விளக்கப்பட்டார். இந்துரையான கேரளப் புராணத்தின் மற்றொரு பக்தர் பத்திரத்தை கேரளத் தன்னிகளை விளக்கப்பட்டார். இதுவரை அகராதிகள் பராமரித்தது மூன்று பாலம் விளக்கப்பட்டார். இந்துரையான கேரளப் புராணத்தின் மற்றொரு பக்தர் பத்திரத்தை விளக்கப்பட்டார். (இந்துரையான கேரளப் புராணத்தின் மற்றொரு பக்தர் பத்திரத்தை விளக்கப்பட்டார்)

2. அகராதிகளில் தென்பெற்று விளைந்த மூட்டலை நாள்புழக்கான

- ஒவ்வொரு பாலையும் இராணுவத் தன்னிகளில் அனுப்பியது கேரளத் தன்னிகளில் முன்னிலையான மூட்டலை நாள்புழக்கான. அனுப்பியது கேரளத் தன்னிகளில் முன்னிலையான மூட்டலை நாள்புழக்கான. இந்துரையான கேரளப் புராணத்தின் மற்றொரு பக்தர் பத்திரத்தை விளக்கப்பட்டார். இதுவரை அகராதிகள் பராமரித்தது மூன்று பாலம் விளக்கப்பட்டார். இந்துரையான கேரளப் புராணத்தின் மற்றொரு பக்தர் பத்திரத்தை விளக்கப்பட்டார். (இந்துரையான கேரளப் புராணத்தின் மற்றொரு பக்தர் பத்திரத்தை விளக்கப்பட்டார்)

3. அவர்கள் என்று இராணுவத் தன்னிகளின் கையற்ற

- இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள். இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள். இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள் இந்த ரோதரிகளின் கையற்ற பல்லுரைகள்.
4. என்று அடுத்து போகிறது போகிறது போகிறது போகிறது

5. பற்றியன்று அடுத்திருக்கிறது போகிறது போகிறது போகிறது
க, திமுறுமின்றி பயணம் தைனியாள் மகமரிலுள்ள மீது விளையாடக் கூறுகின்றீர்க். பாதாக்கத்தின் அகராதிகளின் மீது, அதிகமான எங்கும் வருவிடும் மகமான கூறுகின்றது. குற்றால் திமுறுமின்றில் விளையாடப் பயணம் மிகுதியாகும். ஆனால், மகமான நூற்றுக்கணக் கூறுகின்றது. பாதாக்கத்தின் முன்னோடியில், மகமான கூறுகின்றது. குற்றால் திமுறுமின்றில் விளையாடப் பயணம் மிகுதியாகும். மகமான கூறுகின்றது. மகமான கூறுகின்றது. பாதாக்கத்தின் முன்னோடியில், மகமான கூறுகின்றது. குற்றால் திமுறுமின்றில் விளையாடப் பயணம் மிகுதியாகும். மகமான கூறுகின்றது.
CONFERENCE PAPERS

TAMIL INTERNET 2011
Blog
(வைல்பம் புத்தகம்)
Impact of SOA and Web 2.0 in Tamil Blogs and Social Networks

Ferdin Joe J
PG Scholar/CSE, Einstein College of Engineering, Tirunelveli, India

Abstract
Social Networks and Blogs are the most powerful electronic media which reaches every nook and corner of the planet. They are being viewed and managed by even devices using the embedded technology. Service Oriented Architecture (SOA) and Web 2.0 are the technologies which are the base for social networks and blogs. Tamil blogs are also gaining popularity among readers with ample contribution from Unicode format of data management. Many Tamils all over the globe; especially the Non Residential Indians (NRIs) have a powerful base of reflecting their thoughts on the social issues of India using the Tamil Blogs. Social Networks like Twitter, Facebook etc., are also supporting the Tamil Unicode nowadays and the Tamil Internet User Community is benefited in many ways. The use of SOA and Web 2.0 plays a vital role in Tamil Unicode with Social Networks and Tamil Blogs. In this paper, various Social Networks and Tamil blogs from all over the globe are evaluated. This evaluation study is done to find the impact of SOA and Web 2.0 on the web. The views, target audience and significant features of various Tamil bloggers and Social Network users are also taken into account for this evaluation.

Index Terms: Social Networks, Tamil Blogs

Introduction
Tamil is powered by Unicode in the web, especially in the social networking is in the increase. Policies were framed after the Tamil Internet Conference 2010 by the Government of Tamil Nadu, India. The usage of Tamil Unicode was approved officially for any communication with the Government offices. With the collaboration of Microsoft, Tamil Unicode was given a sort of recognition at the level of silverlight frameworks. Due to these policies, the use of Tamil Unicode has increased proportionally with the increase in the number of internet users. Lot many Tamil Language Researchers who were new to internet usage, utilized this technology as a boon for their research. Encouragement for the usage was enormous by the release of free and open source softwares like NHM writer released in the CD of CDAC, Govt of India, Azhagi word processor and many more. Nowadays Tamil Unicode is used to comment in various social networking sites like orkut, facebook, twitter etc and chat applications of gmail, yahoo etc. Tamil journalists and magazine column writers found this technology easy and quicker way to reach the editorial board. These areas are powered by the service oriented architecture (SOA) and web 2.0 using blogs and social networking sites. These sites reach the people with the Really Simple Syndication (RSS) and Atom feeds in a more effective way. Tamil Unicode is used by the users to express the culture, activities and emotions of Tamils around the globe.
In this paper, I have analyzed many social networks and Tamil blogs. The use of SOA and Web 2.0 is taken as the usage of these sites. From these websites, the impact of Tamil Users are analyzed for the past one year and the major issues of the Tamil Unicode users are listed out as conclusion of the study.

**Study sites & Criteria**

Various sites like savukku.net, athirvu and many bloggers in blogspot and wordpress are taken into account. They were monitored throughout the year and the issues were listed. This listing was done until the release of wikileaks cables and the UNO report on Srilankan war. Most of the bloggers focused on the Srilankan Tamils issue by slamming the Government of India. Then the 2G spectrum scam which is currently blaming the ruling parties of India and the State Government of Tamil Nadu. Apart from these, the issues were raised on the byeelections and the small incidents happened in and across Tamil Nadu.

**Aim of writers**

The main aim of these writers was to change the minds of the people and they have succeeded partially by getting positive comments from exclusive and anonymous users. They use to write this stuff in Tamil Unicode and therefore the reader base has increased but the contents were not reliable. The quality of writing included many third rated comments on National leaders.

**Impact on readers**

The impact of these articles made both positive and negative impact on the reader base. Though they got an increased reader base, they failed to give quality articles. The issues raised on the Srilankan war became a success with the advent of the UNO report on war crime. But in the case of 2G spectrum scam, it is not. In the case of 2G spectrum scam, no one has the rights to comment on either the Government or the politicians. The scam is still under the trial in the special court of Central Bureau of Investigation, India. Based on some filthy youtube released tapes and opposition party demonstrations, the real conclusion could not be made. But the Tamil content writers made their attempt in this issue with ghost referred knowledge. The Parliamentary Accounts Committee (PAC) is headed by Mr. Murali Manohar Joshi from the opposition party in the Parliament of India. The initial audit report gave a doubt of loss of money due to the spectrum auction. But the writers were just slamming on the Government for the illusion created. The Joint Parliament Committee (JPC) is formed as per the demand of the opposition party and the trial is going on. Recently, “The Hindu” has given a news stating that, the effect of JPC is becoming a boon to the ruling party itself and the opposition parties want the committee to retire as they couldn’t achieve what they felt too. On April 28th, 2011, the Parliament Accounts Committee’s report was rejected whole heartedly rejected by the committee and a new Head was appointed. These activities show that the scam is not the offence of only ruling party but there are some others who are behind. In this scenario, the content of Tamil writers in the web came to be known as just illusions and confused the reader base. Similarly, the Government of TamilNadu was criticized for freebies in the 2011 Assembly Elections Poll campaign. Elections are normally the reflection of the pulse of voters. Without knowing the this pulse, the content was published just by slamming the welfare schemes done. We need to wait till May 13th to see whether the voters have rejected the freebies or they welcomed with red carpet. Without this made clear, the media will never have rights to comment on the Government to change the mindset of people.
Conclusion

In this scenario, as a neutral reader, the issues stated in the social networks and blogs need to be analyzed thoroughly by the readers and then the issues should be taken into account. It is really an offense to feed the people with wrong news. Tamil internet users need to take all these concepts in mind before they read any sort of sensitive news in the media. The links posted in the social networks need to be audited by any neutral organization upon appeal. The negative certificate on the content will make the users to understand the difference between reality and illusions. Currently the Tamil media in the form of blogs, social networks, TV channels are not reliable because of the backing of single person opinion being stuffed on reader base. The auditing on the reality of content and the certification of content will definitely lead the readers to get the correct picture on the issue.

References

Nothing in specific. Most of the eminent social networks and blogs were taken for this study.
Tamil classical literature popularly known as the Sangam literature is hailed by scholars around the world as one best literary output of human civilizations. Scholars from various western countries have patronized Sangam literature and produced various studies and translation on this literature. Many Universities around the world have understood the literary significance of Sangam literature and have opened departments to study Sangam literature. With such literary greatness, Sangam literature is not much known to common people of Tamil Nadu. The common people of Tamil Nadu are more exposed to literature such as Thirukkural and Bakthi than Sangam literature.

Except for very few Puram poems much of Sangam literature is unknown to common people of Tamil Nadu. The only exposure most people have towards Sangam literature is the few poems which they had read in their school Tamil syllabus. There has been no popular appeal or drive to encourage people to read Sangam literature in the likes of Thirukkural.

Themes of Love and war essentially make most of the poems, and not well appreciated by children of young ages. Hence the only time people are exposed to this literature they had not developed any special interest for it. In such circumstances, the age of blogging and social network which are the mostly used by the youngsters in the prime of their life can be effectively used to popularize the Sangam literature. Sangam literature has strong emotional connection with present day lives. Hence if Sangam poems are presented in proper way, it is sure strike a chord with present generation.

Obstacles to a common man:

The biggest obstacle for a common man is the archaic language of Sangam Literature. To understand and appreciate the Sangam poems, a person has to know lot of background details like themes, landscape etc. Dr.Kamil Zvelebil has once remarked that only trained readers can completely understand and appreciate Sangam literature. A common man requires considerable education before he can appreciate Sangam poems. More often than not, translating the poem or explaining the poem line by line doesn’t give full experience of the poem. As an example I would like to take a poem and explain how much back ground detail he requires to appreciate the poem

\[
\text{மாாி ஆΆப᾿ அᾹன காᾰகிᾹ,}
\text{பா᾽வ᾿ அﺎசிய பᾞவர᾿ ஈ╯ ஆஞᾶᾌ,}
\text{கᾶட╯ அைளᾲ ஆசᾣஇய╯,}
\text{அᾶட╯}
\text{கயிா அாி எᾞ合肥்,}
\text{கதᾨஆ ஐைறவᾹ}
\text{வாராᾐ அைமயிᾔᾌ அைமக}
\text{சிறியᾫᾌ உள ஈᾶᾌ,}
\text{விஈலஞ╯ கவைளேய.}
\]
If the chief of maritime land

Where,
the wet crab
seeing the stork,

Which looked
Like a water lily in rain,

Gets afraid and
moves inside the hollow roots of
the mangrove trees,

Like a bull which cut loose
from the rope tied by the herdsman,

Doesn’t return
let it be so.

The merchant sells smaller bangles
in this town.

In the following poem we expect the reader of this translation to know the following points to enjoy the poem completely

1. How the water lily looks when it rains - and how it is so similar to the crane
2. Wet crab being compared to Thalaivan
3. Crane compared to gossip of women
4. Running for safety into hollow roots, as imagery for Thalaivan hurrying to Thalaivi’s home to save his love.
5. Bull cutting loses from the rope tied by the herdsman as imagery for Thalaivan breaking the shackles of the pressure from the society.
6. Small bangles to indicate the Tamil akam tradition of love sickness and thinning of hands and loosening of bangles. Small bangles will fit the smaller hand and hence they can throw away the memory of Thalaivan and be safe from the society.

It can be shown with further examples that traditional methods of translation and brief explanation cannot give a reader the full experience of a sangam poem. Hence experimental approaches should be considered to take this classical literature to the common reader.

**Blogging as a medium:**

Blogging as medium of communication has developed over the last few years. This medium gives enough flexibility and freedom and in most cases free of cost. Blogging has become a popular medium for readers who like to gather knowledge on various topics. Most internet users are used to blog medium hence blogging overcomes one of the major pitfalls of Book. Books involve printing cost
hence there are severe restrictions on page counts and illustrations. Blogs therefore can afford to be as long as they want and can use illustrations as required. The blogging medium is also flexible enough to allow audio and video. The flexibility of this medium can be really helpful in presenting Sangam literature to the new age reader.

With no restrictions in page count, more explanation can be given on the theme, landscape and other required back stories can be given in addition to translation to understand a particular poem.

Sangam literature has very visual nature to it. The poets of the Sangam infused elements of nature-flora and fauna effectively in the poems to describe emotions. They play an integral part in understanding the poem. Most readers’ knowledge of flora and fauna especially with names in archaic Tamil is pretty low. Hence pictures or illustration of flora and fauna mentioned in the poems helps the reader understand the poems better.

**Experimental approaches:**

The visual nature of the Sangam nature can be utilized and the poems can be presented in new formats such as explaining poems with series of pictures or more experimental approach of comic book format (with illustrations and subtext). Audio files with rendering of the poem along with explanations can be uploaded. Lectures on the poem can be recorded and uploaded as video blogs. Small animations can be made to explain similes and imagery of the poem.

Few examples of illustrating visual elements of Sangam poems are given below
A more experimental approach of Sangam poem in comic book style

With blogging medium we have no restrictions to creativity. Experimental and creative approaches can help in popularizing the Sangam poems.

**Visual catalogs:**

Visual catalogs of various flora fauna and smiles can be made into searchable online databases. In one such effort visual catalog of 99 flowers of Kurinchippattu was created 2 years ago in my blog karkanirka.org. This effort can be extended to include all flora and fauna of Sangam and also efforts can be put convert into a searchable online database. Such visual catalogs would help the common reader as well as researchers in understanding the Sangam literature in much better way.
Few slides from the visual presentation of 99 Flowers of Kurichippaatu.

**Relating to popular Culture:**

Most themes of Sangam poems are still prevalent in popular culture though Tamil Cinema, either as scenes or as lyrics for film songs. Relating the Sangam poems to prevalent popular culture helps common reader understand the classical poems better since they are able to relate to the emotion of the poem.

Themes such as love at first sight, opposition for love from parents, eloping of lovers, temporary separation of lovers, wife cooking for first time and waiting for approval of husband, girls parents mistaking girls love sickness and inviting priest to perform rituals are still prevalent as movie themes.

Relating the Sangam themes to popular culture makes them understand that Sangam literature isn’t as alien as they had imagined and creates an interest in them to explore more in the classical literature.

**Use of Social Network:**

Social networks like Orkut, Twitter and Facebook have become the most popular medium for knowledge transfer and sharing. Each of these social networks has hundreds of groups on Tamil Language and literature. Sharing the links of the blogs at such groups helps spread the reach of Sangam literature. This method popularly known as viral marketing is very useful tool. Many new readers are exposed to Sangam literature this way. When a reader likes a poem he shares it to his friends and any one is his friends list who likes the post passes on to his friends. This way with minimum effort the reach of the poems can be maximized. The social networks have played an important role in growth of Karka Nirka blog. In April 2010 Karka Nirka Facebook page was started and presently there are 622 followers and growing. On average one blog posted in Face book is passed on/shared by 50 other people.
**Conclusion:**

New approaches are required to popularize the Sangam literature. There are many takers for such efforts. Fresh enthusiasm and creativity can help to revitalize the Sangam Literature and take it to the present generation.
Impact of SOA and Web 2.0 in Tamil Blogs and Social Networks

Ferdin Joe J
PG Scholar/CSE, Einstein College of Engineering, Tirunelveli, India

Abstract

Social Networks and Blogs are the most powerful electronic media which reaches every nook and corner of the planet. They are being viewed and managed by even devices using the embedded technology. Service Oriented Architecture (SOA) and Web 2.0 are the technologies which are the base for social networks and blogs. Tamil blogs are also gaining popularity among readers with ample contribution from Unicode format of data management. Many Tamils all over the globe; especially the Non Residential Indians (NRIs) have a powerful base of reflecting their thoughts on the social issues of India using the Tamil Blogs. Social Networks like Twitter, Facebook etc., are also supporting the Tamil Unicode nowadays and the Tamil Internet User Community is benefited in many ways. The use of SOA and Web 2.0 plays a vital role in Tamil Unicode with Social Networks and Tamil Blogs. In this paper, various Social Networks and Tamil blogs from all over the globe are evaluated. This evaluation study is done to find the impact of SOA and Web 2.0 on the web. The views, target audience and significant features of various Tamil bloggers and Social Network users are also taken into account for this evaluation.

Index Terms: Social Networks, Tamil Blogs

Introduction

Tamil is powered by Unicode in the web, especially in the social networking is in the increase. Policies were framed after the Tamil Internet Conference 2010 by the Government of Tamil Nadu, India. The usage of Tamil Unicode was approved officially for any communication with the Government offices. With the collaboration of Microsoft, Tamil Unicode was given a sort of recognition at the level of silverlight frameworks. Due to these policies, the use of Tamil Unicode has increased proportionally with the increase in the number of internet users. Lot many Tamil Language Researchers who were new to internet usage, utilized this technology as a boon for their research. Encouragement for the usage was enormous by the release of free and open source softwares like NHM writer released in the CD of CDAC, Govt of India, Azhagi word processor and many more. Nowadays Tamil Unicode is used to comment in various social networking sites like orkut, facebook, twitter etc and chat applications of gmail, yahoo etc. Tamil journalists and magazine column writers found this technology easy and quicker way to reach the editorial board. These areas are powered by the service oriented architecture (SOA) and web 2.0 using blogs and social networking sites. These sites reach the people with the Really Simple Syndication (RSS) and Atom feeds in a more effective way. Tamil Unicode is used by the users to express the culture, activities and emotions of Tamils around the globe.
In this paper, I have analyzed many social networks and Tamil blogs. The use of SOA and Web 2.0 is taken as the usage of these sites. From these websites, the impact of Tamil Users are analyzed for the past one year and the major issues of the Tamil Unicode users are listed out as conclusion of the study.

**Study sites & Criteria**

Various sites like savukku.net, athirvu and many bloggers in blogspot and wordpress are taken into account. They were monitored throughout the year and the issues were listed. This listing was done until the release of wikileaks cables and the UNO report on Srilankan war. Most of the bloggers focused on the Srilankan Tamils issue by slamming the Governement of India. Then the 2G spectrum scam which is currently blaming the ruling parties of India and the State Government of Tamil Nadu. Apart from these, the issues were raised on the byelections and the small incidents happened in and across Tamil Nadu.

**Aim of writers**

The main aim of these writers was to change the minds of the people and they have succeeded partially by getting positive comments from exclusive and anonymous users. They use to write this stuff in Tamil Unicode and therefore the reader base has increased but the contents were not reliable. The quality of writing included many third rated comments on National leaders.

**Impact on readers**

The impact of these articles made both positive and negative impact on the reader base. Though they got an increased reader base, they failed to give quality articles. The issues raised on the Srilankan war became a success with the advent of the UNO report on war crime. But in the case of 2G spectrum scam, it is not. In the case of 2G spectrum scam, no one has the rights to comment on either the Government or the politicians. The scam is still under the trial in the special court of Central Bureau of Investigation, India. Based on some filthy youtube released tapes and opposition party demonstrations, the real conclusion could not be made. But the Tamil content writers made their attempt in this issue with ghost referred knowledge. The Parliamentary Accounts Committee (PAC) is headed by Mr. Murali Manohar Joshi from the opposition party in the Parliament of India. The initial audit report gave a doubt of loss of money due to the spectrum auction. But the writers were just slamming on the Government for the illusion created. The Joint Parliament Committee (JPC) is formed as per the demand of the opposition party and the trial is going on. Recently, “The Hindu” has given a news stating that, the effect of JPC is becoming a boon to the ruling party itself and the opposition parties want the committee to retire as they couldn’t achieve what they felt too. On April 28th, 2011, the Parliament Accounts Committee’s report was rejected whole heartedly rejected by the committee and a new Head was appointed. These activities show that the scam is not the offence of only ruling party but there are some others who are behind. In this scenario, the content of Tamil writers in the web came to be known as just illusions and confused the reader base. Similarly, the Government of TamilNadu was criticized for freebies in the 2011 Assembly Elections Poll campaign. Elections are normally the reflection of the pulse of voters. Without knowing the this pulse, the content was published just by slamming the welfare schemes done. We need to wait till May 13th to see whether the voters have rejected the freebies or they welcomed with red carpet. Without this made clear, the media will never have rights to comment on the Government to change the mindset of people.
Conclusion

In this scenario, as a neutral reader, the issues stated in the social networks and blogs need to be analyzed thoroughly by the readers and then the issues should be taken into account. It is really an offense to feed the people with wrong news. Tamil internet users need to take all these concepts in mind before they read any sort of sensitive news in the media. The links posted in the social networks need to be audited by any neutral organization upon appeal. The negative certificate on the content will make the users to understand the difference between reality and illusions. Currently the Tamil media in the form of blogs, social networks, TV channels are not reliable because of the backing of single person opinion being stuffed on reader base. The auditing on the reality of content and the certification of content will definitely lead the readers to get the correct picture on the issue.

References

Nothing in specific. Most of the eminent social networks and blogs were taken for this study.
Tamil Classical Literature in the age of blogging and social network

Palaniappan Vairam Sarathy
http://karkanirka.org

Tamil classical literature popularly known as the Sangam literature is hailed by scholars around the world as one best literary output of human civilizations. Scholars from various western countries have patronized Sangam literature and produced various studies and translation on this literature. Many Universities around the world have understood the literary significance of Sangam literature and have opened departments to study Sangam literature. With such literary greatness, Sangam literature is not much known to common people of Tamil Nadu. The common people of Tamil Nadu are more exposed to literature such as Thirukkural and Bakthi than Sangam literature.

Except for very few Puram poems much of Sangam literature is unknown to common people of Tamil Nadu. The only exposure most people have towards Sangam literature is the few poems which they had read in their school Tamil syllabus. There has been no popular appeal or drive to encourage people to read Sangam literature in the likes of Thirukkural.

Themes of Love and war essentially make most of the poems, and not well appreciated by children of young ages. Hence the only time people are exposed to this literature they had not developed any special interest for it. In such circumstances, the age of blogging and social network which are the mostly used by the youngsters in the prime of their life can be effectively used to popularize the Sangam literature. Sangam literature has strong emotional connection with present day lives. Hence if Sangam poems are presented in proper way, it is sure strike a chord with present generation.

Obstacles to a common man:

The biggest obstacle for a common man is the archaic language of Sangam Literature. To understand and appreciate the Sangam poems, a person has to know lot of background details like themes, landscape etc. Dr.Kamil Zvelebil has once remarked that only trained readers can completely understand and appreciate Sangam literature. A common man requires considerable education before he can appreciate Sangam poems. More often than not, translating the poem or explaining the poem line by line doesn’t give full experience of the poem. As an example I would like to take a poem and explain how much back ground detail he requires to appreciate the poem

மாாி ஆΆப᾿ அᾹன ெகாᾰகிᾹ
பா᾽வ᾿ அ驷ாசிய பᾞவர᾿ ஈBoxLayout=134],134]’
கᾶட᾿ அைளᾲ அைசᾣஇய᾽
கயிᾠ அாி எᾞ近百ிᾹ,
கதᾨΆ ᾐைறவᾹ
வாராᾐ அைமயி.Android, அைமக
சிறியᾫΆ உள ஈᾶᾌ
,விைலஞ᾽ ஈகவைளேய.
If the chief of maritime land

Where,
the wet crab
seeing the stork,
Which looked
Like a water lily in rain,
Gets afraid and
moves inside the hollow roots of
the mangrove trees,
Like a bull which cut loose
from the rope tied by the herdsman,

 Doesn’t return
let it be so.

The merchant sells smaller bangles
in this town.

In the following poem we expect the reader of this translation to know the following points to enjoy the poem completely

1. How the water lily looks when it rains - and how it is so similar to the crane
2. Wet crab being compared to Thalaivan
3. Crane compared to gossip of women
4. Running for safety into hollow roots, as imagery for Thalaivan hurrying to Thalaivi’s home to save his love.
5. Bull cutting loses from the rope tied by the herdsman as imagery for Thalaivan breaking the shackles of the pressure from the society.
6. Small bangles to indicate the Tamil akam tradition of love sickness and thinning of hands and loosening of bangles. Small bangles will fit the smaller hand and hence they can throw away the memory of Thalaivan and be safe from the society.

It can be shown with further examples that traditional methods of translation and brief explanation cannot give a reader the full experience of a sangam poem. Hence experimental approaches should be considered to take this classical literature to the common reader.

**Blogging as a medium:**

Blogging as medium of communication has developed over the last few years. This medium gives enough flexibility and freedom and in most cases free of cost. Blogging has become a popular medium for readers who like to gather knowledge on various topics. Most internet users are used to blog medium hence blogging overcomes one of the major pitfalls of Book. Books involve printing cost
hence there are severe restrictions on page counts and illustrations. Blogs therefore can afford to be as long as they want and can use illustrations as required. The blogging medium is also flexible enough to allow audio and video. The flexibility of this medium can be really helpful in presenting Sangam literature to the new age reader.

With no restrictions in page count, more explanation can be given on the theme, landscape and other required back stories can be given in addition to translation to understand a particular poem.

Sangam literature has very visual nature to it. The poets of the Sangam infused elements of nature-flora and fauna effectively in the poems to describe emotions. They play an integral part in understanding the poem. Most readers’ knowledge of flora and fauna especially with names in archaic Tamil is pretty low. Hence pictures or illustration of flora and fauna mentioned in the poems helps the reader understand the poems better.

**Experimental approaches:**

The visual nature of the Sangam nature can be utilized and the poems can be presented in new formats such as explaining poems with series of pictures or more experimental approach of comic book format (with illustrations and subtext). Audio files with rendering of the poem along with explanations can be uploaded. Lectures on the poem can be recorded and uploaded as video blogs. Small animations can be made to explain similes and imagery of the poem.

Few examples of illustrating visual elements of Sangam poems are given below
A more experimental approach of Sangam poem in comic book style

With blogging medium we have no restrictions to creativity. Experimental and creative approaches can help in popularizing the Sangam poems.

**Visual catalogs:**

Visual catalogs of various flora fauna and smiles can be made into searchable online databases. In one such effort visual catalog of 99 flowers of Kurinchippattu was created 2 years ago in my blog karkanirka.org. This effort can be extended to include all flora and fauna of Sangam and also efforts can be put convert into a searchable online database. Such visual catalogs would help the common reader as well as researchers in understanding the Sangam literature in much better way.
Few slides from the visual presentation of 99 Flowers of Kurichippaatu.

**Relating to popular Culture:**

Most themes of Sangam poems are still prevalent in popular culture though Tamil Cinema, either as scenes or as lyrics for film songs. Relating the Sangam poems to prevalent popular culture helps common reader understand the classical poems better since they are able to relate to the emotion of the poem.

Themes such as love at first sight, opposition for love from parents, eloping of lovers, temporary separation of lovers, wife cooking for first time and waiting for approval of husband, girls parents mistaking girls love sickness and inviting priest to perform rituals are still prevalent as movie themes.

Relating the Sangam themes to popular culture makes them understand that Sangam literature isn’t as alien as they had imagined and creates an interest in them to explore more in the classical literature.

**Use of Social Network:**

Social networks like Orkut, Twitter and Facebook have become the most popular medium for knowledge transfer and sharing. Each of these social networks has hundreds of groups on Tamil Language and literature. Sharing the links of the blogs at such groups helps spread the reach of Sangam literature. This method popularly known as viral marketing is very useful tool. Many new readers are exposed to Sangam literature this way. When a reader likes a poem he shares it to his friends and any one is his friends list who likes the post passes on to his friends. This way with minimum effort the reach of the poems can be maximized. The social networks have played an important role in growth of Karka Nirka blog. In April 2010 Karka Nirka Facebook page was started and presently there are 622 followers and growing. On average one blog posted in Face book is passed on/shared by 50 other people.
Conclusion:

New approaches are required to popularize the Sangam literature. There are many takers for such efforts. Fresh enthusiasm and creativity can help to revitalize the Sangam Literature and take it to the present generation.
Introduction

‘Tamil,’ as A.K. Ramanujan\[i\] says, ‘one of the two classical languages of India, is the only language of contemporary India which is recognizably continuous with a classical past’. Kamil V. Zvelebil\[ii\] says, ‘probably the most significant contribution of Tamil literature, which still remains to be ‘discovered’ and enjoyed by the non-Tamilians and adopted as an essential and remarkable part of universal heritage’. According to Harold Schiffman\[iii\] ‘most Tamils feel that their language and their linguistic culture really are different from most others in India’ Though information about Tamil literature in works by those referred to above is presented effectively, detailed information about Tamil literary works especially Sangam literature presented in one of the most common online reference sources, namely Wikipedia, is far from satisfactory.

The present paper aims at exploring problems connected with presentation of information about Classical Tamil Literary works in the online encyclopedias namely, Tamil and English wikis. These two online reference works being made free, people in different parts of the world have an easy access to them. But unfortunately some of the accounts about Tamil literary works found in them are often found to be highly skeletal, fragmentary, lacking in citations, disappointing the users remaining unverifiable and being incoherent. Hence the paper makes an attempt at analyzing the problems relating to presentation of information through online resources in them.

Encyclopedias

There are the highly traditional encyclopedias comprising entries developed as per norms of an encyclopedic entry in standard encyclopedias such as Encyclopedia Americana and Encyclopedia Britannica which contain useful information on Tamil literature. However, they are not easily available to those who want to use them for updating their knowledge about Tamil literature in general, specific Tamil literary works in particular. Even though the encyclopedia Britannica is available as an online encyclopedia, the stipulation seeking the users’ credit card number even for a trail run for a few days serves as a factor inhibiting its use. In this context, the Wikipedia encyclopedias (Tamil and English Wikis), freely available for anyone who has an access to a computer with an internet connection, in any part of the world, come in handy or easy to reach. But the information about Tamil literature found in them is far from satisfactory as they are highly fragmentary or skeletal. Against some of the Wikipedia encyclopedic entries suffering from presentation problems carry instructions such as the following:

‘This article needs additional citations for verifications. Please help improve this article by adding reliable references. Un-sourced material may be challenged and removed’ or ‘This Tamil related article
Given below is a specimen entry from English Wikipedia on Nānmanikkatikai one of the Eighteen Minor works of the Post-can;kam period.

Nānmanikkatikai

From Wikipedia, the free encyclopedia

Nanmanikkatigai is a Tamil poetic work of didactic nature belonging to the Pathinenkilkanakku anthology of Tamil literature. This belongs to the 'post Sangam period corresponding to between 100 – 500 CE. Nanmanikkatigai contains one hundred songs written by the poet Vilambu Naganaar. This poetic work is famous for its clarity and easy readability and is often a prescribed text for schools in Tamil Nadu. The poems of Nanmanikkatigai are written in the Venpa meter.

The poems of Nanmanikkatigai each contain four different ideas. The name Nanmanikkatigai denotes this fact comparing the four ideas to four well-chosen gems adorning each poem. The following poem describes four different groups of people who cannot sleep well at night, namely, a thief, a lovelorn person, someone who hankers after money and a miser who worries about losing his money:

[edit] References

- http://www.tamilnation.org/literature/
- http://www.tamilnation.org/literature/pathinen/pm0047.pdf Nanmanikkatigai eText at Project madurai

What is given below can be treated as an expanded encyclopedic entry version which is in no way complete[iv]. An entry relating to this post-cankam work can be attempted using various criteria insisted upon for a comprehensive encyclopedic entry.

Nānmanikkatikai

Nānmanikkatikai is one of the Eighteen Minor Works known in Tamil as Patinenkil'kanaikkunulikal. It comprises one hundred and six quatrains, the first two quatrains being invocation verses. The venpā in its variant forms is the metre of Nānmanikkatikai. However, the first verse and the other two namely the 30th and the 61st verses have five lines in each of them. As the first two quatrains are
in praise of lord Vishnu, people describe it as a Vaishnavaite work.

Those who view it as the post-sangam work would place it at the later part of Buddhists and Jains, AD 100 to 600 AD. S. Vaiyapuri Pillai in his History of Tamil Language and Literature (From the Beginning to 1000 AD) assigns 750 AD as the period of Composition of Nānmanikkatikai but many scholars do not subscribe to this view. As certain lines from Kuruntokai, one of the works of Eight Anthologies, has been used in the Nānmanikkatikai, it is possible to surmise that the Nānmanikkatikai belongs to a later or post-sangam period. The lines that have a striking similarity in the two works referred to above are the following:

Kuruntokai – 397
Nānmanikkatikai 23

As the coincidence between certain lines of Nānmanikkatikai and the Tirukkural is striking, commentators believe that the former has certainly borrowed from the latter giving credence to the view that Nānmanikkatikai belongs to an age later than that of Tirukkural. The quotes cited below will illustrate the point made above.

The Tirukkural - 1041

Nānmanikkatikai

One of the verses of the verses of Nānmanikkatikai and its translation version are given below:

No sleep for those who are surreptitious; no sleep for those who have set their mind on their favorite women
No sleep for those who are keen on wealth creation
And those who safeguard such wealth sleep not.

References

The translation of the Nānmanikkatikai verse into English quoted above is the one attempted by the presenter of this paper.
"Nanmanikkatikai" ப க ைத இ த வி கியி

உ வா க

As there is no entry relating to Nanmanikkatikai in the Tamil Wiki, an entry similar to the one given
below can be made:
நா

மணி க ைக பதிென

ச க

ம விய கால

ஆன . ஒ

லா

சில ெவ

ைசவசி தா த
ெவளியி

இ

இ

ைவ

லாசிாிய

விள பி எ

ப

என ெகா வா
றி

ஐ
. நா

மணிக

லாசிாிய ைஜனேர எ

ஆனைவ. இ

ெவ

எ

ற

104 ெவ

இ

பதி பி

104 ெச

க

ம

ற ப கிறா . இ

ைஜன

சமய க

பா களா
ம

ஆன .

ேம உ ளன

வா

தி மாைல
லாசிாிய

ப

பா க

கட

, தமிழகதி

உ ள நாகனா எ

பா களா

ெச

ேம உ ளன. இ

பாட க

றி பதாக

ெவ

உ ளன( 1904). கி. ஆ. ெப. வி வநாத

எ

ேம

கைட ச ககால அ ல

ஒ

நால

வா

ெதாழிைல

.இ

விள பினாகனா . இ

கட

இவ ெபாி
ள .

றா

106 பாட க

ைன, 1954). இவாி

ணவ

ஒ

அ களா

கழக ெவளி

இவ

அைழ க ப வதா

க
ஆசிாிய

காண படவி ைல யாத லா

விள பிகனா .

இ

. இத

மணிக

(பாாி நிைலய , ெச
இர

கண

பா க

பதி

ள

கீ

ஆசிாிய

ப றி

ெபய

ள ைஜன க
பல

இட களி

இ

பதா

விள பினாகனா ,
நயினா

நயினாரெரன ெகா

க

க

எ

இவ ைஜன
காண ப வதா

ெகா ள ப கிறா .

க

பதிென

கீ

கண

க

இளவழகனா , ைசவசி தா த

:நா

பதி

1980, தி வர கனா பதி பக , ெச
கி. ஆ. ெப
த பதி

வி வநாத
1954, 11 வ

,
பதி

மணி க ைக, ஆசிாிய
கழக

தி ெந ேவ

.எ

. பால
மிெட ,

தர

பி ைள எ

த பதி

கிற

1904, ஏ ர

ைன 600 018

மணிக

நா

மணிக

, பாாிநிைலய , ெச

ைன – 600 0108,

2007

Enriching Encyclopedias
An encyclopedic entry will be a long essay comprising a preview or introduction followed by
treatment treatment of each item mentioned in the preview in separate sections. Diagrams or maps or
tables or charts are inserted wherever required. Use of photos or images or pictures as illustrative
materials is attempted wherever possible. A dictionary entry restricts itself with whatever connected
with that word such as phonological information, grammatical information, semantic information,
idiomatic information connected with that word etc. While an encyclopedic entry deals with
whatever connected with the subject referred to by the word. At the end of the article a brief
summary of what has been dealt with in the article is presented. For the benefit of those who are
interested in acquiring more information on the topic chosen for treatment in the entry, a short
bibliography is presented. Certain entries may be as long as a few hundred pages, a table of contents
is usually presented enabling easy reference and location of information needed in the entry.
On line encyclopedias offer the additional advantage of being dynamic : new information relating to
the subject dealt with are made available in the encyclopedia as when they are available, not waiting
for the next static format such as the disc or paper based publication to come out. The Tamil and
English Wikis are as much dynamic as the encyclopedia Britannica online, The Wikipedia is one of
the first user-generated content encyclopedia. The principles of democracy is enshrined in its making
and it would never become obsolete as it is dynamic.

228


Mismatch between Technical Know-how and Tamil Studies’ Scholarship

Scholars who have a depth of knowledge in Tamil literary studies are not able to enrich the encyclopedias like the Tamil and English Wikipedia as they do not have the technical skills needed for serving as collaborative editors for them. But those who are well versed in using the computer for serving as a collaborative editor have only a superficial knowledge and understanding of Tamil language and literature in general, especially the Cankam or literature of the Academies in particular. As a result of the mismatch between possession of computing skills needed for serving effectively as a collaborative editor of Tamil and English Wikis and the ability to write convincingly on Tamil literature with suitable citations and making references necessary to make their accounts about Tamil literature authenticated and well documented. Even among the scholars who have a thorough knowledge of Tamil literature, only a small group of native scholars are capable of using the English language and the Tamil language for this purpose. Some of the foreign scholars who can use the English language effectively for describing the ancient Tamil literature they are unable to give a convincing account of the literary works as their understanding of the ancient Tamil literature. Superficiality of native and foreign scholars either in the ability to use the computer or in their understanding of the Tamil literature not only by the foreign scholars but also by the native scholars or in their mastery of using either the English language for English Wiki or the Tamil language for the Tamil wiki. As a result of this mismatch, some of the articles written for the wikis remain Stubs which need elaboration and citations for increasing their verifiability and coherence.

Conclusion

The paper has focused on the sorry state of affairs prevailing in presenting information about Tamil literature in general, classical Tamil Literature in particular in world’s most common dynamic reference books such as English and Tamil Wikipedia. Sample material with problems of presented along with improved versions of information. Certain procedures or practices relating to making encyclopedic entries are referred to in the section that deals with enriching encyclopedias. The chief reasons for information presented in the Wikis being incoherent are identified as the mismatch between the technologically savvy suffering from superficiality in Tamil studies and those who have a thorough knowledge in Tamil Studies but not being aware of the technical skills needed for presenting information in the Wikis as collaborative editors

Notes

- The Interior Landscape: Love Poems from a Classical Tamil Anthology, (1967)
- The Smile of Murugan: On Tamil Literature of South India
- Language Policy and Linguistic Culture in Tamilnadu, Chapter 6, on Tamilnadu from Linguistic Culture and Language Policy, H. Schiffman, 1996.
- An entry complete with all the necessary components cannot be attempted because of space constrains prescribed for this paper
உலகிலும் மக்கள்: தமிழ் லிட்டரீப்ளம்

நூற்றாண்டுகளுக்கு முன்னரே, ஊர்கள் தமிழில் ஆராய்ச்சியுள்ளனர். இதன்படி, தமிழ் லிட்டரீப்ளம் (Wiki) சார்ந்த வரலாறு நோமியான நூல்களும் தமிழில் பாதுகாப்பிட்டுள்ளன. 

1. அறிஞர்

அளவில் கிடைத்தவைகள் முக்கியமானவை, என்றாலும், தமிழில் நூல்கள் வளர்ந்தது அளவுகளில் குறைந்துள்ளது. முடிவு செய்யலாம், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

2,926 என்ற எண்ணிக்கையில் தமிழில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

27,000 என்ற எண்ணிக்கையில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

2,926 என்ற எண்ணிக்கையில் தமிழில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

27,000 என்ற எண்ணிக்கையில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

2,926 என்ற எண்ணிக்கையில் தமிழில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

27,000 என்ற எண்ணிக்கையில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

2,926 என்ற எண்ணிக்கையில் தமிழில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

27,000 என்ற எண்ணிக்கையில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கையில் (edits) செய்துள்ளது. இது முன்னேற்றக்கூடிய விளக்கம் நொந்துள்ளது. நேரடியாக, வண்ணம் என்றால், வழக்கம் என்றால், பலதரப்பட்டவைகள் கூறும் வண்ணம் தமிழில் குறைந்துள்ளது. 

2,926 என்ற எண்ணிக்கையில் தமிழில் பெருந்தாலது மேலும், 778,600 எண்ணிக்கை�
“ஒன்று ஆழ்த்திகள்” போலே விளக்கத் தெளிவு செய்யவுடன். இது மீண்டும் அனுமதிக்கபட்டுள்ளது, குருகை ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறு

"வெப்பங்கள் ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட 

“குறுகிய ஆண்டுக்குட்பட்ட” போலே விளக்கும் குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட 

"குறுகிய ஆண்டுக்குட்பட்ட” போலே விளக்கும் குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட 

2. விளக்கம்: வரையறை

விளக்கம்: வரையறை: அல்லது குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட குறு 

விளக்கம்: வரையறை: அல்லது குறுகிய ஆண்டுக்குட்பட்ட குறுகிய ஆண்டுக்குட்பட்ட 

விளக்கம்: வரையறை: அல்லது குறுகிய ஆண்டுக்குட்பட்ட குறு 

விளக்கம்: வரையறை: அல்லது குறுகிய ஆண்டுக்குட்பட்ட குறு 

விளக்கம்: வரையறை: அல்லது குறுகிய ஆண்டுக்குட்பட்ட 

விளக்கம்: வரையறை: அல்லது குறு 

விளக்கம்: வரையறை: அல்லது 

விளக்கம்: வரையறை: அல்லது 

விளக்கம்: வரையறை: அல்லது 

விளக்கம்: வரையறை: அல்லது
3. தமிழ் நிகழ்வுறை

தமிழ் நிகழ்வுறை வரலாற்றில் 30, 2003 ஆம் ஆண்டு பெரும் பாதுகாப்பு பெற்றது [7]. நிகழ்வுறையை வரலாற்று தொடர்பில் விளக்கம் மிகு தைல்களின் தொடர்புடைய பண்புகளைச் சார்ந்தது[8]. உயல்முறைப்படுத்தப்படும் வளமுறை தொடர்பாக பண்புகளை பொருளாக மிகுது நிகழ்வுறையை விளக்கத்துடன் தொடர்புடையது. தொடர்புடையது, தமிழ் நிகழ்வுறைச் சார்ந்த பண்புகளைக் காண்பெற்றது எண்மம்படி தமிழ் நிகழ்வுறை வளமுறை பொருளாக மிகுது நிகழ்வுறையை விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுறையைப் பண்புகளைப் பற்றிய தொடர்புடையது விளக்கத்துடன் தொடர்புடையது. தமிழ் நிகழ்வுrifoe
அலும இந்தசர் கூற்றமங்கலத்தில், மானிய தொடர்குறிப்பிட்டுக்கொள்ளவும் வாய்வது. (3) பின் படி இந்து வரும்பொழுது(Transform) நூற்றுலட்சுமாக செய்யும் பண்டைய குறுகிய நோக்கங்களும் (தி: எதிர்ப்பு செய்யும் குற்றுக்கனவுடன் edit wars) குறுகிய நோக்கு வகையானது சாதாரணமானது. ஆனால் உயிர் பல இந்து வரும்பொழுது நோக்கங்களில் எடுக்க வேதியியல் விளக்கம் கையறிக்கப்படுகிறது. இதன் தீர்மானம் தற்காலத்தில் தமிழ் போதும் தொடர்ந்து (http://en.wikipedia.org/wiki/Wikipedia:Five_pillars) விளக்கம் பயன்படுத்தப்படுகிறது. (4) சில முழுநிலை வளர்ச்சி வருவாய்த்துள்ள பண்டைய வகையானது சாதாரணமானது பண்டைய வகையானது. எனவென்றும் பகுதிகளில் குறுகிய நோக்கு வகையானது தொடர்ந்து தொடர்ந்து விளக்கம் பயன்படுத்தப்படுகிறது. (7) இதன் வரும்பொழுது நோக்கு வகையானது தமிழ் வைப்படுத்தும் புடைய நோக்கு வகையானது எடுத்துச் செய்யப்பட்டுள்ளது. (7) இதன் தீர்மானம் தற்காலத்தில் எடுக்க வேதியியல் விளக்கம் கையறிக்கப்படுகிறது.


5. ஜிமி வேல்ஸ (Jimmy Wales), லாரி ஸ்நார்ட் (Larry Snger) அகராதி லக்கோர்ஸ் தொகுக்கு நிகழ்வியியிய வைண்டியியன், நுபேடியா (Nupedia) தொகுதியா 2000 வருடா நிகழ்வியன் அகராதி லக்கோர்ஸ் (Mark Guzidal) வேல்ஸ அகராதி நிகழ்வியியிய வைண்டியியன் (CoWeb) தொகுதியா 2000;


8. அப்பன், வைண்டியா, என்.வி.எல்கேரியன், லாபர்ஸ் விக்கிப்பீடியா நிகழ்வியன், பிரிவு 2010..


<table>
<thead>
<tr>
<th>பொருள் விளக்கம்</th>
<th>பொருள்வாய்ந்தவை</th>
<th>குளிர்</th>
<th>வெளியூறு</th>
<th>பொருள் விளக்கம்</th>
</tr>
</thead>
<tbody>
<tr>
<td>தொலைத் தொடர்</td>
<td>320</td>
<td>22k</td>
<td>0.8M</td>
<td>16K</td>
</tr>
<tr>
<td>ஐரோப்பிய பொருள்</td>
<td>450</td>
<td>35k</td>
<td>13MB</td>
<td>14K</td>
</tr>
<tr>
<td>எச்செய் பொருள்</td>
<td>319</td>
<td>11k</td>
<td>94MB</td>
<td>8.4K</td>
</tr>
<tr>
<td>பழமையான பொருள்</td>
<td>231</td>
<td>16k</td>
<td>93MB</td>
<td>8.3K</td>
</tr>
<tr>
<td>வெளியூறு</td>
<td>368</td>
<td>12k</td>
<td>43MB</td>
<td>4.8K</td>
</tr>
<tr>
<td>ஏர்கண் பொருள்</td>
<td>176</td>
<td>7.3k</td>
<td>67MB</td>
<td>7.8K</td>
</tr>
<tr>
<td>பிரிக்குநல்</td>
<td>79</td>
<td>14k</td>
<td>60MB</td>
<td>1.4K</td>
</tr>
<tr>
<td>நூற்றாண்டு பொருள்</td>
<td>39</td>
<td>0.379K</td>
<td>0.033M</td>
<td>0.0645K</td>
</tr>
</tbody>
</table>

*பிரிவு 1: 200 30 2010 ஏன் திறன்முறைகளுக்கு பிற்கால விளக்கங்கள்*
நிக்கிளியாம் - துவிப் தீர்விதம்

2. ரத்திரவன்

மந்துத்தியல் முன்ன்தான் பட்டு அமைவதார,
அுந்தசிவன் விழா திருப்பு பாண்டைக்கனவுண்டு, உரோ நெடுங்கு, தீர்விதம்.
chandrakala.vetrivel@gmail.com <mailto:chandrakala.vetrivel@gmail.com>

කருணையார்

நிக்கிளியாம் தெற்கு 'அந்திசேர்வு' தவறு பிரிவுக்குவிட்டு நெடுங்கையல் சுற்றுலாத்தூர், காண்டு
கால்கைக்கனவுண்டு இருதான். இது ஒரு பல்கலைக் கழகமுண்டு தொண்டுக்கைக்கனவுண்டு. கருணையார் குறிப்பிட்டுதல்கள் தவறு பிரிவுக்குவிட்டு பல்கலைக் கழகமுண்டு சுற்றுலாத்தூர் காண்டு கால்கையல் சுற்றுலாத்தூர் காண்டு கால்கையல் சுற்றுலாத்தூர் காண்டு கால்கையல் சுற்றுலாத்தூர்

காசியாரினுள் சுற்றுலாத்தூர்

அமையும் புல்கள் ஆராய் காண்டு பார்வைக்குவிட்டு குறிப்பிட்டுதல்கள் எந்திருக்கின்றன. பிரிவுக்குவிட்டு பல்கலைக் கழகமுண்டு சுற்றுலாத்தூர் காண்டு கால்கையல் சுற்றுலாத்தூர்

நிக்கிளியாம் தெற்கு 'அந்திசேர்வு' பதவிக் கால்கையல் சுற்றுலாத்தூர்

தமிழ் தொல்லியல் பல்கலைக் கழகமாக

நிக்கிளியாம் தெற்கு தொல்லியல் பல்கலைக் கழகமாக நெடுங்கையல் சுற்றுலாத்தூர். தொல்லியல் பல்கலைக் கழகமாக, சுற்றுலாத்தூர் பல்கலைக் கழகமாக, ஏலமார், பார்வை 53 சுற்றுலாத்தூர் பல்கலைக் கழகமாக நெடுங்கையல் சுற்றுலாத்தூர் காண்டு கால்கையல் சுற்றுலாத்தூர்

குடம்பிதல் வேளாண் செயல்பாடு

அது புது விழாவுடன் தொடங்கு

அதை, அருங்கூற்று

குடம்பிதல் வேளாண் செயல்பாடு, மனுச அருங்கூற்று துறைவு

இம்முனையால் புதுவிளமைத்துச் செயல்பாடை நூற்பார்

குடம்பிதல் வேளாண் செயல்பாடு புதுவிளமைத்துச் செயல்பாடை

236
கல்லறியும் திறந்திகழ:  

சில நிலையில் காரணம் பெறுவது பொருளாதார நிலையானது செயல்பாடு உடன் பிரதிபோக்குகள். இது பொறுப்பிட்டு சில நிலையில் தொடர்ந்து இருக்கும். உடன் பிரதிபோக்குகள், பொறுப்பிட்டு பெறுவது 41 நிலையில் தொடர்ந்து, தானை, பெரும் தொடர்ந்து, குறுகிய உரையாக்குகளை அனுப்பினங்கள் காட்சிப்படுத்துவது உடன் மீண்டும் தொடர்ந்து குறுகிய உரையாக்குகளை அனுப்பினங்கள் காட்சிப்படுத்துவது.  

சல்லாரின் வாழ்பாடு குறிப்பிட்டு நடுப்பகுதிகள் அறிகுறிகள்

சல்லாரின் வாழ்பாடு 1610 காலத்து உடன், போற்று, அமர்ந்தது முகலாதாரம் பொறுப்பிட்டு காட்சிப்படுத்துவது.  

1. தவறுப்பாடுகள் அம்மிழோன் தவறுப்பாடுகள் ---- குறுகிய உடன், போற்று, அமர்ந்தது முகலாதாரம் பொறுப்பிட்டு காட்சிப்படுத்துவது.  

8. தவறுப்பாடுகள் அம்மிழோன் மாடு 2348 காலத்து உடன், அமர்ந்தது முகலாதாரம் பொறுப்பிட்டு காட்சிப்படுத்துவது  

18. பொறுப்பிட்டு நிறுவனத்தக்கட்டு அம்மிழோன் மாடு 3254 காலத்து உடன், அமர்ந்தது முகலாதாரம் பொறுப்பிட்டு காட்சிப்படுத்துவது.  

1 பொறுப்பிட்டு நிறுவனத்தக்கட்டு 5001 (அம்மிழோன்) 'அம்மிழோன் பொறுப்பிட்டு காட்சிப்படுத்துவது' அனுப்பினங்கள்.  

1. சிறிய மாடு 4286 (அம்மிழோன்) ' அம்மிழோன் பொறுப்பிட்டு காட்சிப்படுத்துவது'  

குறிப்பிட்டு நிறுவனத்தக்கட்டு அரசிய பிரதிபோக்குகள்:

த பொறுப்பிட்டு நிறுவனத்தக்கட்டு அரசிய பிரதிபோக்குகள் விளக்கத்தக்கது. அரசிய 11 குறிப்பிட்டு நிறுவனத்தக்கட்டு தொடர் பொறுப்பிட்டு விளக்கத்தக்கது அரசிய பிரதிபோக்குகள். அரசிய (1) (சுருக்கம், (2) குறிப்பிட்டு (3) பொறுப்பிட்டு (4) பொறுப்பிட்டு குறிப்பிட்டு (5) குறிப்பிட்டு பொறுப்பிட்டு (6) பொறுப்பிட்டு, தொடர், பொறுப்பிட்டு (7) பொறுப்பிட்டு நிறுவனத்தக்கட்டு குறிப்பிட்டு (8) பொறுப்பிட்டு, தொடர் (9) தொடர் குறிப்பிட்டு (10) தொடர், தொடர் நிறுவனத்தக்கட்டு பொறுப்பிட்டு (11) தொடர் குறிப்பிட்டு.  

2. பொறுப்பிட்டு நிறுவனத்தக்கட்டு அரசிய பிரதிபோக்குகள் விளக்கத்தக்கது தொடர் பொறுப்பிட்டு நிறுவனத்தக்கட்டு விளக்கத்தக்கது 11 குறிப்பிட்டு பிரதிபோக்குகள். தொடர் குறிப்பிட்டு தொடர் பொறுப்பிட்டு விளக்கத்தக்கது 8 குறிப்பிட்டு பொறுப்பிட்டு பொறுப்பிட்டு அரசிய பிரதிபோக்குகள். குறிப்பிட்டு நிறுவனத்தக்கட்டு விளக்கத்தக்கது 11-ம் பிரதிபோக்கு விளக்கத்தக்கது. எனினும் பொறுப்பிட்டு நிறுவனத்தக்கட்டு குறிப்பிட்டு விளக்கத்தக்கது குறிப்பிட்டு விளக்கத்தக்கது தொடர் குறிப்பிட்டு. 

3. பொறுப்பிட்டு நிறுவனத்தக்கட்டு அரசிய பிரதிபோக்குகள் விளக்கத்தக்கது குறிப்பிட்டு விளக்கத்தக்கது தொடர் குறிப்பிட்டு. எனினும் பொறுப்பிட்டு நிறுவனத்தக்கட்டு விளக்கத்தக்கது குறிப்பிட்டு விளக்கத்தக்கது.  

4. பொறுப்பிட்டு நிறுவனத்தக்கட்டு அரசிய பிரதிபோக்குகள் விளக்கத்தக்கது குறிப்பிட்டு விளக்கத்தக்கது தொடர் குறிப்பிட்டு. எனினும் பொறுப்பிட்டு நிறுவனத்தக்கட்டு விளக்கத்தக்கது குறிப்பிட்டு விளக்கத்தக்கது.
4 x 5 = 20
4 x 2 x 50 = 400
5 x 5 = 25

அம்மையின் குளும் ரீதியான காண்டிகள் கையெழுத்துகள் அறிவிக்கப்பட்டன.

நான் பெயரிட்டது எந்தவாறு வேறுபட்டது அதன் வழி

'புனித பொழுதுச் சுருக்கம்'

'பெரும் பொழுதுச் சுருக்கம்'

'செத்த சுருக்க சுருக்கம்'

பொழிவு மாணவர் அதனையும் பொழிவு

நான் குறிப்பிட்டேன்.

சங்கம்: குலன்னுடன் செல்லும்போது பொன்முன் நிராகரிகள் மூலம் புதுக்கட்ட நிலையைப் பெரும் முயற்சிக்கு வருவதற்கு இயற்ற இயலுமை அச்சாரமாக வருகிறது. இந்தியாவின் விளக்கமும் பொருள்களின் 'சிறையான சமூகங்கள்' பார்வை நூற்றாண்டு வெளியிட்ட நிறுவனைகள் குறிப்பிட்டு வருகிறது.
E-Commerce

(மேல்பாடுகள்)
E-Governance Activities in Tamil Nadu

E.Iniya Nehru
Senior Technical Director,
National Informatics Centre Tamil Nadu State Centre, Chennai
(E-mail:nehru@nic.in)

A number of G2C services are being rendered electronically to the citizens through a single window mechanism. The list includes different types of certificates such as Land Ownership Certificate, Community Certificate, Birth Certificate, Encumbrance Certificate and Nativity Certificate etc. along with other services such as Scholarship portals, permits, passes, licenses to name a few.

The Major E-Governance Projects implemented recently are:

**Tamil Nadu – E-Services of Transport Department:**

A single portal which enables Citizens to file Learner’s License application online, Register their Grievances, know the Status of their redressal, Appointments to visit RTOs and Know their RTOs has been implemented. It also provides the facility to Dealers to file the New Vehicle Registration applications online, generation of Heavy Vehicle Training Course attended certificate online, filing of applications by the Financier for endorsement of Hire Purchase agreement and hire purchase Termination online. 1577 Driving Schools and 1307 Dealers have already enrolled. More than 1,10,000 New Vehicle Registration applications, 60,000 Learner’s License applications are filed through this system and 15,000 Heavy Vehicle Training Course attended certificates are being generated through this system every Month.

**Tamil Nadu - e-Services for Department of Commercial Taxes**

To facilitate the Dealers of Commercial Taxes the Government of TamilNadu provides the anytime anywhere services like Online filing of VAT returns, Online payment of Taxes with 5 different Banks, Online submission of Form-W refund Claims, Online filing of e-Request for saleable forms, Fast Track Clearance system at Checkpost and Online submission of New Registration application.
The system also provides the facility to all the Citizens to search on the commodity code, Tax Rate, TIN number, VAT clarifications, GO’s, Notifications and Circulars issued by the Government, contact Details, VAT Act, Rules and Auction details. It helps the Department to monitor the status of Returns filed by the Dealers and also to identify the Non-filers. More than 3,10,000 Dealers are enabled for filing e-Returns online. As of now, average of 2,30,000 Dealers are filing their Returns online every month. More than 2 Crores of Sales and Purchase invoice data are being captured every month. More than Rs 1500 Crores of Taxes per Month are being paid through e-payment through the 5 Nationalized Banks.

Tamil Nadu - eDistrict : Scholarship System

The Web based system implemented by the Government of Tamil Nadu, provides a facility for students to file application for scholarship online through their respective Institutions, as a first step towards bringing in transparency in the processing of the scholarship forms at various levels of the Government. The system has the necessary provision for requisite backoffice work flow for processing the application. It facilitates quicker processing of scholarship applications of the student and also it provides the status of the scholarship application through the website/CSC/SMS to the students. Automatic SMS messaging services is also sent to individual students who have provided the mobile number in the scholarship application. 1200 institutions dealing with the Scholarship of the BC/MBC and 54 institutions for SC/ST students are making use of
this facility in the state. About 3.7 lakh students for BC&MBC and 2.45 lakh students for Adi Dravidar scholarship have applied so far.

Tamil Nadu - Government e-Procurement System of NIC (GePNIC)

The Tamil Nadu Government implemented the SKoch-Challenger 2000 and eIndia awarded GePNIC – the total automation of the process of physical tendering activity on internet in a faster, and secure environment adopting industry standard open technologies. It is highly generic in nature and can easily be adopted for all kinds of procurement activities such as Goods, Services & Works, by Government offices across the country. All the registered government departments on this application are double authenticated one with the Login-id and Password and other one with the Digital Signature Certificates and PIN for their ensured security transaction on the internet. The departments create Tender and Corrigendum and publish them on the site. Bidders bid against the eligible tender and bids are encrypted with the bid opener’s public key. The same bids are opened only by the Bid openers DSC.

GePNIC has been implemented successfully in the State Governments of Orissa, Tamil Nadu, West Bengal, Uttar Pradesh, Haryana, Chandigarh UT Jharkhand, NICSI, Mahanadi Coalfields Limited (MCL) Orissa, PWD Punjab and Viskhapatnam Port Trust. It is also being implemented for
procurements under Pradhan Mantri Gram Sadak Yojana (PMGSY) of Rural Development Ministry in 21 states, covering the North Eastern states. Around 52,080 tenders, worth over Rs 85,089 Crores, have been processed successfully from 2008 to February 2011

Tamil Nadu – eDistrict Project of Social Welfare Department

An initiative of Social Welfare department covering 215 Citizen Service Centres and 10 Block Offices of the Pilot district of Krishnagiri District to provide five different services of interest to the citizens. Services of Marriage Assistance to Widow Daughter, Orphan Girls, Widow Remarriage, Inter Caste and Child protection scheme are some of the services one can avail through the registered CSCs or facilitation centres at Block/District/Taluk. Being a workflow based application the transparency in processing the application for the marriage assistance is provided to the citizens who can verify the status of their submitted application using the ID number provided in the acknowledgement receipt of their application. Facility to provide the SMS messaging service to the citizens after the approval of their application is also implemented. More than 885 applications have been received to date and are in different stages of processing.
Tamil Nadu – Pregnancy and Infant Cohort Monitoring & Evaluation System

An Online monitoring system that helps to monitor the health status of Pregnant Women registered with any PHC in the rural areas of Tamil Nadu has been successfully implemented from 2008 across 1500+ PHCs. All the 385 Block Medical Officers and 42 District Health Officers are making use of the system for the effective monitoring of the PHCs in monitoring the health conditions of the Pregnancy women and Infant cohort. More than 28 lakh records of Ante Natal checkup details are maintained.

The system captures the details of the pregnant women at various stages like ante-natal care, delivery, post-natal care etc. Similarly it captures details of infants like growth, immunization etc.
Tamil Nadu - Technical Institutions e-Governance Portal for Directorate of Technical Education

To enable the Citizens to get basic details of all the 430+ Polytechnic Colleges and 450+ Engineering Colleges and to enable the Diploma Students studying in these Polytechnic Colleges to know the Attendance details and Semester Examination results for the current Academic Year, this web application was designed and hosted by the DoTE of Government of Tamil Nadu. This portal has 3 different sections for the Citizen, for the Polytechnic Colleges and for the Student of Polytechnic Colleges to get to know the entire details of their interest of Polytechnic and Engineering colleges. Institutions and DoTE are provided with the staff profile and institution profile of all the institutions. More than 5.45 lakh of citizens visited the site from its launch in March 2010.
National Informatics Centre has been providing informatics support to Central Ministries, State Government and District Administration. E-Governance initiatives have already made a large impact in various sectors including agriculture, rural development, judiciary, health, education, transport and administration.
New Media and Tamil - using softwares, tools and Technology

S. Gunasegaran
New Media Trainer
ACTA Trainer (Singapore WDA Approved), 2D animator/3D visual specialist
Temasek Polytechnic, Singapore

Abstract

Recent developments in new media and Tamil computing have changed the overview of creating Tamil contents and engaging applications especially using revolutionary iphone and iPad technology. The Internet has brought the world closer and helped creative industries to make realistic achievements in the e-learning environment. As with introduction of Unicode and technical capabilities to include Tamil as part of iPhone and iPad, the teaching and learning approaches are changing.

This paper looks into ways to create engaging contents, using digital music, animation and high definition audio and videos. It will showcase the usage of new learning environment using Adobe, Toon Boom, Second Life, Blackboard, Moodle and Eon Reality to engage with new Tamil around the globe. It will also examine the need to create more digital games and tools to address the learning needs of young learners of Tamils. It will focus on using blogging, wiki, chatting and tweeting to promote the language effectively.
பொது மாணவர்கள், அரசியல் தொழில் அற்றவர்களும் குறுக்குமாறும், பெண்களும் மாணவர்களும் குழந்தைகள், பெண், மாணவர்கள் மாணவல் சபையுடன் கூடியும். கலை தொழில் குறுக்குவர் பல்வேறு வகையான வணிகத் தலைமையினர், அதிக பொருள் செய்திகளை வெளியிட்டு முடிக்கின்றது என்று முடிக்கிறது.

podcasts, RSS feeds, social media, text messaging, blogs, wikis, virtual world, என தின்பட்டால் உள்ள பார்வையான வரலாறு மாணவரும், கல்வியற்றிருந்து கல்வி முழுவதும் வரலாறு, பெரும்பாலும் பொருளறிக்கைவியல், பெண் மற்றும் பாலமூல பொருளறிக்கையியல், வாழ்வியல், விளையாட்டு, புதிய மாணவரும் பொருளறிக்கை பாடல் காரணமாக வரலாறு குறைந்ததும் குறைந்ததும் குறைந்ததும் குறைந்ததும் குறைந்ததும் புதுவிய மாணவர் வாழ்வு தரும்.

குறிப்பிட்டு அவர்களின் கல்வியும் பார்வையும் விளையாட்டு வாழ்க்கையும். அவர்களின் மாணவர்கள் (blogs) குறிப்பிட்டு வெளியேடுக்கள்.குறுக்குவர் வரலாற்றுக்கு விளையாட்டு முன்னேறுவது குறைந்ததும் குறைந்ததும் குறைந்ததும் குறைந்ததும் குறைந்ததும் குறைந்ததும் வாழ்வு தரும்.

கல்வியறிக்கையின் குறிப்பிட்டு விளையாட்டு வாழ்க்கையை விளையாட்டு வாழ்க்கையை. தமிழக கல்வியும் குறுக்குமாறும் வணிகத் தலைமையினர் குறைந்ததும் குறைந்ததும் குறைந்ததும் வாழ்க்கையை. பெண் பொருளறிக்கை மாணவர் குறைந்ததும் குறைந்ததும் குறைந்ததும் வாழ்க்கையை. பெண் பொருளறிக்கை மாணவர் குறைந்ததும் குறைந்ததும் வாழ்க்கையை.
Microsoft Office in Window7

Google Website

Toon Boom animation

Eon Reality

podcasting

New media and Tamil

Virtual Classrooms
Social network sites

My space, facebook, twitter, Orkut, bebo, blogspot, live journal, myspace, freewebs, Typepad, friendster, blogger, Multiply, live spaces, Yanoom, blackplanet, myyearbook, freewebs, xanga, blogger, wordpress, live journal, blackplanet, myspace, facebook, twitter, friendster, Orkut, bebo, blogspot, live journal, myspace, freewebs, Typepad, xanga, Multiply.

Cloud computing

Microsoft Windows, Apple Mac OS, iPad, iPhone, Android.

Constrains and Remedies

Microsoft, Sony, Nitendo

Tamil in the future

Conclusion
References

- Websites
  - http://namnaadi.edumall.sg
  - http://sangamam.edumall.sg
  - http://www.tamilvu.org/
  - http://spp.moe.edu.my
  - http://www.pazhahutamil.com/login/
  - http://blangahrisetamil.blogspot.com/
  - http://kidsone.in/tamil/
  - http://www.sangapalagai.com/
  - http://beta.toonboom.com/
  - www.microsoft.com
  - www.apple.com

Reference Books

- Salmon, Gilly(2002) E-tivities: The Key to Active Online Learning, London, Kogan Page

- Paul Chin, Using C&IT to Support Teaching
- Reynol Junco & Jeanna Mastrodicasa: Connecting to the net. generation
- Les Lloyd, Editor: Technology and Teaching
தமிழ் கணினி பாவைன

சிவா அஹ்மாத்
தமிழ் கணினி பாவைன் (இலங்கை தமிழ் கணினி பாவைன்)
Email: tamilambu@yahoo.com

தமிழ் கணினி பாவைன் பலவைல் கணினி பாவைன் நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கைக் காலகானை, தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கைக் காலகானை அதன்முதல் 2000ஆம் ஆண்டு முதல் வழங்கும் பலவைல் வாழ்க்கைக் காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கைக் காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கைக் காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கைக் காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கைக் காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கைக் காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கை காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கை காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கை காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கை காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம் ஆண்டு வழங்கும் பலவைல் வாழ்க்கை காலகானை நடவறை தமிழ் கணினி பாவைன் என்ற கால வாழ்க்கைகாலகானை நடவறை தமிழ் கணினி பாவைன் வழங்கும் பலவைல் வாழ்க்கை காலகானை அதன்முதல் 2000ஆம்
குறுக்கு வவசவையிடும் வரலாற்றுறவில் கிட்டத்தட்டு பிரார்த்தனை மிக்கள் நிறைந்து இதுவும் பிடித்து.

1. மகளில் முன்னேறும் பொழுதுபோக்கு கட்டளை
2. குறிப்பிட்டு சென்று புரோபாக்கன
3. குறிப்பிட்டு சென்று புரோபாக்கன் முன்னேறும் பொழுதுபோக்கு கட்டளை
4. அதிகமான பதவிமாடு மகளில் பொழுதுபோக்கு கட்டளை
5. பொது பொழுதுபோக்குக் கட்டளை

சரணமிகு முன்னேறும் பொழுதுபோக்கு கட்டளை

குறுக்கு பொழுதுபோக்குக் கட்டளையில் மகளில் முன்னேறும் பொழுதுபோக்கு கட்டளையில் சாத்திக்கு பயன்படுத்தப்படும் பொழுதுபோக்கு மறைவை அதிகாரிகள், புராண வித்தொச்சி புரட்சிக் குழுவை பயன்படுத்தப்படும் குழுவை மறைவை அதிகாரிகள் வைத்து கொண்டிருக்கின்றன. இதன் காரணமாக, அவர்கள் பொழுதுபோக்கு மறைவை அதிகாரிகள் பயன்படுத்தும் குழுவை மறைவை அதிகாரிகள் வைத்து கொண்டிருக்கின்றன.

திட்டமிகு எல்லாம் பொழுதுபோக்கு கட்டளை

திட்டமிகு பொழுதுபோக்கு கட்டளையில் மகளில் பொழுதுபோக்கு கட்டளையில் சாத்திக்கு பயன்படுத்தப்படும் பொழுதுபோக்கு மறைவை அதிகாரிகள், புராண வித்தொச்சி புரட்சிக் குழுவை பயன்படுத்தப்படும் குழுவை மறைவை அதிகாரிகள் வைத்து கொண்டிருக்கின்றன. இதன் காரணமாக, அவர்கள் பொழுதுபோக்கு மறைவை அதிகாரிகள் பயன்படுத்தும் குழுவை மறைவை அதிகாரிகள் வைத்து கொண்டிருக்கின்றன.

திட்டமிகு எல்லாம் பொழுதுபோக்கு கட்டளையில் தண்டனை விளக்கம் எனினும்.
Electronic Commerce

Dr. B. Neelavathy

Electronic commerce, commonly known as e-comm, e-commerce or eCommerce, consists of the buying and selling of products or services over electronic systems such as the Internet and other computer networks. The amount of trade conducted electronically has grown extraordinarily with widespread Internet usage. The use of commerce is conducted in this way, spurring and drawing on innovations in electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern electronic commerce typically uses the World Wide Web at least at some point in the transaction's lifecycle, although it can encompass a wider range of technologies such as e-mail, mobile devices and telephones as well.

A large percentage of electronic commerce is conducted entirely electronically for virtual items such as access to premium content on a website, but most electronic commerce involves the transportation of physical items in some way. Online retailers are sometimes known as e-tailers and online retail is sometimes known as e-tail. Almost all big retailers have electronic commerce presence on the World Wide Web.

Electronic commerce that is conducted between businesses is referred to as business-to-business or B2B. B2B can be open to all interested parties (e.g. commodity exchange) or limited to specific, pre-qualified participants (private electronic market). Electronic commerce that is conducted between businesses and consumers, on the other hand, is referred to as business-to-consumer or B2C. This is the type of electronic commerce conducted by companies such as Amazon.com. Online shopping is a form of electronic commerce where the buyer is directly online to the seller's computer usually via the internet. There is no intermediary service. The sale and purchase transaction is completed electronically and interactively in real-time such as Amazon.com for new books. If an intermediary is present, then the sale and purchase transaction is called electronic commerce such as eBay.com.

Electronic commerce is generally considered to be the sales aspect of e-business. It also consists of the exchange of data to facilitate the financing and payment aspects of the business transactions.

History

Originally, electronic commerce was identified as the facilitation of commercial transactions electronically, using technology such as Electronic Data Interchange (EDI) and Electronic Funds Transfer (EFT). These were both introduced in the late 1970s, allowing businesses to send commercial documents like purchase orders or invoices electronically. The growth and acceptance of credit cards, automated teller machines (ATM) and telephone banking in the 1980s were also forms of electronic commerce. Another form of e-commerce was the airline reservation system typified by Sabre in the USA and Travicom in the UK.

From the 1990s onwards, electronic commerce would additionally include enterprise resource planning systems (ERP), data mining and data warehousing.
In 1990, Tim Berners-Lee invented the World Wide Web web browser and transformed an academic telecommunication network into a worldwide everyman everyday communication system called internet/www. Commercial enterprise on the Internet was strictly prohibited until 1991. Although the Internet became popular worldwide around 1994 when the first internet online shopping started, it took about five years to introduce security protocols and DSL allowing continual connection to the Internet. By the end of 2000, many European and American business companies offered their services through the World Wide Web. Since then people began to associate a word "ecommerce" with the ability of purchasing various goods through the Internet using secure protocols and electronic payment services.

**DOT-COMS**

Internet use gave a large jump toward the turn of the century, from being common in 26 percent of households in 1998 to 55 percent in 2003. Usage rates continue to climb in the United States and worldwide. This widespread use caused the rise—later followed by the collapse—of many Internet-based businesses, called “dot-coms” for their adoption of the suffix “.com” at the end of their names, referring to their Web site addresses. They used the three Cs method of business—commerce, content, and connection—offering one of the three to possible customers. Although the dot-coms formed the basis for today’s e-commerce, inflated expectations and inexperience in online business transactions lead to the dot-com bubble of 2000 and 2001, when many purely online businesses imploded, costing investors millions. Some of the more famous dotcom busts include Flooz.com, 360Hiphop, Pets.com, Kibu.com, and GovWorks.com, which was featured in the documentary *Startup.com*.

After the dot-com bubble, the surviving companies dropped the coms from the end of their names and went on, some becoming successful businesses. For most companies, however, a combination of physical-based customer service and products with online components offering similar services has proven to be a more trustful method of incorporating e-commerce. In response to the dot-com bust and the continued growing interest in online trade, the Federal Trade Commission, or FTC, began to elaborate on their previous online business regulations.

Chief among the FTCs regulations is the policy that all online advertisement must tell the truth and not mislead customers. As in physical markets, all online claims must be substantiated. Disclaimers can be particularly complex on Web sites, and the FTC requires that all disclaimer information must be easily accessible and readable. In response to worries of online security issues such as account and identity theft, the FTC has also made it clear that online companies should notify customers when collecting personal data, and several Privacy Protection Acts created during the dot-com era were made to enforce that policy.

**Strategies**

One of the first challenges involved in moving to online commerce is how to compete with other e-commerce sites. A common problem in addressing this challenge is that e-commerce is often analyzed from a technical standpoint, not a strategic or marketing perspective. E-commerce provides several technical advantages over off-line commerce. It is much more convenient for the buyer and the seller, as there is no need for face-to-face interaction and Web-based stores are open 24 hours a day. Also, e-commerce purchasing decisions can be made relatively quickly, because a vendor can present all
relevant information immediately to the buyer. These factors lend themselves to a transactional approach, where e-commerce is seen as a way to reduce the costs of acquiring a customer and completing a sale.

In contrast, most successful e-commerce Web sites take a relational view of e-commerce. This perspective views an e-commerce transaction as one step among many in building a lasting relationship with the buyer. This approach requires a long-term, holistic view of the e-commerce purchasing experience, so that buyers are attracted by some unique aspect of an e-commerce Web site, and not by convenience. Since consumers can easily switch to a competing Web site, customer loyalty is the most precious asset for an e-commerce site.

While the primary focus of most Internet activity is on the business-to-business and business-to-consumer facets of e-commerce, other transaction methods are included. The success of eBay and its consumer-to-consumer portal for auction-based transactions has dramatically changed how people and companies conduct business. In addition to having a significant effect on business-to-business transactions, retailers are beginning to tap into this new and dynamic approach to commerce.

**Widgets and e-Commerce**

A widget is a transferrable piece of code that can move itself in and out of Web site data, collecting information or executing a particular function for a metadata program. Some of the most visible widgets are the advertisements seen on most Web sites. These are in fact pieces of code from a third-party business that are being used to communicate marketing messages.

Widgets are one of the most important tools for e-commerce, used most often for distribution of information and online promotional activities. A 2008 article by Ori Soen with *TechNewWorld* explores the new possibilities widgets offer companies interested in e-commerce. Not only are widget-advertisements inexpensive and relatively easy to employ, they can be combined with present marketing efforts and visual productions with the added effect of animation, if desired.

The problem most cited with present-day widget use by e-commerce companies is that online users no longer pay attention to widget advertisement. Most business Web sites accessible today have a multitude of widgets, and the advertisements are often diluted. Like emerging problems with TV commercials, users have learned to simply stop paying attention. Soen, however, sees this as an opportunity for companies to develop more innovative marketing techniques, better online animations, and more effective branding strategies aimed at online users.

Still, e-commerce Web sites are often crowded, and Soen suggests a different focus for widget advertisement: social networks. Social applications, such as MySpace and Facebook, are another field open to creative widget use, but they also offer a more open demographic, namely, people who are more likely to be attracted to creative widgets and—more importantly—have the ability to spread the word to their friends about advertisements that have caught their eye, giving companies two ways to promote instead of one.

Widgets serve a third purpose for e-commerce companies: the ability to collect important data concerning what advertisements are most effective to customers. When widgets are combined with
analysis tools, they can be very useful gatherers of marketing information. They can judge how long a potential customer spends with the widget, and to what extent they interact with the animation. Promotional activities and marketing analysis can be effectively combined.

**Personalization**

One of the key practices to a successful e-commerce company is personalization. Rachelle Crum's 2008 article, “Personalization: Telling E-tail Customers What They Really Want,” lists several ways businesses can personalize their customers' online experiences.

When customers buy products online, they often receive a short list of other items they may be interested in. This is known as recommendation, and because of the ease and access in online business, it is relatively easy for businesses to include in their e-commerce activity. Customers are much more likely to order from the company when they receive a personalized list of products.

Tailored Web pages are another important part of personalization. Many companies have designed their e-commerce businesses to use the data from returning customers to create specific Web pages advertising new products the customer may be interested in, deals that may appeal particularly to the customer, and other information tailored especially for them.

**Mobile Web**

Certain devices, such as the iPhone, are becoming popular for their ability to access the Internet remotely. New technology has allowed remote Web access through phone and other handheld devices to become faster and easier to use. Some e-commerce companies have begun developing Web applications specifically for mobile Web users. This entails creating streamlined Web pages that condense information and allow customers to find what they looking for quickly and without hassle. Since these streamlined applications can be easier to navigate than normal Web pages, and can be accessed from nearly any location, some predictions say the mobile Web will become a powerful tool in the e-commerce field.

There are several different ways companies can make e-commerce more available to mobile Web users. Web designers can simply remove graphics from the Web site for mobile applications, giving users a simplified, text-only site to navigate. Style sheets can also be used, to create other versions of online stores, tailored to specific devices. Or, if a company wanted to devote more time to the project, a second Web site could be created solely for mobile Web users.

**Web Site Creation Tips**

Beyond technology tools such as widgets and mobile Web devices, there are simple ways to improve Web pages and how they read and look. Slight changes in the way a Web page integrates marketing, product information, and visuals can create an enormous difference in the perceptions of customers. Most Web surfers spend a very short time inspecting online stores before moving on, and the right words or the right information, displayed correctly, can make a great difference.

As David Needle says in his 2008 article for Smallbusinesscomputing.com, there are three different kinds of written cues or signposts that companies can place in their Web sites. The first type of signpost is navigation-oriented. This involves the way the online store is constructed—where the links to
products are and where they go, how far down the Web site pages scroll, and links to related sites. “Bread crumbs,” or easy ways for customers to return to the sites they have previously seen, are an excellent tool to give online stores structure and usability.

The second type of signpost is microcontent-related. This refers to Web site headings, URLs, and titles that organize online information. All information in the company's online store should be clear and easy to read and understand.

The third type of signpost is metadata, which is data concerning the information of the Web site itself, such as how many users have accessed it and the keywords within the site that would come up in a search engine or analytical program.

**Barriers to Success**

Despite the growing number of e-commerce success stories, plenty of e-commerce Web sites do not live up to their potential. There were two primary causes of e-commerce failures during the early 2000s.

First, most Web sites offer a truncated e-commerce model, meaning that they do not give Web users the capability to complete an entire sales cycle from initial inquiry to purchase. As analyzed by Forrester Research, the consumer sales cycle has four stages. First, consumers ask questions about what they want to buy. Second, they collect and compare answers. Third, the user makes a decision about the purchase. If the purchase is made, the fourth phase is order payment and fulfillment (delivery of the goods or services). The problem is that many Web sites do not provide enough information or options for all four phases. For example, a site may provide answers about a product, but not answers to the questions that the consumer has in mind. In other cases, the consumer gets to the point where he or she wants to make a purchase, but is not given an adequate variety of payment options to place the actual order.

The second problem occurs when e-commerce efforts are not integrated properly into the corporate organization. A survey by *InteractiveWeek* magazine found that in most companies e-commerce is treated as part of the information system (IS) staff’s responsibility, and not as a business function. While sales and marketing staff generally assist in the development of e-commerce Web sites, final profit and loss responsibility rests with the IS staff. This is a major source of breakdowns in e-commerce strategy because the units that actually make products and services do not have direct responsibility for selling them on the Web. One promising trend is that more companies are beginning to decentralize the authority to create e-commerce sites to individual business units, in the same way that each unit is responsible for its part of a corporate intranet.

**Success Factors**

After studying many aspects of electronic commerce, several consulting and analytic firms created guidelines on how to implement and leverage it successfully. In particular, two organizations have developed lists of critical success factors that seem to capture the state of thinking on this topic. First is the Patricia Seybold Group, which publishes trade newsletters and provides consulting services related to using information technology in corporations. This firm identified five critical e-commerce success factors:
Support customer self-service. If they so desire, Web users should be enabled to complete transactions without assistance.

Nurture customer relationships. Up-front efforts should focus on increasing customer loyalty, not necessarily on maximizing sales.

1. Streamline customer-driven processes. Firms should use Web technology to reengineer back-office processes as they are integrated with e-commerce systems.

2. Target a market of one. Each customer should be treated as an individual market, and personalization technology should be employed to tailor all services and content to the unique needs of each customer.

3. Build communities of interest. A company should make its e-commerce Web site a destination that customers look forward to visiting, not simply a resource people use because they have to conduct a transaction.

A quick review of two successful e-commerce sites, the Amazon.com bookstore site and Dell Computer's Web site, illustrate how many of these principles combine to help develop a strategic e-commerce capability.

Amazon.com, which has one of the highest sales volumes of any Web-based business, has optimized its site for the nature of its products and the preferences of its customers. The site is highly personalized; each visitor to the site, once registered, is greeted by name. The site content also is customized. Using software based on pattern recognition, Amazon.com compares a particular customer's purchase history to its overall record of transactions and generates a list of recommended books that seem to fit his or her interests and tastes. The company has a very integrated customer service support system, so that any customer service representative can access all data on the transactions, purchasing information, and security measures of each customer. The system also supports communications using e-mail, fax, and telephone.

Finally, Amazon.com helps to build a community of users through its Associates Program. Under this program, a Web site can host a hyperlink directly to the Amazon.com site. Any time that a visitor to that site buys books through Amazon.com, the Web site owner receives a share of the transaction revenues. This is a very inexpensive way for Amazon.com to extend its marketing and advertising reach across the Web. Dell Computer also uses personalization and customization tools. For every major corporate customer, Dell creates a special Premier Page, which shows all products covered under purchasing contracts with that firm, as well as the special pricing under those contracts. This ensures that employees of that firm always get the right price for each purchase. Ford Motor Company reports that by encouraging employees to buy PCs from its Premier Page, the company saved $2 million in one year.

Dell also has integrated its e-commerce Web site with all back-office systems, so that when a customer orders a custom-configured PC, that information is automatically transferred to the production system to ensure that the unit is built according to specifications. This also improves customer service; Dell will proactively notify any customer if a production problem or inventory shortage will delay delivery.
Electronic commerce, as used by U.S. firms, has already undergone several generations of evolution. Early experiences helped to stabilize e-commerce technology and set the development path for more sophisticated and useful technologies. Later experiences provided guidelines on strategic approaches and operational models that will help to improve e-commerce success.

Three key issues will determine the long-term viability of electronic commerce. These are:

1. Technological feasibility, or the extent to which technology—bandwidth availability and information reliability, tractability, and security—will be able to sustain exponentially increasing demands worldwide.

2. Socio-cultural acceptability, or the extent to which different global cultures and ways of doing business will accommodate this new mode of transacting, in terms of its nature (not face-to-face), speed, asynchronicity, and unidimensionality.

3. Business profitability, or the extent to which this way of doing business will allow for profit margins to exist at all (e.g., no intermediaries, instant access to sellers, global reach of buyers).

As technology continues to develop and mature, the ability to assess the impact of electronic commerce will become more cogent. Moreover, the significance of privacy, security, and intellectual property rights protection as prerequisites for the successful worldwide diffusion, adoption, and commercial success of Internet-related technologies—especially in places with less democratic political institutions and highly regulated economies—is continually increasing. The differentiation between the Internet (the global network of public computer networks) and intranets (corporate-based computer networks that involve well-defined communities and potentially more promising technology platforms for fostering Internet-related commerce) became significant in the late 1990s and early 2000s. Intranet development has surpassed the Internet in terms of revenue—by 2005 more than half of the world’s Web sites were commercial in nature.

Adverse Possibilities of e-Commerce

Ned Kock, in his book *Encyclopedia of E-collaboration* (2008), gives several possible negative effects of e-commerce, if the trend continues at the same rate it is currently growing.

Global companies with highly developed online stores may already possess the extra edge to attract potential customers. This may leave beginning companies, eager to enter the online market, without much chance to make an impact. International competition may become skewed and lead to an unhealthy type of oligopoly in the e-commerce world.

Some also fear that e-commerce will allow companies to evade certain tax laws, especially when it comes to international trade. New regulations might need to be set for customs concerning online exchanges.

Others wonder how e-commerce will change the job market. While online business offers jobs to those with newer IT skills, it can also displace many traditional jobs.

Advantages of Electronic Commerce

The greatest and the most important advantage of e-commerce, is that it enables a business concern or individual to reach the global market. It caters to the demands of both the national and the
international market, as your business activities are no longer restricted by geographical boundaries. With the help of electronic commerce, even small enterprises can access the global market for selling and purchasing products and services. Even time restrictions are nonexistent while conducting businesses, as e-commerce empowers one to execute business transactions 24 hours a day and even on holidays and weekends. This in turn significantly increases sales and profit.

Electronic commerce gives the customers the opportunity to look for cheaper and quality products. With the help of e-commerce, consumers can easily research on a specific product and sometimes even find out the original manufacturer to purchase a product at a much cheaper price than that charged by the wholesaler. Shopping online is usually more convenient and time saving than conventional shopping. Besides these, people also come across reviews posted by other customers, about the products purchased from a particular e-commerce site, which can help make purchasing decisions.

For business concerns, e-commerce significantly cuts down the cost associated with marketing, customer care, processing, information storage and inventory management. It reduces the time period involved with business process re-engineering, customization of products to meet the demand of particular customers, increasing productivity and customer care services. Electronic commerce reduces the burden of infrastructure to conduct businesses and thereby raises the amount of funds available for profitable investment. It also enables efficient customer care services. On the other hand, it collects and manages information related to customer behavior, which in turn helps develop and adopt an efficient marketing and promotional strategy.

**Disadvantages of Electronic Commerce**

Electronic commerce is also characterized by some technological and inherent limitations which has restricted the number of people using this revolutionary system. One important disadvantage of e-commerce is that the Internet has still not touched the lives of a great number of people, either due to the lack of knowledge or trust. A large number of people do not use the Internet for any kind of financial transaction. Some people simply refuse to trust the authenticity of completely impersonal business transactions, as in the case of e-commerce. Many people have reservations regarding the requirement to disclose personal and private information for security concerns. Many times, the legitimacy and authenticity of different e-commerce sites have also been questioned.

Another limitation of e-commerce is that it is not suitable for perishable commodities like food items. People prefer to shop in the conventional way than to use e-commerce for purchasing food products. So e-commerce is not suitable for such business sectors. The time period required for delivering physical products can also be quite significant in case of e-commerce. A lot of phone calls and e-mails may be required till you get your desired products. However, returning the product and getting a refund can be even more troublesome and time consuming than purchasing, in case if you are not satisfied with a particular product.

**Conclusion**

Thus, on evaluating the various pros and cons of electronic commerce, we can say that the advantages of e-commerce have the potential to outweigh the disadvantages. A proper strategy to address the technical issues and to build up customers trust in the system, can change the present scenario and help e-commerce adapt to the changing needs of the world.
Natural Language Processing

(இயற்கை மொழி பராமானியம்)
An Efficient Tamil Text Compaction System

Tamil Computing Lab (TaCoLa), College of Engineering Guindy, Anna University, Chennai.
haisweety18@gmail.com, jijutodo@gmail.com, madhankarky@gmail.com

Abstract

Tamil is slowly becoming the online language and mobile text messaging languages for many Tamils around the world. Social networks and mobile platforms now extensively support Unicode and applications for keying Tamil text. The number of characters in a text message is limited in some social nets and mobile text messages. The need for compacting the text becomes essential as it translates to saving online storage space, cost and many more factors. The paper proposes a text compaction system for Tamil, a first of its kind in Tamil. The system proposed in this paper handles common Tamil words, acronyms/abbreviations and numbers. Morphological analyzer [1] and Morphological generator are used to stem inflexion words and replace them to compact using a mapping repository. The proposed work is tested with over 10,000 words and it is found that the final result is reduced to 40% of the original text. The paper concludes by discussing possible extensions to this system.

1. Introduction:

In all languages, using compact or short form of words in text messages, emails, and blogs is rapidly increasing. It is particularly popularly amongst young urbanities as it allows for voiceless communication, useful in noisy environment that would defeat a voice conversation and also buffered communication since the message the sender wants to convey can be accessed by the receiver at any time. Compacting text is thus necessary because of limited message length in blog sites and tiny user interface of mobile phone. Getting the shortest word has no rule and it is mainly aimed at understanding. That is, those words should be understood by everyone. We can obtain the compact words by omitting letters, replacing prefix and suffix of through suitable symbols and numbers. This causes the compacted system to be credited with creating a language. The paper proposes a Text Compaction system for Tamil, the primogenital in Tamil.

2. Background:

Tamil is perhaps the only classical language, whose glorious literatures date back to the pre-Christian era, has remained in continuous use for more than many millennia now. Due to the untiring efforts of scholars, researchers and enthusiasts, it has also evolved creatively over the years to the extent that it is also used today profusely in computers, internet, mobile phone etc. Diverse creative efforts have been taking place that would pave the way for a quantum jump in the usage of Tamil in Information Technology. “Tamil Virtual University”, “Centre for Research and Applications of Tamil in Internet”,
“Tamil Software Development Fund” is to quote a few. These efforts paved the way for the motivation of proposing Tamil compaction system in Tamil.

Many compaction systems have been developed for English and other languages. Lee Ming Fung in [2] proposed a Short form Identification and Categorization model based on maximum entropy to identify short forms from actual words and acronyms/abbreviations and categorize the short forms into the short forms formed from letter omission and those formed through phonetic substitution of parts of words. In the proposed system the compact words are formed in a diverse variety of ways such as omission, truncation and phonetic substitution. Acronym Identification and detection has been much researched. Acrophile in [3] automatically searches acronyms from acronym-expansion pairs from domain specific databases. By acronyms expansion pairs, we refer to a pairs each containing acronyms and their full expanded form or meaning. The paper makes use of acronym expansion pairs to replace the full expanded form with the acronyms.

3. Text Compaction Framework:

The figure below presents the various components of the framework.

![Text Compaction Framework Diagram]

3.1 Input Processing

The input text is tokenized based on a delimiter and is passed on to the Morphological Analyzer. The analyzer removes the suffix (if present) added to the word and delivers the root word (RW). For example if the input to the analyzer is கணிఆபபாறி the output is given as கணி.

3.2 Identification of the type

The proposed paper handles three categories of words; common Tamil words, Abbreviations /acronyms, numbers. Now, the category to which the RW belongs is to be identified. The RW is checked to decide the category of abbreviations /acronyms. This is done by comparing the root word with the keys of the hash map (2.3). If the comparison results are true then the RW is considered as the abnormal word (AW) i.e. it belongs to the category of acronyms/abbreviations, else, it is treated as the normal word (NW) i.e. it belongs to either the first or third category.
3.3 Extraction of the compact word

If the word is identified as a normal word, it is passed to a tree which is built dynamically from the set of words that has already been stored in the dictionary. The NW is then searched in the binary search tree. On finding the NW in the binary search tree, the compact word is retrieved with an efficient mapping algorithm that maps each of the normal word with its compact word.

Say suppose the word is an abnormal word, its compact word is retrieved in the following manner. A linked hash map is built for all the abbreviated words. The hash map uses the first word the abbreviated word as its key. Again with the help of an efficient mapping algorithm, the compact word is retrieved. In case the NW is a number name it is replaced with the numerals based on the place value system.

3.4 Output Processing

The compact word that is being extracted is passed on the Tamil tool Morphological Generator to add the suitable suffix to cater to the rules of the language.

4. Results and Analysis:

The paper proposes the following layout for displaying the results to the user. It has two text areas: the one on the left is for entering the input text and the other on the right for displaying the output. The user can also view the no of characters that have been reduced in the output text.

Efficiency of the system can be calculated as (no of characters in the input text / no of characters in the output text) X 100%. The proposed work is tested with over 10,000 words and it is found that the final result is reduced to 40% of the original text.
5. Conclusion and Future work:

The paper describes the Tamil Compaction System, a framework for shrinking the text such that its meaning remains the same. Different subsystems and components of the framework are described in detail. Results from the implementation of this Tamil compaction system framework is provided and is compared against the compacting third party applications of social networking sites that are implemented for English language. Improving the mapping for words which are frequently used, conceptual reducing, integrating numerical analyser will take this system to its next level.

References:

- *Acrophile* (L.S Larkey, P Ogilvie, MA Price, B Tamilio, 2000) a system that automatically searches acronym expansion pairs.
- *Short Message Service (SMS) Texting Symbols: A Functional Analysis of 10,000 Cellular Phone Text Messages* by Robert E. Beasley, Franklin College.
Abstract

Cricket is one of the most followed sports in the Indian subcontinent. There is a wide requirement for natural language descriptions, which summarize a cricket match effectively. The process of generating match summaries from statistical data is a manual process. The objective of this paper is to propose a framework for automatic analysis and summary generation for a cricket match in Tamil, with the scorecard of the match as the input. Data analytics is performed on the statistical match data, to mine all frequently occurring patterns. The paper proposes a parameter called Interestingness, which quantifies the interestingness of the match. The paper also proposes a customization model for the summary. We propose an evaluation parameter called humanness, which quantifies the similarity between the output and a manually written summary. Discussing the results and analyzing the summaries generated for matches based on scorecards, this paper concludes with proposing some extensions for future developments.

1. Introduction

The number of websites which facilitate people to follow and analyze sports has increased manyfold. Among them, there are an exceptionally large number of sites devoted to Cricket. Mostly these involve participation of experts, who present their views and summaries in English about cricket matches. There are no such sites which provide similar services in Tamil. In this case it is also desirable if there is an alternative for human creativity. As a solution the paper proposes an automated Tamil summary generation framework which is capable of analyzing and generating a Tamil summary about a cricket match, provided the scorecard as the input. This paper discusses the overall architecture and implementation details of such a framework.

The large amount of data in this domain makes it possible to apply data mining and data analytics techniques. The input scorecard is analyzed to construct feature vectors, which are then subjected to data mining. Based on the various parameters identified, the interestingness of the match is quantified.

The summary generation part involves extraction of key players and events from a match. Appropriate sentences are then synthesized to express these selected events. The sentence constructs and the vocabulary used are chosen based on the linguistic ability specified by the user. Then the sentences are combined into a meaningful summary.

The results of the system, i.e. the summaries, are evaluated based on the Humanness parameter. This parameter gives the degree of similarity between the generated summary and the manually written summary, with which it is compared. This value helps us decide, the level of creativity achieved by
the system. 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. 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 extensions to the current framework and directions for further studies in the field of Tamil Summary Generation Systems.

2. Background

In the literature there are existing works on summary generation from statistical data. Alice Oh et al. generated multiple stories about a single baseball game based on different perspectives using a reordering algorithm [1]. Ehud Reiter et al. in their book building natural language generation systems explain the difference between natural language generation and natural language processing and also describe the various steps involved in the natural language generation process with examples [2]. Jacques Robin et al. presented a system (called STREAK) for summarizing data in natural language. It focuses on basketball game to design and evaluate the system [3]. L. Bourbeau et al. came up with the FoG (Forecast Generator) using the streamlined version of the Meaning-Text linguistic model. This system was capable of generating weather forecasts in both English and French [4].

3. Summary Generation Framework

The Tamil Cricket Summary Generator consists of the following major components:

- Data Gathering and Modeling module
- Data Mining and Data Analytics module
- Summary Generator
- Evaluator

Figure (1) given below depicts the Summary Generation framework.

Figure 1: Tamil Cricket Summary Generator Framework
3.1 Data Gathering and Modeling Module

Data gathering is the first step of the system. The data to be gathered is present in internet. This module has a custom designed parser, for the tag structure of the site. The user must provide the URL from where the particular match’s data can be obtained. The module checks whether the match has already been processed. If not the parser parses the page and retrieves the statistical data. The statistical data is then modeled in the form of the predefined feature vectors.

3.2 Data Mining and Analytics Module

Modified version of Apriori algorithm is used to find the association rules from the feature vectors. After performing mathematical analysis using correlation of variance (CoV), CoV is plotted against average to give an idea about how consistent the player is. The interestingness of the match is calculated based on the weighted average of the scores assigned to the factors identified, they include the Winning margin, Team history, Individual records made, High run rate, Series state, Relative position in international ranking, Reaction in social networks etc.

3.3 Summary Generator

The summary generator part of the framework consists of the following sub modules Content Determiner, Aggregator, Tamil Morphological Generator and Layout Determiner. The events to be included in the summary are not predefined and are not the same for every match. Based on the interestingness of the total match, the interestingness of the individual events and the expert level chosen by the user, particular events are chosen to be included in the summary. The content determiner is responsible for identifying those facts which are worth mentioning in the summary.

Aggregation of relevant events from other matches in the summary will make it more readable and interesting. The aggregator performs this function. It chooses events based on their similarity and coherence and aggregates them with the key events selected in the content determiner module.

As a next step, the sentences used to describe the events are synthesized. The sentence which is the most apt to the current event under consideration is selected. The vocabulary used in the sentence and the depth to which an event is discussed is also varied based on the expert level of the user. The nouns in the key events are passed to the morphological generator along with the desired case endings and the generated variants are added to the sentences.

The layout determiner module chooses the layout of the summary to be generated. The layout is varied based on the interestingness of the match. The sentences are aggregated in the fashion of the layout selected and the final output summary is passed to Evaluator.

3.4 Evaluator

The summary generated by the system is evaluated based on its degree of similarity with human written summaries. The summaries are compared based on two parameters, the Nouns Mentioned and the Events Mentioned.

The nouns and the events in the summaries are extracted along with their absolute positions. The events in the summary are modeled as a set consisting of, one or more Performers (the persons who takes part in the event), Numeral (the numeric part involved in the event e.g. 4 wickets) and a
Descriptor (the action connecting the Performer and the Numeral). Their absolute positions refer to the sentence number in which they are mentioned. Then these absolute positions are normalized based on the total number of sentences present in the summary. Three different scores are calculated they are,

- **Similarity Score**: The ratio of the number of nouns and events mentioned in both the summaries to the total number of nouns and events mentioned at least in one summary.
- **Count Score**: The ratio of the number of nouns and events mentioned in the system generated summary to the number of nouns and events mentioned in the human written summary.
- **Closeness Score**: The degree of closeness, in terms of the normalized positions of the nouns and events mentioned in both the summaries.

A weighted average of these three scores yields the final humanness score.

### 4. Implementation

To implement the proposed framework, espncricinfo a reliable and prominent site for Cricket data is chosen as the source of input. The framework was implemented in java. The URL of the match for which the summary is to be generated is obtained from the user. The feature vectors designed for modeling a match are stored as rows with unique identities, in the back end oracle database. The patterns which are generated as a result of data mining are validated based on the support and confidence parameters. As a design decision all nouns are stored in English and are translated on the fly using a constantly updated look up database. This decision was taken to allow interoperability and easy extension of the system to other languages in future. The sentence pattern files are stored external to the system, so as to allow modifications without changes in the system. The summary generated for the match is stored in the back end, indexed with the unique identity assigned already. The user interface is designed to be simple and robust. It allows the users to search matches based on various parameters and also to save their preferences.

### 4.1 Results

Score cards of 90 One Day International matches where retrieved and their summaries were generated. These include matches between 9 countries. Both individual matches and series were considered. A large number of hidden patterns in cricket domain have been retrieved based on the algorithm used. The patterns have been validated and the ones which are interesting have been reported. The factors contributing to the interestingness of the match have been identified and the weights associated with them have been found. The consistency of a player has been modelled and consistency analysis of a player is done to analyse his performance.

The difference in the language used and the events mentioned in the summary is pronounced when the user opts for an expert level. Similar facts occurring in the past have been identified and added to the summary. Each summary was compared with two human written summaries, one an expert summary and other an average summary, their cumulative scores were considered. The humanness score of the summaries tend to be in the range of 70% to 85%. The recurrence of layouts is also minimal, which reflects the fact that the summaries generated are not monotonous.
5. Conclusion and future work

In this paper we have proposed the framework for an Automated Tamil Cricket Summary Generator. The current implementation of the system can be enhanced by adding machine learning capabilities to make the summaries more human and interesting. The system can be extended to produce summaries in multiple languages apart from Tamil. The system can be enhanced to generate summaries about the match in real time. As a next level the system can be modified for summary generation in other sports too.

![Figure 2: Screenshot of the Tamil Cricket Summary Generation System](image)

The framework can be used as a guideline to develop summary generation systems, which can be applied for any domain where frequent numerical reports are used. (Weather Prediction, Industrial Quality Testing etc)

References

Lyric Mining: Word, Rhyme & Concept

Co-occurrence Analysis

Karthika Ranganathan, T.V Geetha, Ranjani Parthasarathi & Madhan Karky

Tamil Computing Lab (TaCoLa),
College of Engineering Guindy, Anna University, Chennai.
karthika.cyr@gmail.com, madhankarky@gmail.com

ABSTRACT

Computational creativity is one area of NLP which requires extensive analysis of large datasets. Laalalaa [1] framework for Lyric analysis and generation proposed a lyric analysis subsystem that required statistical analysis of Tamil lyrics. In this paper, we propose a data analysis model for words, rhymes and their usage in Tamil lyrics. The proposed analysis model extracts the root words from lyrics using a morphological analyzer [2] to compute the word frequency across the lyric dataset. The words in their unanalyzed form are used for computing the frequent rhyme, alliteration and end-rhyme pairs using adapted apriori algorithm. Frequent co-occurring concepts in lyrics are also computed using Agaraadhi, an on-line Tamil dictionary. Presenting the results, this paper concludes by discussing the need of such an analysis to compute freshness, pleasantness of a lyric and using these statistics for Lyric Generation.

Keywords : Tamil Lyrics, Morphological Analyser, Apriori algorithm.

I. INTRODUCTION

Tamil is one of the world’s oldest languages and has a Classical status. Numerous forms of literature exist in Tamil language of which, lyrics play a vital role in taking the language to every household in form of original film soundtracks, jingles, private albums, and commercials. With over thousands of lyrics being created every year, we do not have proper tools to model and analyse lyrics. Such an analysis framework would enable one to see various patterns of words, combinations and thoughts used over time. The analysis framework will also make it possible to generate fresh lyrics where the freshness can be associated with the concepts and thoughts associated with the lyric.

In this paper, we discuss about Tamil lyric Analysis on the basis of word usage, rhyme usage and the co-occurrence of word. The frequency of word usage is identified by considering a morphological root of the word using morphological analyser, instead of considering terms. For analysing the frequent rhyme, alliteration and end-rhyme pairs, we adapted Apriori algorithm [4]. To identify the co-occurring concepts in lyrics, we used “Agaraadhi”, an on-line Tamil dictionary Framework [3] and a new algorithm has been proposed to compute the frequent usage of co-occurring concepts in lyrics.

The rest of this paper has been organized as follows. In Section 2, we explain about the algorithm and tools. In Section 3, we explain the methodology and in Section 4, we discuss our results. Conclusions and future extensions to this work are presented in section 5.
2. MORPHOLOGICAL ANALYSIS AND APRIORI ALGORITHM

Morphological analysis is the process of segmenting words into morphemes and identifying its grammatical categories. For a given word, morphological analyser (MA) generates its root word and its grammatical information. The role of morphological analyser in the proposed work is to identify the noun and verb morphology of a given word, examples are as follows

Example 1, for noun morphology:

இராமைன (Ramanai)
இராமஒuni0.A9_uni0.ஙஞ (Raman) + ஐ (ai)
Entity + Accusative Case

Example 2, for verb morphology:

சென்றான (Senraan)
சென் (sel) + ன் (R) + அன் (Aan)
Verb + Past Tense Marker + Third Person Masculine Singular Suffix

In the proposed work, the frequency of a word is identified by considering the variations of a noun in terms of its morphology. For instance, the variations of Ramanai such as Ramanaal, Ramanukku, Ramanin, Ramanadhu are also counted for the word Raman.

The Apriori Algorithm is an influential algorithm for mining frequent item sets for Boolean association rules [4]. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. It has objective measures: support and confidence. The support of an association pattern is the percentage of task-relevant data transactions for the apparent pattern. Confidence can be defined as the measure of certainty ortrustworthiness associated with each discovered pattern.

3. LYRIC ANALYSIS

(i) Word Analysis

The frequency of words is used to associate a popularity score for each word. This score is proposed to be used for lyric generation part of the framework proposed in [1]. In this work, the popularity score of a word has been identified from lyrics. In lyrics, the words are mainly attached with the suffix. So, the root words are taken into consideration for determining its frequency count. The root words are identified using morphological analyser. The algorithm to find the word usage is illustrated below:

Algorithm

Let \( L_0 \) is the set of lyric dataset and \( L_S \) denotes the set of sentences of lyric dataset and \( L_W \) denotes the set of words in the lyric dataset. \( W_C \) denotes the word count across all lyric dataset. Let \( m \) be the total number of sentences in lyrics and \( n \) be the total number of words in lyrics.
a) Given a Lyric dataset \( L_D \)

b) For each \( L_{Si} \leftarrow 1 \) to \( m \)

   Split the sentence \( L_S \) into words \( L_W \)

c) For each \( L_{Wj} \leftarrow 1 \) to \( n \)

   \( R_W \leftarrow \text{ProcessMorphAnalyser}(L_{Wj}) \)

d) Let \( R_W \) be the root word

   if \( R_W \) exist, then add into the word count list \( W_C \)

   else add into the word count list \( W_C \)

e) Return \( W_C \).

Here \( \text{ProcessMorphAnalyser}(L_{Wj}) \) returns the root of the given word.

(ii) Rhyme Analysis

Alliteration (Monai) is the repetition of the same letter at the beginning of words. The rhyme (Edhugai) is defined as the repetition of the same letter at the second position of words. The end rhyme (iyaibu) is defined as the repetition of the same letter at the last position of words. The example of alliteration, rhyme and end rhyme is given below:

Examples:

- உயஔuniஆBBFஐzzஈஆஔtm_ra_halant
- உஔuniஆBAண_uniஆBCD

and

- இதயஔuniஆBAE_uniஆBCD
- காதஔuniஆBBஊ_uniஆBCD

rhyme in alliteration (monai) as they start with the same letter.

- துவபி and டும் rhyme in rhyme (edhugai) as they share the same second letter.

- பாதசக்கம் and பாதகுண்டகம் rhyme in end – rhyme (iyaibu) as they share the same last letter.

We have adapted apriori algorithm to find the frequency count of rhyme, alliteration and end rhyme pairs of Tamil lyrics which has been illustrated below:

Algorithm:

Let \( L_D \) be the set of lyric dataset and \( L_S \) denote the set of sentences of Lyric dataset. Let \( m \) be the total number of sentences in lyrics. Let \( P_{C1} \) denote the count of alliteration and \( P_{C2} \) denote the count of rhyme and \( P_{C3} \) denote the count of end – rhyme.

a) Given lyric dataset \( L_D \).

b) For each \( L_{S} \leftarrow 1 \) to \( m \)

   Join the pair of sentences \( (L_P) \)

c) For each \( L_P \)

   Consider the first \( (L_{P1}) \) and last words \( (L_{P2}) \)

d) Rhyme \( (L_{P1}, L_{P2}) \)

e) return \( P_C \).
Algorithm: Rhyme (L\textsubscript{P1}, L\textsubscript{P2})

a) Let k denote the \textit{i}\textsubscript{th} character. Let M\textsubscript{L} denote the alliteration (monai) list and R\textsubscript{L} denote the rhyme (edhugai) list and E\textsubscript{L} denote the end – rhyme (iyaibu) list.

b) For \forall i,

\begin{align*}
&\text{if } k = 1, \quad \text{if } L\textsubscript{P1}(k) = L\textsubscript{P2}(k), \text{ then add into } M\textsubscript{L} \text{ list and increment } P\textsubscript{C1} \\
&\text{if } k = 2, \quad \text{if } L\textsubscript{P1}(k) = L\textsubscript{P2}(k), \text{ then add into } R\textsubscript{L} \text{ list and increment } P\textsubscript{C2} \\
&\text{if } k = i - 1, \quad \text{if } L\textsubscript{P1}(k) = L\textsubscript{P2}(k), \text{ then add into } E\textsubscript{L} \text{ list and increment } P\textsubscript{C3}
\end{align*}

(iii) Co-occurrence concept Analysis

Co-occurrence is defining the frequent occurrence of two terms from a text corpus on the either side in a certain order. This word information in NLP system is extremely high. It is very important for cancelling the ambiguous and the polysemy of words to improve the accuracy of the entire system [5].

In this method, to improve the efficiency of co-occurrence, we have been considering the concept of each word. The concept for each word has been identified using the Agaraadhi, an on-line Tamil dictionary. The example for concept word which has been in lyric is given below:

Example: The word "நிலஒ" which has the concept ஓவெநு, மாதஒ, ஓவெநுணைணைச்சுலா, ஓவெநுணைணைச்சுலா, ஓவெநுணைணைச்சுலா, ஓவெநுணைணைச்சுலா.

By considering these concepts, we have been determining the co-occurring words using our own algorithm is described below:

Algorithm:

Let L\textsubscript{D} denote the set of Lyric dataset and W denote each word in lyric. Let C\textsubscript{W} denote the set of concepts for each word. Let W\textsubscript{C} denote the word count.

a) If the word C\textsubscript{W} identify, then consider the next word.

\begin{align*}
&\text{Increment the count } W\textsubscript{C}
\end{align*}

b) Else the word W and consider the next word.

\begin{align*}
&\text{Increment the count } W\textsubscript{C}
\end{align*}

c) return W\textsubscript{C}

4. RESULTS

The lyric corpus of more than two thousand songs were analysed for the word usage, rhyme usage and Co-occurrence concepts usage. The analysed results are given below:
Table 1 shows the list of top 10 usage words in lyrics.

<table>
<thead>
<tr>
<th>WORDS</th>
<th>USAGE</th>
<th>WORDS</th>
<th>USAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ஐது</td>
<td>2009</td>
<td>மர</td>
<td>1062</td>
</tr>
<tr>
<td>காணா</td>
<td>1941</td>
<td>எ க</td>
<td>987</td>
</tr>
<tr>
<td>மான்</td>
<td>1645</td>
<td>கரையா</td>
<td>965</td>
</tr>
<tr>
<td>இளா</td>
<td>1556</td>
<td>பு</td>
<td>857</td>
</tr>
<tr>
<td>கச்சை</td>
<td>1153</td>
<td>தின்பலன்</td>
<td>793</td>
</tr>
</tbody>
</table>

Table 2 shows the list of top 10 rhyme words in lyrics.

<table>
<thead>
<tr>
<th>EDHUGAI</th>
<th>USAGE</th>
<th>MONAI</th>
<th>USAGE</th>
<th>IYAIBU</th>
<th>USAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>கூத்து, கூத்து</td>
<td>107975</td>
<td>கூத்து, கூத்து</td>
<td>33492</td>
<td>கூத்து, கூத்து</td>
<td>111552</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>80125</td>
<td>கூத்து, கூத்து</td>
<td>26289</td>
<td>கூத்து, கூத்து</td>
<td>83435</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>61204</td>
<td>கூத்து, கூத்து</td>
<td>16478</td>
<td>கூத்து, கூத்து</td>
<td>63543</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>33731</td>
<td>கூத்து, கூத்து</td>
<td>15405</td>
<td>கூத்து, கூத்து</td>
<td>18411</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>32570</td>
<td>கூத்து, கூத்து</td>
<td>14001</td>
<td>கூத்து, கூத்து</td>
<td>16478</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>25747</td>
<td>கூத்து, கூத்து</td>
<td>11640</td>
<td>கூத்து, கூத்து</td>
<td>14001</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>24867</td>
<td>மாணி, மாணி</td>
<td>10993</td>
<td>மாணி, மாணி</td>
<td>12524</td>
</tr>
<tr>
<td>மாணி, மாணி</td>
<td>19297</td>
<td>குழலி, குழலி</td>
<td>9985</td>
<td>குழலி, குழலி</td>
<td>10524</td>
</tr>
<tr>
<td>மாணி, மாணி</td>
<td>18935</td>
<td>குழலி, குழலி</td>
<td>9976</td>
<td>குழலி, குழலி</td>
<td>9367</td>
</tr>
<tr>
<td>மாணி, மாணி</td>
<td>14486</td>
<td>மாணி, மாணி</td>
<td>9962</td>
<td>மாணி, மாணி</td>
<td>4818</td>
</tr>
</tbody>
</table>

Table 3 shows the list of top 10 co-occurring concept words in lyrics.

<table>
<thead>
<tr>
<th>CO-OCCURRING WORDS</th>
<th>USAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>அசைவு, அசைவு</td>
<td>638</td>
</tr>
<tr>
<td>சிறைவு, சிறைவு</td>
<td>530</td>
</tr>
<tr>
<td>ஓள, ஓள்</td>
<td>506</td>
</tr>
<tr>
<td>கூத்து, காளு</td>
<td>478</td>
</tr>
<tr>
<td>கூத்து, காளு</td>
<td>469</td>
</tr>
<tr>
<td>மாணி, மாணி</td>
<td>434</td>
</tr>
<tr>
<td>கூத்து, கூத்து</td>
<td>419</td>
</tr>
<tr>
<td>கூத்து, காளு</td>
<td>367</td>
</tr>
<tr>
<td>கூத்து, காளு</td>
<td>302</td>
</tr>
<tr>
<td>கூத்து, காளு</td>
<td>287</td>
</tr>
</tbody>
</table>
By analysing those data, this shows that the most of lyrics which predicts the emotion of happiness and love. In the adapted apriori algorithm, the support which represents the total number of pair words with the total number of combination of sentences and the confidence which described the total number of pair words with the total number of pair of sentences. The results may vary if the number of lyrics used for the analysis is increased.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we discussed the usage of words and rhymes in the lyrics dataset. The adapted apriori algorithm has been used to detect the frequency count for rhyme, alliteration and end word pairs. This analysis has been mainly used in the lyric generation and computing freshness scoring for lyrics. Frequent co-occurring concept is also been identified for the development of lyric extraction, semantic relationship, word sense identification and sentence similarity. Possible extensions of this work could be the detection of emotions in lyrics by genre classification and identify genre specific rhymes and concept co-occurrence.

REFERENCES

Template based Multilingual Summary Generation

Subalalitha C.N, E.Umamaheswari, T V Geetha, Ranjani Parthasarathi & Madhan Karky
subalalitha@gmail.com
Tamil Computing Lab (TaCoLa)
College of Engineering Guindy Anna University, Chennai.

Abstract

Summarization of large text documents becomes an essential task in many Natural Language processing (NLP) applications. Certain NLP applications deal with domain specific text documents and demand for a domain specific summary. When the essential facts are extracted specific to the domain, the summary proves to be more efficient. The proposed system builds a bilingual summary for an Information Retrieval (IR) system named CoRee, which tackles Tamil Language and English Language text documents [1]. As the input documents are tourism domain specific documents, the summary is extracted based on specially designed seven tourism specific templates 7 both for Tamil and English. The templates are filled in with the required information extracted from the UNL representation and a bilingual summary is generated for each text document irrespective of the language of input text document. The efficiency of the summary has been tested manually and it has achieved 90% efficiency. This efficiency depends on factors other than summary generation such as enconversion accuracy and dictionary entry coverage. The proposed system can be extended for many languages in future.

1. Introduction

Automatic summary generation has been a research problem for over 40 years [2]. Summarizing the texts helps in avoiding information overload and also saves time. Multi lingual Natural Language applications have emerged in great number in recent years. This makes the need for a multi lingual summary generation a quintessential task. Alkesh patel et al have come up with a multi lingual summary generation by using structural and statistical factors [2]. David Kirk Evans has generated multi lingual summary using text similarities existing in the sentences [3]. Dragomir Radev et al have developed a multi lingual summary generation tool named MEAD using centroid and query based methods. They have also used many learning techniques such as decision trees, Support Vector Machines (SVM) and Maximum Entropy [4].

All the above works on multi lingual summarization have not used a interlingua document representation. We propose that a multi lingual summary can be generated with much more ease by using a interlingua document representation language called, “Universal Networking Language” (UNL) [5]. UNL converts every term present in a natural language text document into a language independent concept, thereby making the applications built using it a language independent one. The proposed work extracts a domain specific summary, as the UNL documents used are tourism domain specific. Tourism specific templates are framed and the sentences fitting the templates are chosen and formed as a summary.
The rest of the paper is organized as follows. Section 2 gives a brief introduction about UNL. Section 3 describes the proposed summarization technique. Section 4 discusses the evaluation of the proposed work. Section 5 reveals the enhancements needed to the proposed work and Section 6 gives the conclusion of the paper.

2. Universal Networking Language

UNL is an intermediate language that processes knowledge across language barriers. UNL captures the semantics of the natural language text by converting the terms present in the document to concepts. These concepts are connected to the other concept through UNL relations. There are 46 UNL relations like \texttt{plf(Place From)}, \texttt{plt(Place To)}, \texttt{tmf(Time from)}, \texttt{tmt(Time to)} etc [1]. This process of converting a natural language text to UNL document is known as Enconversion and the reverse process is known as Deconversion. The UNL document is normally represented as a graph where the nodes are concepts and edges are UNL relations. An example UNL graph is shown for the example 1.

Example 1: John was playing in the garden.

![Figure 1: UNL graph for Example 1](image-url)

The nodes of graph namely, “\texttt{John(iof>person)}”, “\texttt{Play(icl>action)}” and “\texttt{garden(icl>place)}” represent the terms John, playing and garden present in the example 1. The semantic constraints in the concepts, “\texttt{iof>person}”, “\texttt{icl>action}” and “\texttt{icl>place}” denotes the context in which the concepts occur. The edges namely, “\texttt{agt}” and “\texttt{plc}” indicates that, the concepts involved are agents and place. From the above discussion, it is shown that the UNL inherits many semantic information from the natural language text and portrays in a language independent fashion.

The proposed work uses Tamil language text documents and English language documents enconverted to UNL for summary extraction which is described in the next section.

3. Template based Information Extraction

As discussed earlier, the summary is generated using the tourism specific templates. Figure 2 shows the over view of the proposed summary generation framework. The Framework consists of both language dependent and independent parts. The functionalities involving UNL are language independent and the inputs supplied to the framework to generate bilingual summary are the language dependent parts. The bilingual summary generation is explained in the coming sections.
The seven templates describe about the tourism specific information of a place such as god, food, flora and fauna, boarding facility, transport facility, place and distance. The correct information for these templates are extracted as discussed below. The usage of semantics helps greatly in eliminating the ambiguities that may arise while picking up a concept to fill the slot. For instance, the word “bat” may denote a cricket bat or the mammal bat.

This type of ambiguity is resolved by the semantic constraint, as the cricket bat will get the semantic constraint, “obj<thing (object thing)”, whereas the mammal gets the semantic constraint, “icl>mammal”. Table 1 displays few semantics used for the respective templates.

The extracted tourism specific concepts are converted to the target language terms for building a summary using the sentence patterns which is explained in the next section.

4 Multi Lingual Summary generation

The information (concepts) extracted from the UNL graph using the templates are converted to the target language term using the respective UNL dictionary. For instance, to generate the English summary, the concepts comprising the semantic constraints are converted to English terms using the English UNL dictionary which consists of mapping between English terms and UNL concepts. These terms which when filled into the appropriate English sentence patterns, gives a English summary. The same procedure is done for building a Tamil summary. For each UNL graph irrespective of its source language, a summary in Tamil and English are generated.
<table>
<thead>
<tr>
<th>Template</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>God</td>
<td>iof&gt;god, iof&gt;goddess, icl&gt;god</td>
</tr>
<tr>
<td>Food</td>
<td>icl&gt;food, icl&gt;fruit</td>
</tr>
<tr>
<td>Flaura and Fauna</td>
<td>icl&gt;animal, icl&gt;reptile, icl&gt;mammal, icl&gt;plant</td>
</tr>
<tr>
<td>Boarding facility</td>
<td>icl&gt;facility</td>
</tr>
<tr>
<td>Transport facility</td>
<td>icl&gt;transport</td>
</tr>
<tr>
<td>Place</td>
<td>icl&gt;place, iof&gt;place, iof&gt;city, iof&gt;country</td>
</tr>
<tr>
<td>Distance</td>
<td>icl&gt;unit, icl&gt;number</td>
</tr>
</tbody>
</table>

Table 1: Semantics used for each templates

The terms obtained from the UNL dictionary will be a root word. For instance, the term, “eating” will be entered as eat (icl >action) in the UNL dictionary. So the terms obtained from the UNL dictionary needs to be generated to its original form using Morphological generator. The summary generation requires only tourism specific concepts, so the generation is almost not required. But we have used a morphological generator for Tamil, as the place information and distance information in Tamil with the case suffixes இ᾿ (il), இᾢᾞᾸᾐ (ilirunthu), உᾰᾁ (ukku) etc needs to be generated. For the example UNL graph shown in figure 3, the generated transport template in Tamil which is part of the summary is given in example 2.

Example 2: சென்னை (iof>city)

The concept chennai (iof>city) in the above graph, is generated as “சென்னை” by adding the case suffix “இ’” and the concept bus (icl>vehicle) is generated as “சுண்டும்” by adding the case suffix, “இ’”.

Figure 3: UNL graph given as input for example 2
5. Performance Evaluation

The proposed work has been tested with 33,000 Tamil and English text documents enconverted to UNL graphs. The performance of the methodology proposed has been evaluated using human judgement. The accuracy of the summary generated has achieved 90%. Apart from the summary generation factors such as tourism specific concept extraction, the accuracy also depends on the quality of enconversion and dictionary entry. By improving these factors, the accuracy can further be improved.

6. Conclusion and Future work

The proposed work generates a tourism specific bilingual summary using the intermediate document representation, UNL and tourism specific templates. The bilingual summary is generated in a simple and efficient manner compared to the earlier work done for multi lingual summary generation. The only over head involved is developing a enconverter framework.

As future enhancements, sentence patterns can be replaced by selecting the sentences having high sentence score based on its sentence position and the frequency of concepts. Query specific summary can also be generated on line, as the summary discussed here is a tourism specific generated off line using the templates. The evaluation of the generated summary can also be done by comparing it with the human generated summary. By doing this, many factors to make the machine generated summary compatible with human generated summary may evolve.

Reference

- Radev, Allison, Blair-Goldensohn et al (2004), MEAD – a platform for multidocument multilingual text summarization
- Prof. R. Nedunchelian, “Centroid Based Summarization of Multiple Documents Implemented using Timestamps” First International Conference on Emerging Trends in Engineering and Technology, IEEE 2008
Abstract

With the advent of computational tools for creativity, it becomes inevitable to design data structures which cater to the specific needs of the creative form considered. A lyric generator has to retrieve words fast based on Part of Speech and rhyme. This paper aims at building special indices for the LaaLaLaa Lyric Generator framework based on POS and rhyme to facilitate faster retrieval. The retrieval times of the proposed model and the conservative word indexed model are compared. The indexing is based on the KNM(Kuril, Nedil, Mei) pattern and the letters that occur in the rhyming spots of the words. The data structure is organized as hash tables to ensure best retrieval complexity. Separate hash tables for each POS and rhyming scheme are created and populated. Here, the key would be the meter pattern with the letters occurring at the rhyming spots and the value would be the list of all those words which fall under the key’s constraint. When the word indexed and meter rhyme indexed retrievals were compared, the latter reduced the average retrieval time drastically. There were not steep variations in the retrieval times as was in the former approach. This remarkable efficiency was traded-off with space.

1. Introduction

Tamil, one of the oldest languages, has a very rich literary history dating back to two thousand years. We have more than two thousand lyrics being written in this language in the form of film songs, advertisements, jingles, private albums etc. With the advent of computational tools for creativity, it becomes inevitable to design data structures which cater to the specific needs of the creative form considered. Tools for poem generation, story generation and lyric generation have been proposed.

A poem has to have three qualities – meaningfulness, poeticness and grammaticality [1]. A lyric is a poem which has constraints of having to satisfy a tune and a theme. This paper talks about building special indices for the LaaLaLaa lyric generator framework[2] to aid faster retrieval of words based on POS and rhyme.

A lyric generator would require the retrieval of words of a particular meter and particular letters at rhyming spots. This would maximize the poeticness of the lyric generated with the increase in rhymes. This retrieval process, when carried out on an un-indexed word database, is too expensive as it would take separate processing for meter, and each of the three rhymes in Tamil. To facilitate faster retrieval of words satisfying these constraints, the word database has to be indexed based on
meter-pattern and rhyme. This makes indexing of the word database based on the abovesaid constraints necessary.

This paper is organized as eight sections. The second section discusses about a few existing works in this area. The third section gives an overview of the Rhyme schemes in Tamil. An overall view of the system is given by the fourth section while the fifth section talks about the approach proposed for indexing. This work concludes with the results obtained and scope for future work in this area.

2. Background

Though significant number of works has been done in the arena of poetry generation in other languages, there is only less number in Tamil.

The “Automatic Generation of Tamil Lyrics for Melodies” [3] identifies the required syllable pattern for the lyric and passes this to a sentence generation module which generates meaningful phrases that match the pattern. This system generates rhyme based on maximum substring match and fails to make use of the three rhyming schemes that are specific to Tamil language. “LaaLaLa - A Tamil Lyric Analysis and Generation Framework” [2] generates Tamil lyrics for POS tagged pattern with words from a rhyme finder according to rhyming schemes in Tamil. Nichols et al[5] investigate the assumption that songwriters tend to align low-level features of a song’s text with musical features. K. Narayana Murthy[4] suggests having a non-dense TRIE index in main memory and a dense index file stored in secondary memory.

Anna Babarczy et al[6] suggested a hypothesis that a metaphoric sentence should include both source-domain and target-domain expressions. This assumption was tested relying on three different methods of selecting target-domain and source-domain expressions: a psycholinguistic word association method, a dictionary method and a corpus-based method. Hu, Downie and Ehmann[7] examine the role lyric text can play in improving audio music mood classification. Mahedero et al[8] argue that a textual analysis of a song can generate ground truth data that can be used to validate results from purely acoustic methods. Mayer et al[9] present a novel set of features developed for textual analysis of song lyrics, and combine them with and compare them to classical bag-of-words indexing approaches and results for musical genre classification on a test collection in order to demonstrate our analysis. “Semantic analysis of song lyrics” studies the use of song lyrics for automatic indexing of music. Netzer et al explore the usage of Word Association Norms (WANs) as an alternative lexical knowledge source to analyze linguistic computational creativity. Logan et al. use song lyrics for tracks by 399 artists to determine artist similarity[12].

3. Rhyme Schemes and Rhyme Patterns

**Rhyme Schemes:** English has number of rhyme effects like assonance, consonance, perfect, imperfect, masculine, feminine etc. This arises from the variation in stress patterns of words, lack of clear cut description about the spots where rhymes can occur.

In Tamil, the grapheme and phoneme are bound stronger than in English. There are 3 characteristic rhyme schemes in Tamil – Monai (மொனை), Edhugai (எட்டுகை) and Iyaibu (இயயை).

Two words are said to rhyme in monai if their first letters are the same, in edhugai if their second letters are the same and in iyaibu if their last letters are the same.
Examples: பறை and பᾲை rhyme in monai as they start with the same letter.
அஒединி and விரிசிப்பில் rhyme in edhucai as they share the same second letter.
பாககழை and பபபுகககζ rhyme in iyaibu as they share the same last letter.

As one may infer, two words can rhyme in more than one pattern also.

Examples: அஒединி and கூஒடி rhyme in edhucai and iyaibu.
கஒகஒககை and கஒகைககஒ rhyme in all the three schemes.

**Meter Pattern:** One way of classifying alphabets of the Tamil language is based on the time interval (மாஒகைதி - maathirai) for which they are pronounced. One maathirai corresponds to the time taken to wink the eyelid. The types of letters in this classification are

- **Nedil (N)** (ெநஒறி) - Those alphabets which are pronounced for the time interval of 2 maathirai.
- **Kuril (K)** (கூஒறி) - Alphabets which take 1 maathirai to be pronounced.
- **Mei (M)** (ெமஒறி) - Alphabets which are pronounced for 0.5 maathirai.

Meter pattern of a word refers to its Kuril Nedil Mei pattern.

For example, the meter pattern of the word மாஒனி is NKM as மா is a Nedil(N), ல is a Kuril(K) and ஊ is a Mei(M).

The indexing methodology proposed needs to take care of both meter pattern and rhyme to facilitate faster retrieval of words for the LaaLaLaa Lyric Generation and Analysis framework[2].

4. **Overview of the System**

![Diagram of the system](image)

*Figure 1. Overview of the system.*

Each word from the word database is converted to an object. Meter pattern and the alphabets at the Rhyming positions of each word are found out by Meter Pattern Extractor and Rhyme Extractor respectively. Using the data obtained, the Index builder builds the Rhyme Meter Index aiding faster retrieval than the normal un-indexed retrieval.
5. Indexing Algorithm

<table>
<thead>
<tr>
<th>Part of Speech</th>
<th>MeterPattern1</th>
<th>Letter1</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Letter2</td>
<td>Words</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MeterPattern2</td>
<td>Letter1</td>
<td>Words</td>
</tr>
<tr>
<td></td>
<td>Letter2</td>
<td>Words</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MeterPattern1</td>
<td>Letter1</td>
<td>Words</td>
</tr>
<tr>
<td></td>
<td>Letter2</td>
<td>Words</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MeterPattern2</td>
<td>Letter1</td>
<td>Words</td>
</tr>
<tr>
<td></td>
<td>Letter2</td>
<td>Words</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2. Indexing Logic**

This indexing has been designed to facilitate fast retrieval of words specifically for lyric generation. For instance, the system would need a word of a particular meter pattern with a particular letter rhyming in monai scheme.

The data structure is organized as hash tables with separate tables for each Part Of Speech and Rhyming Scheme. For example, there is a separate table for Nouns’ monai, Nouns’ edhugai, Nouns’ iyaibu and so on. Here, the keys are the Meter Pattern and the Letter at the particular Rhyming spot. The values are the words corresponding to the particular Meter pattern and letter.

**The algorithm**: The word database is scanned word by word and is indexed based on meter pattern and rhyme. Here,

\[\alpha \leftarrow \text{number of words in the word database};\]
\[w_i \leftarrow \text{ith word in the word database};\]
\[\beta \leftarrow \text{meter pattern of } w_i;\]
\[\text{monaiKey} \leftarrow \text{key of } w_i \text{ corresponding to monai};\]
\[\text{edhugaiKey} \leftarrow \text{key of } w_i \text{ corresponding to edhugai};\]
\[\text{iyaybuKey} \leftarrow \text{lkey of } w_i \text{ corresponding to iyaybu};\]

\[\text{begin}\]
\[\forall j = 1, 2, 3 \ldots \text{ do}\]
\[\beta \leftarrow \text{MeterPattern of } w_j;\]
\[\text{monaiKey} \leftarrow \text{firstLetter (} w_j \text{) } + \beta;\]
\[\text{edhugaiKey} \leftarrow \text{secondLetter (} w_j \text{) } + \beta;\]
\[\text{iyaybuKey} \leftarrow \text{lastLetter (} w_j \text{) } + \beta;\]
\[\text{Add the word to the list of words with the respective keys in the respective Hashtables.}\]
\[\text{end};\]
For each word in the word database, meter pattern is extracted first followed by letters at the rhyming spots namely, first, second and last positions (for monai, edhugai and iyaibu respectively). Keys for monai, edhugai and iyaibu are found out using the abovesaid equations. The word is added to the monai, edhugai and iyaibu hashtables with keys monaiKey, edhugaiKey and iyaibuKey respectively if those keys don’t appear previously in the tables. Else, the word is added to the list of words with that key.

Hash-tables are chosen for the implementation as they have the retrieval complexity of O(1).

6. Results

![Figure 3. Word Indexed Vs Meter Rhyme Indexed Approach](image)

Indexing has brought about a drastic increase in the speed of retrieval of the words rhyming with a given word. Using the Word Indexed approach, the time complexity was O(α) where α was the total number of words. But after Meter Rhyme indexing, the complexity has become O(1) due to the use of hash table which is a very efficient data structure for retrieval. Rhyming words for a sample of 500 words were retrieved using both the approaches and the above mentioned graph was obtained. The Word indexed system took 875.47 millisecond in an average while the Meter Indexed system took 1.90 millisecond only. From the graph it can also be inferred that there are steep variations in the Word-Indexed approach while the Meter-Rhyme Indexed approach does not show such steep variations and is consistent. In terms of time efficiency, Meter-Rhyme indexed approach is evidently superior compared to Word-Indexed approach. In terms of space, it is not so efficient as each word will occur not once, but nine times in various Hash tables.

7. References


Tamil Document Summarization Using Latent Dirichlet Allocation

N. Shreeya Sowmya\textsuperscript{1}, T. Mala\textsuperscript{2}
\textsuperscript{1}Department of Computer Science and Engineering, Anna University
\textsuperscript{2}Department of Information Science and Technology, Anna University
Guindy, Chennai
\textsuperscript{1}shreeya.mel@gmail.com \hspace{1cm} \textsuperscript{2}malanehru@annauniv.edu

Abstract
This paper proposes a summarization system for summarizing multiple tamil documents. This system utilizes a combination of statistical, semantic and heuristic methods to extract key sentences from multiple documents thereby eliminating redundancies, and maintaining the coherency of the selected sentences to generate the summary. In this paper, Latent Dirichlet Allocation (LDA) is used for topic modeling, which works on the idea of breaking down the collection of documents (i.e) clusters into topics; each cluster represented as a mixture of topics, has a probability distribution representing the importance of the topic for that cluster. The topics in turn are represented as a mixture of words, with a probability distribution representing the importance of the word for that topic. After redundancy elimination and sentence ordering, summary is generated in different perspectives based on the query.

Keywords- Latent Dirichlet Allocation, Topic modeling

I. Introduction
As more and more information is available on the web, the retrieval of too many documents, especially news articles, becomes a big problem for users. Multi-document summarization system not only shortens the source texts, but presents information organized around the key aspects. In multi-document summarization system, the objective is to generate a summary from multiple documents for a given query. In this paper, summary is generated from the multi-documents for a given query in different perspectives. In order to generate a meaningful summary, sentences analysis, and relevance analysis are included. Sentence analysis includes tagging of each document with keywords, named-entity and date. Relevance analysis calculates the similarity between the query and the sentences in the document set. In this paper, topic modeling is done for the query topics by modifying the Latent Dirichlet Allocation and finally generating the summary in different perspectives.

The rest of the paper is organized as follows. Section 2 discusses with the literature survey and the related work in multi-document summarization. Section 3 presents the overview of system design. Section 4 lists out the modules along with the algorithm. Section 5 shows the performance evaluation. Section 6 is about the conclusion and future work.
II. Literature Survey

Summarization approaches can be broadly divided into extractive and abstractive. A commonly used approach namely extractive approach was statistics-based sentence extraction. Statistical and linguistic features used in sentence extraction include frequent keywords, title keywords, cue phrases, sentence position, sentence length, and so on [3]. Cohesive links such as lexical chain, co-reference and word co-occurrence are also used to extract internally linked sentences and thus increase the cohesion of the summaries [2, 3]. Though extractive approaches are easy to implement, the drawback is that the resulting summaries often contain redundancy and lack cohesion and coherence. Maximal Marginal Relevance (MMR) metric [4] was used to minimize the redundancy and maximize the diversity among the extracted text passages (i.e. phrases, sentences, segments, or paragraphs).

There are several approaches used for summarizing multiple news articles. The main approaches include sentence extraction, template-based information extraction, and identification of similarities and differences among documents. Fisher et al [6] have used a range of word distribution statistics as features for supervised approach. In [5], qLDA model is used to simultaneously model the documents and the query. And based on the modeling results, they proposed an affinity propagation to automatically identify the key sentences from documents.

III. System Design

The overall system architecture is shown in the Fig. 1. The inputs to the multi-document summarization system are multi-documents which are crawled based on the urls given and the output given by the system is a summary of multiple documents.

![Fig. 1 System Overview](image)

System Description

The description of each of the step is discussed in the following sections. The architecture of our system is as shown in Fig. 1.
1. Pre-processing

Pre-processing of documents involves removal of stop words and calculation of Term Frequency-Inverse Document Frequency. Each document is represented as feature vector, (ie.,) terms followed by the frequency. As shown in Fig. 1, the multi-documents are given as input for pre-processing, the documents are tokenized and the stop words are removed by having stop-word lists in a file. The relative importance of the word in the document is given by

\[ \text{Tfidf}(w) = \text{tf}(w) \times (\log(N)/\text{df}(w)) \]  

where, \( \text{tf}(w) \) – Term frequency (no. of word occurrences in a document)  
\( \text{df}(w) \) – Document frequency (no. of documents containing the word)  
\( N \) – No. of all documents

2. Document clustering

The pre-processed documents are given as input for clustering. By applying the k-means algorithm, the documents are clustered for the given k-value, and the output is the cluster of documents containing the clusters like cricket, football, tennis, etc., if the documents are taken from the sports domain.

3. Topic modeling

Topic models provide a simple way to analyze large volumes of unlabeled text. A "topic" consists of a cluster of words that frequently occur together. In this paper, Latent Dirichlet Allocation is used for discovering topics that occur in the document set. Basic Idea- Documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words.

Sentence analysis

Multi-documents are split into sentences for analysis. It involves tagging of documents by extracting the keywords, named-entities and the date for each document. Summary generation in different perspectives can be done from the tagged document.

Query and Relevance analysis

The semantics of the query is found using Tamil Word Net. The relevant documents for the given query are retrieved. The relevance between the sentences and the query is calculated by measuring their similarity.

Query-oriented Topic modeling

In this paper, both topic modeling and entity modeling is combined [3]. Based on the query, the topic modeling is done by using Latent Dirichlet Allocation (LDA) algorithm. Query is given as prior to the LDA and hence topic modeling is done along with the query terms. Query may be topic or named-entity along with date i.e. certain period of time.
4. Sentence scoring

The relevant sentences are scored based on the topic modeling. For each cluster, the sum of the word’s score on each topic is calculated, the sentence with the word/topic of high probability are scored higher. This is done by using the cluster-topic distribution and the topic-word distribution which is the result of the Latent Dirichlet Allocation.

5. Summary generation

Summary generation involves the following two steps

5.1 Redundancy elimination

The sentences which are redundant are eliminated by using Maximal Marginal Relevance (MMR) technique. The use of MMR model is to have high relevance of the summary to the document topic, while keeping redundancy in the summary low.

5.2 Sentence ranking and ordering

Sentence ranking is done based on the score from the results of topic modeling. Coherence of the summary is obtained by ordering the information in different documents. Ordering is done based on the temporal data i.e. by the document id and the order in which the sentences occur in the document set.

IV. Results

Table 1 shows the topic distribution with number of topics as 5, the distribution includes the word, count, probability and z value. The topic distribution is for each cluster.

Table 1 Topic model with number of topics as 5

<table>
<thead>
<tr>
<th>TOPIC 0 (total count=1061)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORD ID</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>645</td>
</tr>
<tr>
<td>2806</td>
</tr>
<tr>
<td>2134</td>
</tr>
<tr>
<td>589</td>
</tr>
<tr>
<td>1417</td>
</tr>
<tr>
<td>2371</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>

V. Conclusion and Future Work

In this paper, a system is proposed to generate summary for a query from the multi-documents using Latent Dirichlet Allocation. The multi-documents are pre-processed, clustered using k-means
algorithm. Topic modeling is done by using Latent Dirichlet Allocation. The relevant sentences are retrieved according to the query, by finding the similarity between the sentences and the query. Sentences are scored based on the topic modeling. Redundancy removal is done using MMR approach.

Topic modeling can be extended to find the relationship between the entities, i.e. the topics associated with the entity as a future work.

References

Language Ideology, Inscriptions, Spoken Tamil and Technology

(இயற்பியல் விசாரணை, கல்வியமைப்பு களிப்பு,
சொகுத் தமிழு - இயற்பியல் தமிழகத்தில் பயன்பாடு)
Mapping Language Change in Tamil:  
Corpus analysis and Computer Database Making

Appasamy Murugaiyan  
EPHE- UMR 7528 Mondes iranien et indien, Paris

Introduction
In this paper, I would like to illustrate how ‘corpus analysis’ can help us mapping the process of language change and language use. The vast Tamil language corpora, dating more than two millennia, make it very envious to any outsider. One of the most challenging debates that we all have been witnessing in the last fifty years or so in the Tamil milieu have to do with Tamil LANGUAGE. However, so many questions arise related to, the language of Tamil Sangam anthologies, language of inscriptions, language of bhakti literature and then the emergence of modern Tamil. What is the relationship between these different varieties of Tamil? How they distinguish from each other? How to analyse and understand the grammatical structure of each of the genre of Tamil varieties- are some of the issues that deserve a definitive answer. The problem is that most of the studies that have been done so far give a vague idea mostly based on what may be called a ‘native speakers’ intuition. Yet, a corpus based empirical picture is due to be made.

1. Corpus analysis and Computational methodology in Tamil:
Mapping language change has been a major concern of corpus linguists. Historical corpora provide evidences for language change in many ways. Such a corpus-based study would not only enrich the history of Tamil language but also would contribute to the theoretical models for language change. I will illustrate the importance of the corpora based empirical studies in mapping language change and use in Tamil primarily based on few citations (Murugaiyan 1993, 2004 and 2011).

The field of corpus linguistics has proved beyond doubt that the corpus is a fundamental tool for any type of research on language and particularly if we want to find answers to some relevant issues on the language use and language change. At the outset, it is important to make a distinction between computational methodology for linguistic analysis and computational linguistics (CL). CL is concerned by fields like Artificial Intelligence and aims at developing formal models based on aspects of human cognition and implement them as computer programmes. Our concern here is how computer can be used in linguistic analysis on a specific Tamil corpus through a series of research questions. Computer-based methodology allows us 1) to work on vast corpora, which would otherwise be impossible and 2) to seek answers to many questions on different diachronic and synchronic aspects of Tamil linguistics. In many instances, CL relies on the results of corpus linguistic analysis using computer methodologies.

The paper is structured as follows: §2 introduces four / three questions from Tamil linguistics, §3 surveys the currently existing and or used POS Tag for (Indian languages) Tamil and raises in §4 the
question of structure of the database and the type of corpus and §5 concludes on a Tamil inscription database.

2. Few issues in Tamil linguistics:

I will illustrate the importance of the corpora based empirical studies in mapping language change and use in Tamil primarily based on few citations (Murugaiyan 1994, 2004 and 2011). Of which the first two have to do with the historical linguistics of Tamil and the last two with the modern Tamil linguistics.

2.1. Word order variation in inscriptive Tamil

Languages vary widely in many ways, including their canonical word order. It is known that some word orders are much more common than others are. Change or variation in word order type is one of the most important areas in the study of historical linguistics and language change.

We can roughly identify two different views on the word order in Old Tamil: (1) SOV is considered as the basic order and 2) view suggests a free word order. However, scholars have overtly recognized variation in the SOV order but have not made a detailed attempt to discuss this variation (Zvelebil 1967.71, 1997.43), except Herring (Herring 2001). Nevertheless, a corpus based empirical analysis of inscriptive Tamil, conducted by me, shows a third possibility: the constituent order is neither free nor strictly of SOV type and the variation in constituent order is motivated by pragmatic factors (Murugaiyan A 2011). In other words, the position of different constituents in a sentence is conditioned by information structure and many other contextual factors. In such cases, the word order is not used to encode grammatical relations within a sentence (like -subject vs. object-). For this pilot survey, I analysed a total number of 35 Hero stone inscriptions dating from 450 to 650 CE.

2.2. Experiencer (or Dative subject) constructions in Sangam corpus

In most of the Modern South Asian languages, the verbs of physical, psychological and cognitive processes, known also as affective verbs, mark the principal actant (experiencer) in dative, accusative or genitive cases. This type of constructions known as dative subject constructions (or oblique-experiencer) since 1960’s is considered as a major feature of the Indian linguistic area (Masica C. 1976). During the last three decades, there have been numerous studies on this major areal feature, which is shared by Indo Aryan, Dravidian, Munda and Tibeto-Burman family of languages (Verma and Mohanan 1990, Shibatani and Pardeshi 2001, Bhaskararao Peri, 2001).

Despite the great attention paid to experiencer constructions, we are yet to know about their origin and their diachronic developments in the different language families of South Asia. It is still commonly suggested that it has spread from the Dravidian family of languages. In order to bring in new data on the historical development of experiencer constructions in the South Asian typological and areal linguistics, I worked on a small corpus of Sangam literature. My survey showed that the dative experiencer construction is absent in classical Tamil (Murugaiyan, A. 2004) even though it is widely attested in modern Tamil (yāṉ viyarttaṉaṉ in Sangam, compared to eṇakku viyarkkiatu in modern Tamil).
2.3. Non-dative dative subjects

We notice in Modern Tamil, and in some of the Dravidian languages, the experiencer or the dative subject is not always marked in dative case. Some notions like *virumba* ‘to desire’, *porāmai* ‘(be) jealous’ depict constructions that range from a typical nominative-accusative structure (N1_{Nom} - N2_{Acc}) to nominative-oblique (N1_{Nom} - N2_{Obl}) structure. This type of Nominative and Accusative case marking represents higher degree of transitivity and is characteristic of agent role but not that of experiencer.

The major question here is how to define or identify a “(proto) typical experiencer” (for instance, absence of volition – control - contact with the second participant (effectiveness) and capable of receiving physical, psychological, cognitive experiences). Experiencer is a scalar notion like transitivity or agent. According to the syntactic structure of the construction, it seems possible to identify different kinds of experiencer like “agent-like”, “patient-like” and “typical experiencer”. The morphological case marking of the first and second participants (arguments) of these constructions correlates to some extent with the ‘variation’ in the semantic role of the experiencer.

2.4. Differential Object Marking and the accusative case in Tamil

Tamil is a morphologically rich language and many of the major grammatical functions are marked by case markers. For example, in Tamil the direct object is marked in accusative case. In general, it is held both by the traditional and modern Tamil grammars that the accusative marker “ai” is present obligatorily with animate nouns and is not obligatory with inanimate nouns. The case marking of direct objects is typologically one of the major concerns in the theoretical setting of Differential Object Marking (DOM) and goes beyond the animate / inanimate dichotomy. A number of studies on Indian languages have shown that several parameters like humanness, definiteness, individuation and affectedness of the noun on the one hand and the semantics of the verb on the other, contribute to the marking or non-marking of the direct objects (A. Murugaiyan 1993).

The four cases mentioned above raise few interesting questions. How these morphosyntactic, syntactic and semantic features in Tamil can be successfully encoded in a computer-aided corpora analysis? What would be the adequate standard of data annotation for Tamil? The point to be remembered here is that our objective is to map language change and use but not machine learning.

3. Parts-of-Speech Tags for Tamil and Indic languages

Over the past decades, large number of electronic language data have been created and annotated in the area of NLP for Indian languages. The POS Tag set and the construction of database are closely related to the objectives fixed by the researcher.

Several IL POS (Indian languages Parts of Speech) tagsets are often designed by a number of research groups working on Indian languages. The existing tagsets differ considerably from one another as they have been motivated by specific research agenda. They follow in general the PENN Tree Bank model and or the recommendations made by EAGLES. The number of tagsets varies from 34 up to 64 and 123 (Evaluating SKT tagsets). They differ considerably in terms of the selection of morphosyntactic categories and the scale of granularity. These different works, in course of time, have led to the creation of standard POS Tag set for Indian languages based on the EAGLES model. The proposal made by Baskaran et al. (2008) aims at fixing standard coding frame for Indian languages in
general. This model is expected to “capturing the shared linguistic structures in a methodical way [...] and ensure cross-linguistic compatibility”. This is no doubt part of computational linguistics focusing on machine learning.

This being so, in computer aided linguistic analysis we face different challenges. In most cases, we need language specific, fine-grained multi level frame. The picture becomes still more complex while we study a specific aspect of language change and use (cf. §2) on a large corpus. We need to decide, among other things, setting up various types of corpora: synchronic and diachronic on the on hand and according to the type of literature under scrutiny (for example Sangam anthology, Bhakti literature, Tamil inscriptions, Modern Tamil and so on).

4. Linguistic corpora and Database Making

If we wish to answer empirically to questions raised in §2 above and similar others, the corpus based study is a fundamental requisite. The first phase of corpus creation is data entry, which involves rendering the text in electronic form. The Sangam corpus is electronically available in standard format. On the contrary, the inscriptive corpus is not available electronically and the printed versions cannot be digitized via OCR. We still need to standardize the complex orthographic rules and the different scripts used in the inscriptions. The whole corpus of Tamil inscriptions has to be captured manually.

The corpus annotation as we have seen early is language oriented and should reflect the type of questions we are addressing in the corpora. Now turning back to the sample questions raised in §2, it is important to know precisely what information should we need to encode in a linguistic database? Should a coarse tag set be enough or should we need detailed set of information at morphosyntactic, syntactic, semantic and discourse levels.

For instance, to account for the word order variation and its correlation with pragmatics or discourse structure, one should include pragmatic information in the corpus annotation and the type of word order for each ‘sentence’.

As for experiencer (dative subject) constructions in Classical Tamil, a large variety of lexical and grammatical devices are used to express physical, biological, cognitive and mental perceptions in classical Tamil. In the following examples, we have a compound verb composed of a noun and support or light verb.

\( \text{pasi} + \text{paṉu} \) (hunger+endure) \hspace{1cm} \text{feel hungry} \ [\text{puṇ.260.6}]

\( \text{pasi} + \text{kūra} \) (hunger+grow) \hspace{1cm} \text{hungry} \hspace{0.5cm} \text{grow} \ [\text{naṇ.29.3}]

\( \text{pasi} + \text{teṉu} \) (hunger+burn) \hspace{1cm} \text{Hungry} \hspace{0.5cm} \text{burn} \ [\text{aka.291.3}]

\( \text{añar uṉa} \) (suffering+to have) \hspace{1cm} \text{to suffer} \ [\text{kuṇ.76.6}]

The database should on the one hand provide all lexical bases, both noun and verb, and on the other hand all light verbs and the syntactic and semantic rules of combination that licence the instantiation of the predication. A large historical corpus should be able to trace the evolution from ‘nominative’ to ‘dative’ structure.

The case marking variation noticed in dative subject constructions in modern Tamil is due to the lexical and semantic nature of the experiencer predicates:
1) nāṉ Sandōśam paṇugīčē 'I am happy'
   I joy feel.pres.1s
2) *nāṉ magicci paṇugīčē 'I am happy' (not accepted)
   I joy feel.pres.1s
3) vīraṉ semmāttiyaṉ virumbupikiṉāṉ
   pn.nom pn.acc like.pres.3m.s
   Viran likes Semmathi
4) vīraṉukku semmāttimīdu viruppam
   pn.dat pn.loc liking
   Viran like Semmathi

In (1) & (2) the two lexical items with almost identical meaning do not fall under the word formation rule. In (1) the noun sandōśam is an IndowAryan loan word. In (3) and (4), the predicates are of two different grammatical categories. The finite verbal predicate in (3) marks the direct object in accusative case and shows agreement with experiencer. On the contrary, in (4) with a nominal predicate, the direct object is marked in locative and the experiencer is marked in dative case.

These predicates, nominal and verbal, are divided into three groups: cognitive, psychological and physiological. There is a close correlation between the experiencer construction and, among other things, (1) the grammatical category of the predicate and (2) the semantic nature of the predicate. Even though we notice some generalisation between the semantic role and the morphological case marking, this relation is a language specific phenomenon. The point here is that at what level we incorporate the various lexico-semantic features in the corpus annotation.

Finally, the differential object marking is noticed in Tamil and in many other languages. In order to account for the marking or non-marking of the direct object in accusative case we have to distinguish several features: human / non-human, definite / indefinite, referential / non-referential, degree of affectedness of the direct object on the one hand and the semantics of the verb on the other. In examples (5) and (6) the mere presence or absence of the accusative case change completely the meaning. In (6), the object, ‘loan’ known as object of creation, exists prior to the utterance of the sentence. On the contrary, in (5), the speaker is trying to contract a loan, and it does not exist really in the discourse context.

5) avariṉam kaṉaṉ kēṉṉēn saṉṉaikku vanduvittār
   he.locative loan ask.past.1.s quarrel.dative come.adp.aux.past.3.s
   I requested him to lend me some money, but he started to pick on me

6) avariṉam kaṉaṉai kēṉṉēn saṉṉaikku vanduvittār
   he.locative loan.acc ask.past.1.s quarrel.dative come.adp.aux.past.3.s
   I asked him to repay my money back, but he started to pick on me

5. Conclusion

As mentioned earlier, nature of database and the inventory of tagsets depend on the type of research agenda. This type of multi-layered analysis on large corpora is possible only with computer-aided methodology. The four cases mentioned in §2 and the few examples given in §4 are a few among
hundreds of other questions that we have to account for in Tamil. Most of the existing POS tag sets do not provide us with a fine-grained annotation scheme. But in compute-aided corpora-based linguistic research, a fine-grained annotation paradigm is essential. In my experimental database on Tamil Inscriptions, I have about 120 tagsets. Three types of rule-based annotations—morphosyntactic, syntactic and semantic—are done manually. I have aimed at a fine-grained analysis and so I have opted for a high number of tags. I have also included information like word order types, verbal valency and other minute details that would help to map different changes at morphological, syntactic and semantic levels. I am using, for each sentence the interlinear glossed text (IGT) format, which includes source language text, a morpheme-by-morpheme gloss, and a translation into French or English.

The ongoing research is to illustrate how linguistic corpora can be used as readily available evidence for mapping language development and language variation in time and space. A huge computerized historical corpus would certainly allow a comparative view of the Tamil language at different moments in the history. These data would help us not only to capture different stages of linguistic developments but will also help to test modern theories about variation and change. The construction of huge electronic corpora in Tamil presents many constraints related to linguistic theories. However, a close collaboration with computational specialists would certainly help to overcome many of the shortcomings and certainly would enhance computational methodology for corpus-based linguistic analysis in Tamil.

References

- Murugaiyan, A (Forthcoming) Identifying Basic Constituent Order in Old Tamil: Issues in historical linguistics with Special Reference to Tamil Epigraphic texts (400-650 CE), Proceedings of the World Classical Tamil Conference, 2010 July, India.
Language Ideology and Technology

E. Annamalai
The University of Chicago

Technology, which is a device that increases the ratio of output to the input, works for the language also as it does for any human production. The first technology applied to language was writing invented around 3500 BCE in Mesopotamia (modern Iraq). It made possible to enormously increase the reach of the language -i.e. the content in the language- in space and time. It also provided relative stability to language, which reduced variation. These added characteristics of the written language were necessary for trade, which was the cultural context for the emergence of writing. The idea of the language changed from being a tool of cooperative communication to that of record keeping to eliminate mistrust. From this, the ideology emerged that the written word is more trustworthy than the spoken.

The next major technology is the invention of woodblock printing in China developed to print language on paper around 650 CE. Its improvement into movable metal types around 1440 CE in Germany made copying of printed texts possible at a low cost and in less time, which in turn made their reach increase exponentially in space, physical and social. This ushered in the era of print capitalism (Anderson 1991), one aspect of which was production and dissemination of language materials as commodities of the market. The printed texts became the private property of printing establishments or individual authors. The content and the particular form of the language in which it is coded came to be owned by the producers of it while the abstract language remained the public domain. When the written language through print lent itself to be controlled by the institutions of the society such as school, media, courts, it was possible to shape the language according to their ideologies such as purism, precision (as in law or science).

When the audio recording and replay technology was improved to use electrical and magnetic devices in the early decades of the 20th century, the added characteristics of the written and printed language could be transferred to the spoken language. This did not bring about any new ideology to the language except that the ideologies such as purism and precision were difficult to carry out in the spoken language, as it was not controlled by societal institutions through their system of rewarding the adherents of their language ideology.

The latest in technology that bears on language is the digital technology, which equally applies to the spoken and the written language. While the earlier technologies moved from their use for non-linguistic content such as pictures and music to language, the digital technology moved from numbers to language. While writing and printing technologies represent the language unit, viz. the letters, directly, the digital technology converts language units, viz. letters and phones, into digits for processing. Digital technology gives enormous scope for editing while composing. The implication of this is that the characteristic of finality and permanence, which the earlier technologies gave to texts, turns out to be fragile in this technology. Digital technology is useful not only for composing new
language content, but also for copying and storing the old. It reduces drastically the time and cost for transmitting the materials globally. It takes away the control of language from social institutions and gives it back to individuals.

What this new technology is doing and can do to Tamil? The writing technology gave new documentation register and literary code to Tamil, as seen in the emergence of Sangam literature and contemporaneous Tamil Brahmi inscriptions of record. The printing technology stabilized the Tamil alphabet both in their graphic form (which was evolving over 2000 years) and number (the five grantha letters were added and three contextually dependent letters (caarpezuttu) were dropped with the advent of this technology). The digital technology has brought, and is capable of bringing more, new effects on Tamil. Decontrolling of Tamil brings the written Tamil closer to the spoken Tamil with regard to the effects and gives legitimacy to this convergence. The societal institutions lose their commanding role in shaping Tamil. As anyone with a vocal cord could speak at will, any one with literacy skill and access to digital technology could write at will. Anyone can be a writer without vetting by a teacher or an editor. The spontaneous writing could be more effective on language than spontaneous speaking because the language as written at will by any and all individuals gets into the public domain accessible to any, unlike the language spoken.

It follows that there will be more language ideologies at play is shaping Tamil, which are subscribed to by the individuals; they will not be just the ones promoted, and penalized for non-compliance, by the elite in control of the societal institutions. Purism ideology is also prevalent among the individual practitioners of the new technology. Purism of Tamil includes elimination of loan words and acquired letters, avoidance of spoken forms in vocabulary and grammar and, to a lesser extent, reclaim of historically antedated grammatical constructions. This ideology is also enforced using the same digital technology on the writings of others with a different ideology in communally created content such as Wikipedia. Nevertheless, the multiplicity of ideologies in shaping Tamil cannot be excommunicated from this technology.

Any technology is not in itself and by itself an aid to modernize a language. It depends on its user, who decides what it is used for. The print technology helped not only the use of prose as a language of literature, but also helped recoding of the oral literature in the visual medium on paper. The books of folk literature printed exceeded numerically the books of modern fiction and poetry (Blackburn 2005) when print technology came into being in Tamil Nadu. So is the religious literature compared to the secular literature. The digital technology used for astrological predictions is in demand as it is for weather predictions. While technology cannot be appropriated for what is valued as modern, its potential for this task should not be wasted away.

An ideology that digital technology is capable of implementing on Tamil is parity of written and spoken Tamil and reduction of distance between them. I shall not go into discussing here the rationale for this ideology and its importance for the survival of Tamil as a modern language with vitality (Annamalai 2011a). There are many areas to implement this ideology. I shall mention some of them, many of which relate to teaching Tamil to learners of different backgrounds as to their exposure to Tamil before, during and after learning.
Inclusion of spoken Tamil in Tamil pedagogy is on the increase. It is an important part in the Tamil teaching programs designed for non-native adults outside Tamil Nadu and Jaffna, especially in the Universities in the U.S., where the goal of Tamil learning is as much oral communication as it is philological and literary inquiry. This extends to children of Tamil ethnicity, who have lost their heritage language and want to revive it, as in Mauritius. Their interest is communicative within the community as well as cultural and political for reasons of identity. The traditional Tamil language curriculum in schools under the rubric of learning the mother tongue as a minority in countries like Singapore focuses on literacy skills and introduction to literature. This is being modified in Singapore to include speaking of Tamil within the curriculum in order to make Tamil more relevant for students in their lives (Seethalakshmi). Like these students, the second generation Tamils of post-colonial diaspora has limited listening comprehension in Tamil and they want to add spoken skill to their Tamil competence. Even in a curriculum that focuses on the reading skill, reading modern fiction and magazines will be hard without the knowledge of spoken Tamil, where conversations between characters, jokes etc are written in this variety.

The basic need when spoken Tamil becomes part of the Tamil curriculum is identification of standard spoken Tamil, which is spoken in inter-dialect communicative situations by the schooled. To identify it empirically, we need a searchable database of spoken Tamil. Such a database could be built relatively easily using digital technology from dialogues in movies and television shows. This database is a necessary, if not sufficient, tool to compile the grammatical, lexical, semantic and phonetic parameters of the standard spoken Tamil. The data needs to be processed to sift out dialect and formal features, which are mixed up in the standard speech as well as the mixing of English in it; the phonetic data needs to be brushed up to upgrade the non-standard pronunciation found commonly in the public programs used to build the database. It should be possible to write algorithms to do these jobs mechanically. Pedagogy requires not speech as it, but as standardized. I shall not go here into discussing what the standardized spoken Tamil is (Annamalai 2011b).

The standard spoken Tamil data needs to be transcribed in Tamil script. The publically available Google tool for speech recognition and instant transcription needs to be improved substantially for Tamil. When this is developed, it should be possible to go into making popular, inexpensive tools that instantly convert the utterance of a student into a written sentence, and conversely a written sentence of spoken Tamil into an utterance. This will help the student recognize speech visually and aurally, which facilitates the recall of the language in the learning situation.

Converting speech into writing assumes a standard spelling system of the standard spoken Tamil. This is yet to be developed by linguists to be used in the pedagogical context to begin with. A fundamental principle of any spelling system is that it is not an exact phonetic transcription of speech, but a convention of writing from which speaking could be deduced with straightforward rules of correspondence. The spelling system of every language has such rules. In the case of Tamil, an additional requirement is that the spelling system of the spoken Tamil also serves as the base for the student to migrate to the conventional spelling of the written Tamil with straightforward rules. No such spelling system exists now. The way the spoken Tamil written is notoriously inconsistent between authors and in the same author at different places. The spelling system used by on Tamil teacher is consistent but it is not the same across teachers, though each is relatable. The spelling
system for spoken Tamil needs to be standardized urgently. I have developed a system as the starting point to initiate this process (Annamalai 2009). If the standardized spelling necessitates some additional letters in the Tamil alphabet, we need to get cultural acceptance of it. Then they will need to be provided in Unicode.

The other need is to relate the spoken and written Tamil at the lexical level with regard to the spelling of words. This should be bidirectional from the spoken to the written and from the written to the spoken. This tool will reduce the distance between the two by facilitating migration between the two. There are now tables available to relate the written to the spoken with general rules of deletion and change. They need to be fine tuned to cover complex relations, variations and exceptions and then algorithms are to be written so that one can get the spoken form from the written on any device, handheld or desktop, just like checking the meaning of a word. The converse of relating the spoken with the written is more difficult primarily because of the fact that one form of the spoken will relate to two of the written. A solution to this problem has to be found such as using the absence of one of the two possible forms in a built-in dictionary to reject it.

There is no dictionary of standard spoken Tamil available now. It is possible to compile one using its standardized spelling for the head entries followed by the written word in conventional spelling with gloss in English for learners who are not proficient in Tamil. The converse of it would be to have parallel head entries first in conventional spelling followed by spoken spelling. It is possible to produce mechanically one version of the dictionary from the other and rearrange the entries alphabetically. An alternative to this has the additional advantage of using the spoken input when referring to a dictionary. In this, a spoken word keyed in in its spelling will first identify its written equivalent, which will lead to its meaning in the digital dictionary. It should also be possible to speak the word a student, who is not competent in writing Tamil, hears an unknown word when watching a video or touring Tamil Nadu and wants to know its meaning. A digital dictionary should be able to point out the meaning either directly from the spoken cue or through the written word coded in the dictionary by automatically linking the spoken word to the written form.

Technology is available to do the above things if the language ideology with regard to spoken Tamil mentioned above is a driving force. Using digital technology for the spoken Tamil is not just for doing research on it. It is also for teaching Tamil effectively and attractively. The hope is that it will attract more students to learn Tamil, who are now put off because of the perceived difficulty of learning separately the two varieties of Tamil. It is the desire of all of us to mitigate the difficulties of learning Tamil by the younger generation of Tamils and also the non-native speakers of Tamil.

References


Blackburn, Stuart. 2005. Print, Folklore and Nationalism in Colonial South India. New Delhi: Permanent Black

Seethalakshmi and Vanitamani Saravanan. 2009. தரமான பார்க்கும் ஆசிய வியான். Singapore: Author (at National Institute of Education)
Tamil: A Family of Languages

Harold Schiffman
South Asia Studies. University of Pennsylvania

Abstract

As is well-known, the Tamil language is one of the oldest languages on the Indian subcontinent, dating from the early centuries of the Common Era, if not before. It is commonly divided into a number of ‘stages’, beginning with the earliest period (Sangam Tamil), followed by Medieval Tamil (or Middle Tamil), and then the Modern Literary Language, which dates from about the 13th century. I would also add a fourth stage, that represented by the spoken language of today, which differs sometimes quite radically from its written form. This presents a formidable challenge to non-Tamils who wish to learn to both read and speak Tamil, since Tamil society offers little help to those wishing to speak, even though most authentic communication between live speakers goes on in Spoken Tamil, and learners who wish to learn something about Tamil culture will not get far without a knowledge of the spoken language. Ignoring the modern spoken language also hides the tremendous diversity among dialects of Tamil, especially those that differ radically from what I call “Standard Spoken Tamil” or SST (Schiffman 1998). Sri Lanka Tamil is one of those dialects that are not mutually intelligible to many other speakers. For this reason, I propose that we should cease treating Tamil as one language, and begin to think of Tamil as a family of languages, related of course through history, from the oldest stages to the most modern. In historical linguistics terms, we would treat the oldest stage as Proto-Tamil, and later stages as ‘daughter languages’ or even ‘granddaughters’.

The Accessibility Problem

Foreigners who wish to learn Tamil are confronted with enormous challenges. Tamil culture tends to value the study of Classical Tamil and its ‘daughter’ languages (Medieval Tamil, modern Literary Tamil) but not its ‘grand-daughters’ i.e. the spoken dialects used by all Tamils for most of their interpersonal communication. Foreigners who attempt to learn spoken Tamil are discouraged from doing so in various ways:

- * scolding* the learner for ‘corrupting’ the language
- ‘correcting’ the spoken form by repeating the LT form
- Ridiculing the learner by laughing at him/her for using spoken forms

As an example of the first strategy, when I was doing research on spoken Tamil in my first visit to India in 1965-66, students who were influenced by DMK came to me and asked me to cease and desist from studying spoken Tamil, because it ‘contributed to the downgrading’ of the language. An example of the third type, ridicule, happened to me while passing through customs from Singapore into Malaysia—I spoke Tamil to the customs agent, who had a Tamil name and looked to be of Indian descent. Her response was “You talk just like my Granny!”

312
The Sociology of Language

One of the things that Tamils are famous for is their ‘love of their language’ which, however it can be measured, has got to be more intense than any other expression of ‘language loyalty’ found on the Indian subcontinent. More Tamils have died for their language than any other language group, and this intense loyalty has of course attracted attention by scholars who study other parts of India. I came to India to study Tamil syntax and write a dissertation on that subject, but I soon found myself being asked to comment on the Tamils’ language loyalty, which of course in 1965 had resulted in various forms of extreme (and sometimes violent) resistance to the imposition of Hindi as the national language.

When asked to write something about this topic, I did so, but soon found myself inadequately prepared to approach this topic without preparation in a field that was far from what the discipline of Linguistics had prepared me for. Fortunately for me, I was drawn into what is known as the Sociology of Language by the appearance of a book on the German language in America (Kloss 1963), which drew me into the topic of my own linguistic heritage as an American of German descent, and with more reading, I decided to offer a course on the topic of ‘Language Policy.’ My experience with Tamil helped widen my approach to this topic, and I taught the course both at the University of Washington and the University of Pennsylvania for almost 35 years. I soon discovered that there was an extensive body of literature on this subject, including but not limited to the work of Fishman, Ferguson, Haugen, Hymes, and many others, which helped me to understand topics such as ‘language loyalty’ and to present them to students.

The study of language policy soon became my primary research interest, and because I had once concentrated in Slavic Linguistics and had visited the Soviet Union and gotten a taste of its linguistic diversity and its language policy, coupled with the fact that I had also lived in France for two years, which has its own kind of linguistic chauvinism, gave me a range of experience to deal with various kinds of language policy issues.

What keeps a language alive? The point I want to make about this is based on research about what keeps a language vital and alive, and what leads to language shift, and language death. Much has been written about what the effective methods for language vitality and preventing language death, but one of the most important claims for effectiveness was made by Fishman (1991) who proposed that ‘intergenerational transfer’ is the most crucial factor in keeping a language alive. That means that all other strategies, such as using the language in education, or recording the words of the last viable speaker, or any other strategy that may be proposed, are all ineffective and useless, unless intergenerational transfer takes place. For the case of Tamil (and in fact for any ‘threatened’ language), this means that Tamil must be learned and spoken in the home. If it is not spoken in the home, but only learned at school, it will not survive, and some other language, probably English, will replace it.
The Case of Singapore. One of the crucial cases that illustrates this problem most clearly is the case of Singapore. In Singapore, as is well known, Tamil is one of the ‘official’ languages and receives support from the educational system along with other ‘mother tongues’ of the Singapore population. But Tamil is increasingly not the language spoken at home in many Singapore Tamil families, and many researchers now fault the educational system, which until recently insisted on teaching only Literary Tamil, but giving no support to the spoken language. Without this, English takes over as the dominant language (or its Singapore variant, Singlish). In research I conducted in Singapore in 1994, Tamil students in the system revealed to me that they did not feel that Tamil was their language; it belonged to someone else, they said, and they saw no use in learning it, especially since it had no economic value in Singapore. They also voiced the complaint that no matter how hard they tried, they could never satisfy their teachers, who always faulted them on their poor knowledge of Tamil. This research is buttressed by research by others who have studied the Tamil system in Singapore, e.g. Gopinathan, Seethalakshmi, Saravanan, and others.

One might argue that Singapore is different from Tamil, and that in Tamilnadu, where Tamil is the ambient language, this is not a problem. But increasingly, I find that Indian Tamils who have been educated in English medium schools do not handle Literary Tamil well, and speak a kind of spoken Tamil that is heavily mixed with English.

What I am saying is that we make an error if we assume that supporting the study of one kind of Tamil, such as Classical Tamil or modern Literary Tamil, will solve all our problems, and keep Tamil alive. We need in fact to treat Tamil as a family of languages, and support all of the members of that family. The branch of the family that gets the least support typically is that of spoken Tamil and all its variants. This is the true ‘mother tongue’ of all Tamils, the one they learn first, the one in which their emotions are centered.

This notion is often ridiculed by Tamil experts, but in the discipline of Linguistics, we know that by the age of six, about the time when children start school and start to acquire literacy, they have solidified their knowledge of their mother tongue, the one they learn at home at their mother’s knee. Our educational systems tend to assume that their only responsibility is to teach the literary language, but in fact the literary language will not be learned if the spoken language is not used as a resource, instead of treated as a liability. Educational systems that try to ‘kill off’ the spoken language (whether English, Tamil, French or any other language) will also probably kill the language that they are attempting to teach.

Spoken Tamil as Resource. Let me give an example from my own experience of teaching Tamil. One of the hardest things for learners of Tamil who have no home background in Tamil is the syntax of relative clauses in Tamil. In English, we can take the example of two sentences, that are combined using a relative pronoun such as ‘that’ or ‘which’ or ‘who’ and make one sentence:

1. That boy came yesterday.
2. I saw the boy.

→I saw the boy WHO/THAT came yesterday.
In Tamil, of course, this is done differently. There is no relative pronoun, but instead, the verb of one of the sentences is converted into an adjective, and placed before the co-referential noun:

1. *anda payyan neettu vandaan.* That boy came yesterday.
2. *naan payyane paatteen.* I saw that boy.

→ *naan neettu vanda payyane paatteen* I saw the boy who came yesterday.

These are obviously two very different kinds of syntactic structures, and what I have found in my thirty years of teaching Tamil to both Americans with no background in Tamil, and some Tamils whose parents were born in India, is that Americans with no background have a very hard time with these structures—they have to be drilled over and over to master them, whereas students with a Tamil background at home have no problem with these sentences, either in spoken Tamil or in Literary Tamil.

In other words, knowledge of the spoken language is not only not useless, it is an asset, and needs to be built-upon, rather than exterminated.

**So where shall we start?**

First of all, we need data in the form of a database of spoken Tamil materials, in order to be able to conduct some much-needed research on what is the most appropriate form of Tamil to teach. I suggest that we organize to create this database using the following sources:

- **Tamil ‘social’ films.** There are hundreds if not thousands of Tamil films whose soundtracks could be digitized, transliterated, and made searchable for examples of all kinds of spoken Tamil, but mostly what I call ‘standard’ Tamil. (Schiffman 1998)
- **Tamil radio plays and television sitcoms** also constitute a source for more spoken Tamil.
- **Linguistic Survey of India.** The LSI, compiled early in the last century, has transcription of spoken Tamil of various sorts, and recently gramophone records of some of these samples have been digitized and will soon be made available for study.
- **The English Dictionary of the Tamil Verb** has sound files for more than 9,700 spoken Tamil example sentences. These sentences could be easily adapted for use in a ‘matched-guise’ type study.
- **Field recordings.** Many researchers, myself included, have made tape recordings of Tamil speakers in various dialect areas, and could pool these resources to add to the database.

**The Matched Guise Test:** One of the effective ways to study spoken Tamil and discover how mother-tongue speakers conceive of various forms, including which examples of spoken Tamil constitute the most useful and ‘acceptable’ form to be used as models for students to imitate and learn, is the research methodology known as ‘Matched Guise’ testing. Matched guise tests originated in Canada in the 1960’s as a way to determine Canadians’ attitudes towards the ‘other’ language, i.e. attitudes of Anglo-Canadians towards French, and attitudes of Franco-Canadians towards the English of Anglo-Canadians. Over 100 such studies have been done on languages around the world, comparing the language attitudes of bilinguals and bidialectal people, and recently some studies have also been carried out in Singapore as a way to study attitudes of Tamils towards various forms of Tamil (Seethalakshmi et al. 2005, 2006). More needs to be done on this among Tamils in Tamilnadu, since attitudes toward Tamil among Tamils in Singapore seem to be different from those found in Tamilnadu.
The effectiveness of matched-guise testing comes from the fact that subjects are asked to evaluate, not the *language* of other speakers, but the *speaker* him/herself.

Matched-guise tests are constructed utilizing bidialectal or bilingual speakers, who are recorded speaking in each of their two variants. Then the recordings, typically of five bilinguals, are scrambled—they are mixed with those of other speakers, and played to subjects who are themselves bilingual/bidialectal. Typically, the subjects fail to recognize that the same speaker has been recorded twice, so when they hear the samples, they think they are hearing ten different speakers. Asked to judge the speakers on variables such as level of education, what kind of job they might have, what kind of earning power they might have, as well as other social variables, the subjects rate the ‘guise’ of one of the bilinguals more highly than the other guise.

In Canada, English ‘guises’ are ranked higher than French guises, even by French speakers; subjects even rank English speaker-guises as taller, although French guises are usually ranked as ‘more friendly.’ In all the matched-guise studies I have looked at, there is always a differential in these areas—there is never ‘equality’ of rank, despite any attempts in various societies to create social equality.

In case the notion that speakers can tell if someone is tall by listening to their voice, consider the following cartoon, which seems to assume that speakers can imagine all kinds of physical characteristics of other speakers by their voices!
Conclusion
So who will volunteer to join me in this endeavor? There are a number of problematical issues we will need to deal with:

1. We will need to convince many people in the Tamil establishment that Spoken Tamil is something worth studying and collecting data for.

2. We will need to seek funding to support the collection of data, its transcription, and a method for accessing forms online. This will involve ‘tagging’ of forms, since Spoken Tamil is (in some ways) morphologically more complex and less transparent than Literary Tamil.

3. We will need to get legal permissions to copy the sound tracks of Tamil films, and other material that already exists ‘out there.’

Bibliography


Website: http://ccat.sas.upenn.edu/~haroldfs/bibliogs/ MACHGUIS.HTM
Tamil Literature from Sangam to Modern Period: A Continuum with colorful changes: What does a search of the Tamil Electronic data reveal us?

Vasu Renganathan
University of Pennsylvania (vasur@sas.upenn.edu)

Introduction

A systematic study of the Tamil language from Sangam to Modern period from a historical perspective may reveal that there does exist a continuum of changes that occurred from one stage to another in Tamil language. Without such a study, any synchronic description of Tamil would only reflect its complexity in an overwhelming way. In other words, The Tamil language, the way it is now with a museum of complex forms, expressions and grammatical constructions, both in written and spoken variety, demonstrates a vast number of linguistic characteristics at phonological, morphological and syntactic levels, that require a comprehensive diachronic study to fully understand them in a coherence way. In this respect an extensive electronic database of Tamil texts from all of the stages along with a powerful query tool to search texts from various dimensions is indispensable. This paper is an attempt to illustrate how such an electronic database for Tamil (http://www.thetamillanguage.com/sangam) can be used extensively to study some of the morphological and syntactic behaviors of Tamil from a historical point of view.

Upon exploring the Tamil electronic database consisting of a variety of data ranging from the Sangam to Modern Tamil, especially by employing the principles of historical linguistics, one may immediately be able to notice that the changes that underwent throughout the history of Tamil language exhibit a systematic, regular and what one may attribute as a set of colorful changes in it. Phonological, morphological and syntactic changes that took place to this language one after another in a sequential manner contributed to the dearth of complexity as we see now as modern Tamil (both spoken and written) – a language that many have attempted to study it using many grammars and dictionaries in many different points of views! What may one illustrate it in a minuscule is that when words or combination of words and suffixes undergo all possible phonological rules on them, either successively at one point of time or periodically at different stages, what results is a set of the most complex forms that can be understood in terms of many dichotomies such as social versus regional dialect; spoken versus written variety; high versus low register; casual versus platform speech and so forth. Thus, attempting to learn this language that contains such a complex set of shades of variations does indeed pose a greater level of difficulty than normal for any second language learner. Not only does it become a big challenge to any second language learner in having to comprehend and use these multiple facets of this language, but it also becomes an immense task for an instructor/evaluator as to how one can judge the competency of a learner who attempts to master it! Thus, by not familiarizing oneself with the myriad of complexities within the Tamil language, either from a historical or purely from a synchronic point of view, one may tend to attribute each of these varieties as belonging to a separate language; and subsequently consider the variations therein as haphazard and random. Upon
studying the Tamil language variations from a historical point of view, one may easily note that such variations are vibrant and quite regular, and notably they conform to a logical sequence of changes. In this respect, no form of Tamil, either it is spoken or written, is neither random nor spontaneous in nature. Not to mention the fact that any study of diaspora Tamil of any region, without such a systematic account from a historical point of view, would only result to provide an unscientific description of the language of the respective region.

Language change occurs as a result of both internal as well as external causes. Internal causes are a) application of more than one phonological rule on agglutinative words; b) undergoing many naturally occurring linguistic processes such as, grammaticalization, reanalysis, metaphorization etc., in the language (See Renganathan 2010). The external causes, on the other hand, can be attributed to such factors like ‘foreign language contact’, ‘bilingualism’, ‘language dominance’, and so on to name a few. Not to mention the fact that over the period of a long history, Tamil language did undergo many changes both due to internal as well as external causes. Prakrit, Sanskrit, Persian, Portuguese, and more recently the English language contributed enormously to the development/distortion of Tamil language in a number of different ways. Interestingly, many Tamil language movements, both conscious as well as unconscious, such as ‘language purism’, ‘official language planning’, ‘language standardization’, ‘Tamilization’, ‘coining new vocabularies’ and so on contributed to the retention of many of these variations within it without having to undergo any extinction in any subtlest manner possible. Many of the so called indigenous and historically relevant Tamil words and morphological and syntactic forms - although not all of them - from the Sangam era are still extant in modern Tamil in one way or another: in one dialect or another, in one speech form or another, or in one register or another. This particular behavior of the Tamil language poses as a big threat not only for its continued consideration as an individual language, but also for its continued use of indigenous and historically significant forms under various labels as ‘pure Tamil’, ‘Sangam Tamil’, ‘Chastised Tamil’ and so on. Ironically, the major threat comes mostly from the judgments of second language learners for whom these historically relevant changes and existence of complex variations pose as a major hurdle in learning the language in a casual manner.

Delving into the complexity – A case in point is the use of the verb en ‘to say’:

Almost all of the grammatical categories in Tamil have a systematic history behind them, and accounting all of them may require enormous amount of time and energy. An attempt is made in this section to trace the various use of the Tamil quotative marker enṉu ‘that’ and its historical development, especially by making use of the electronic data extensively. Use of the verb en ‘to say’ can be taken as one of the instances for the contribution of complex forms in Tamil. This verb has underwent a wide range of alterations throughout the history of the Tamil language, but yet, it is still in use in the modern language the way it was during the Sangam period - perhaps with more number of characteristics which were not prevalent at its earlier stages. Unlike any other verb, this verb exhibits many structural gaps in modern written variety, but, significantly, not in spoken Tamil. Learning to master all of the uses of this verb, especially in spoken Tamil, is definitely one of the major challenges to any second language learner for the main reason that it not only underwent the process of grammaticalization, but also shows an agglutinative structure that is very difficult to comprehend and use by any non-native speaker of the language. This verb was used both as a regular lexical form
as well as a grammatical form representing the ‘complementizer’ in the Tamil language. Both the forms of enṉu ‘that’ and enpatu/enṉal ‘the fact that’ had their equivalents both in old and modern Tamil.

A search of the database using a number of combinations, including எஒuni0.A9_uni0.ஙஞ, எஒuni0.A9_uni0.ஙஞஒuni0..ீ_uni0.ஙீ, எஒuni0.A9_uni0.ஙஞறிஒuni0.9F_uni0.ஙஞ, எஒuni0.A9_uni0.ஙஞஒuni0..ீ_uni0.ஙீெகாஒuni0..3_uni0.ஙஞ and so on would reveal that besides the many of the finite forms of this verb, what underwent a significant change at a later period are the forms of negative adverbial (ennātu ‘without telling’) and nominal derivative (ennāmai ‘not saying’), which do not show any equivalent in modern literary Tamil.

Notably, the Modern Tamil equivalents of the suffixes -ātu and -āmai such as -āmal (eg. collāmal ‘without saying’ *ennāmal) and -ātatu (collātatu ‘that which was not said’ *ennātatu) respectively tend to occur with the verb en only in spoken Tamil but not in the corresponding literary variety. What turns out to be the crux of the issue here is the obscure nature of the spoken Tamil equivalents of the verb en ‘say’ in present, past and future forms, which normally occur as a single or clustered consonant: எ எ எ (பாஒuni0.9ஏ_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞ pākṇēṇṇēn ‘I say that I see’, எ எ எ (பாஒuni0.9ஏ_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞ colṇāṇṇān ‘he says that he tells’); எ எ எ (பாஒuni0.9ஏ_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞ pākṇēṇṇēn ‘I said that I see’); and எ எ எ (பாஒuni0.9ஏ_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞேறஒuni0.A9_uni0.ஙஞ koṇuppēmpēn ‘I will say that I would give’) respectively (Cf. Search: ngr, எஒuni0.A9_uni0.ஙஞகிற). The obscure form of this suffix, its complex clause construction in an agglutinative form, along with the non-existence of some of the conjugations of this verb in written Tamil contribute enormously to the complexity of spoken Tamil.

Notably, these, supposedly, commonly occurring forms in spoken Tamil do not have any parallel in written Tamil, as a result it generates a structural gap in the corresponding written variety of Tamil. What one can attribute to this phenomenon is that the spoken Tamil exhibits a perfect continuum from Sangam to the present time as it continues to retain the structure that one can attest from old Tamil, but this is not the case with the corresponding written variety of Tamil, which exhibits a structural gap in terms of not exhibiting the equivalents of āmal and āmai with the verb ‘en’.

321
If the form 'enkāmal' is nonexistent in modern written Tamil, but only found in spoken Tamil as in -kāma, a question arises as to when and how the form 'enka' as an infinitive form of this verb lost its use in the history of Tamil language? The other alternative point of view would be to consider this form as an innovation in spoken Tamil but not in modern Tamil. Note that the Sangam Tamil form 'enāṭi' enka occurs as an ‘optative form’ to mean ‘let it be said’, but not as infinitive form of the verb ‘en’.

However, neither the Sangam Tamil forms such as ennāmai or ennātu, nor the relatively more recent forms such as enāmal or enātu do not seem to have any parallels in written Tamil, but as we noticed above, they do occur in spoken Tamil with their root forms of the verb 'ṇ', 'ṇṇ' and 'ṁp' in a relatively large number of conjugations.

This is particularly true for the fact that one can observe from the search results of the electronic database using the forms such as 'ṇṇi' and 'ṇṇu', which especially use of the aspecual auxiliaries such as iṇu (definitive auxiliary) and koṇ (reflexive auxiliary). Along with the progressive auxiliary form koṇṇi, these forms seemed to have been attested only starting from the medieval bhakti literature, especially from Tirumular’s Tirumantiram, as sited below.

Surprisingly, like in the earlier cases of negative verbal participle and verbal derivative form, these constructions also do not exhibit in parallel in modern written variety, but only found widely in spoken Tamil.
ellārum varu-v-ōm-ē-u-ē-āka! ‘Eveyone proclaimed affirmatively that they would come)

What do these exceptional forms imply is that ‘spoken Tamil’ and ‘written Tamil’ seem to have followed two different historical paths from Sangam to modern Tamil and in this respect the spoken Tamil seems to show a richer structure than the written Tamil, especially in terms of retaining more number of archaic forms than the corresponding written version. This is in opposition to those instances of modern Tamil where new structures evolved and no traces of which can be found either in Sangam or in medieval Tamil. An example may the case of experience subject construction, which is new to modern Tamil, but not in Sangam Tamil, as in yāṉ viyarttaṉaṉ ‘I was sweat’ as opposed to eṉakku viyarkkiṉatu (cf. Murugaiyan 2004). Yet another feature from a historical point of view is loss of medieval and Sangam forms which do not have any trace in modern Tamil. A case in point is the use of imperative suffixes –min (kēlmin ‘listen’) and –anmin (kūṉanmin ‘do not utter’) etc., which do not have any occurrence in any identical forms in modern Tamil (cf. Renganathan 2010). Identifying the point of time in which these changes occurred is an endeavor that requires analyses of text of different genre in a thorough manner.

Yet another advantage of studying word forms that underwent many changes historically using electronic data is that it is possible for one to trace the trajectories of the cause of certain changes over the period of time in a systematic manner. One of such phenomena is authors’ handling of a particular style causing the development of new categories. One of them that may be sited here is the formation of the modern Tamil modal auxiliary lām. It may be stated that various use of the combination of the infinitive suffix –al with the neuter future form of verb āku ‘become’ in ancient Tamil later caused the formation of lām. An extensive search using the keys such as ṣaṃaṅ, ṣaṅ, ṣaṅaṁ, ṣaṅaṁ, ṣaṅaṁ, ṣaṅaṁ etc., one may notice that the modal auxiliary lām came into existence in modern Tamil by the linguistic process of reanalysis due to various use of this structures by poet saints. Consider for example the expression kēṇu uṇṇu enṇal tuṇintu colal ākum – Manimekalai and its modern Tamil equivalent kēṇu irukkum enkiṇwu tuṇi ntu collalām (Modern Tamil) ‘One may say for sure that there would be a devastation’, where the syntactic construction colal ākum ‘saying is possible’ is found to be occurring with many different combinations synonymously, as in colal ām - after phonological reduction of ākum to ām; colla lām with a reanalysis of verb forms and so on (see Renganathan 2010: pp. 171-73 for a detailed study of this change). By toggling between the selections of the bhakti, Sangam and modern literatures using the above search keys, one can notice the various use of this combinations more in bhakti texts than in Sangam texts.
Search techniques and need for a tagged Corpora:

Perhaps an advance search technique with many combinatory possibilities is needed to successfully derive all of the intended and unattested forms from all of the genres of Tamil language. Ideally, one may want to search text in many complex ways, like ‘words that end in particular suffix (-viṇu; kiṇ; koṇ etc.),’ sentences with a particular combination of words (dative subject and psychological verbs; subject with the suffix āl and modal verbs like -oṇṇ, -iyal etc.) and so on. Even though such sophisticated search possibilities is yet to be made available for Tamil using any conceivable tagged corpus as discussed in detail in Renganathan(2001), Baskaran et al (2008) etc., with the current database, however, storing text in Unicode does offer some work around. For example, if one intends to retrieve all of the word forms with the suffix -iṇu, āl, -ukku and so on, one can use the Unicode glyphs of the initial vowels, as in ◌ிᾌ, ◌ா᾿, ◌ுᾰᾁ respectively to accomplish this task. This method can be considered as a substitute for any equivalent method of information retrieval using tagged corpus, which would normally contain all of the affixes parsed and stored separately in a more systematic manner. Absence of any such tagged corpus and an intelligent parser for all of the genres of Tamil texts from Sangam to Modern Tamil, one requires to use this kind of alternative search methods to accomplish the task. Among many others, the other significant advantages of using electronic data may be making dialect geographies from a historical point of view, attempting to find the chronology of authors and texts and so on.

References

Open Source Tamil Computing

S. Gopinath and E.I. Nehru
National Informatics Centre, Chennai

Abstract

For many of us English is the natural choice for commodity Computing such as Internet Web browsing, email, Instant Messaging, Word Processing etc. As computer proliferates in every aspects of our daily life, it is a need to have some kind of Multilingual Computing. For example an individual would like to send an email to his friend in a local language (like sending a greeting email in Tamil), an e-Governance application needs to be benifited for persons who are not familiar with English, or a product brochure to be made available on Web in a local language. More importantly the multilingual data should be interoperable and to be long lived across computer systems. The fundamental requirements for the Language Computing are

1. Coding system for Local Language Script Set,
2. Input Methods, Output Methods,
3. Software Libraries, APIs, Fonts
4. Applications

The aim of this article/paper is to summarize the various aspects of Multilingual Computing in OpenSource Plaform with emphasis on Linux and Tamil.

Contemporary Linux systems comes withMultilingual features and in most cases they can be enabled very easily.

1. Unicode

Internally, coding of character system to be done to make meaningful processing of characters. A decade ago Standard ASCII (7bit) is a very popular encoding system for English characters. Later extended ASCII (8 bit/1 byte) is used for representing English characters and several Latin characters, Indic characters etc. But a true multilingual system needs to represent all possible script set (including Math Symbols, braille characters etc.) independently so that any electronic text shall contain all sorts of characters.

In late 1980s, Joe Becker (Xerox), Lee Collins and Mark Davis of Apple conceptualized multilingual character set encoding. Joe Becker published a draft for International/Multilingual Character Encoding System and tentatively called Unicode. This proposed 16 bit character model (each character is to be represented by 16 bits). This proposal is popularly called as Unicode88. Later when Unicode 2.0 was proposed, the character width is no longer restricted to fixed 16 bits and hence its posible to represent many ancient character sets like Egyptian Hieroglyphs. As of 2011 Unicode 6.0 standard prevails.
Unicode defines codespace containing code points in the range 0x0 to 0x10FFFF (represented in Hexadecimal) which means there are 1,114,112 code points in the Unicode Codespace. Each character is assigned a code-point. Unicode codespace is divided into 17 planes and numbered from 0 to 16. The Plane 0 is called “Basic Multilingual Plane” which contains codepoints numbered from 0x0 to 0xFFFF. Code Points for Tamil Character Set, other Indic Character Set, Cyrillic, Geometrics Shapes, Math Symbols, Arabic etc are allocated in this plane (hence this Plane-0 is called Basic Multilingual Plane). Other planes are

1. Plane 1 - Supplementary Multilingual Plane
2. Plane 2 - Supplementary Ideographic Plane
3. Plane 14 - Supplementary Special Purpose Plane
4. Plane 15,16 - Private Area Use

The Tamil Language is allocated codepoints from 0x0B80 to 0x0BFF (128 codepoints) in BMP Plane. Please remember, that the codepoints has nothing to do with physical representation of characters in computer binary bits (or qbits in Quantum Computing). The physical representation also called as Unicode Character Encoding is done through a methodology called Unicode Transformation Format and it is defined by the Unicode Constortium. In Posix Systems and in Internet, UTF-8 Encoding Scheme is popular. In this article/paper, by default we frequently refer UTF-8 encoding scheme.

The UTF-8 was initially proposed for Plan9 Operating System. It is a multibyte character encoding system which uses one byte to 4 bytes depending upon the codepoint. The encoding for the code point from 0 to 127 is same as ASCII encoding for ASCII space 0 to 127 and hence it is already compatible with ASCII for the code points 0-127. The main advantage for the UTF-8 is its self synchronising and does not depends on endianness of the computer system. No special marking in the data stream is required for UTF-8. The main disadvantage is that it needs more bytes and it requires extra processing.

One of the basic question that arises is this. We have only 128 codepoints alloted for Tamil in the Unicode and we have 247 number of characters in Tamil Character Set. Is this a right way ?. How do we manage to represent all characters in Tamil Character Set?. This is very debatable. The brief explanation for the question how to accomodate 247 characters in 128 code point is as follows.

The most of the characters in Tamil are conjunct characters and share the same set of glyphs (or visual representation) in most of the characters and they are highly structured. Hence, it is possible to represent the Tamil Characters with very little compromise within 128 code point. But conjunct characters takes more bytes space than vowels or consonants (consonant equivalents).

2. Input System

It is quite obvious that keyboard is the de-facto input device for humans (Yes.. we have voice and other input systems for Humans!). In PC based architecture, keyboard is a raw device, and has no language centric mechanisms expect that in most of the keyboards we have “English” letters inscribed on keys. One of the questions beginners of Tamil Computing asks, is there any keyboard where I can find all 247 characters ?. The answer is similar to that of in the previous section that we indeed do not require individual keys for all the 247 characters (infact one can make that eiher physical or virtual
onscreen keyboards, but it would be expensive and bulky) as many are conjunct characters and many share common glyphs (diacritics) and hence the normal general purpose contemporary keyboard is suffice. This methodology requires some kind of software popularly called Input Methods to assemble the characters when multiple keystrokes are required to form conjunct characters.

Pioneering works has been done by the creators of m17n libraries which is quite popular in Posix systems. The m17n library is a multilingual text processing library for C language. In Linux the popular Input Methods available are Smart Common Input Method (SCIM) and iBUS (Intelligent Input Bus). Most of the current Linux uses iBUS by default.

3. Output Methods

The whole fanfare in Multilingual Computing is on the output. Unless one sees the output in appropriate script, the multilingual computing is not fulfilled. Its the most obvious part of Multilingual Computing. Normally, we talk about fonts whenever we discuss about Multilingual Computing. Fonts contains information about how to draw the shapes on the screen. The renderer engine which part of the output system takes the font and renders the shapes. The fonts are usually distributed in a file (often called as a font file). There are 3 types of fonts. They are Bitmap fonts, Outline fonts and Stroke based fonts. The bitmap fonts or raster fonts contains pixel images of the glyphs. The Outline fonts or vector fonts contains the vector images and mathematical formula can be applied to get the different sizes of a glyph. The popular fonts types such as TrueType, OpenType, Adobe fonts are Outline fonts.

One of the functions of m17n library is displaying and rendering multi-lingual texts. It requires complex processing to render scripts such as Tamil, Devanagiri etc. A character may have a single glyph or few glyphs spaced at one another. A re-ordering may also be required. The technology for such rendering is known as "Complex Text Layout” or CTL. The m17n library database contains what is called Font Layout Tables (FLT) which bridges the fonts and the rendering engine.

The present day Linux distribution such as Debian, RedHat are available with TrueType and OpenType Font rendering engine. The Lohit fonts are quite popular in Linux Community. Lohit Fonts Project is sponsored by RedHat and its a Fedora Hosted. CDAC distributes OpenType Fonts for Indic Languages. By default most of the Linux distros has Tamil fonts by default or can be added very easily.

4. Software Libraries/APIs

The GNU C Library (glibc) supports UTF8 encoding and contains functions for multilingual computing. The utilities such as sort, grep which are based on glibc can be used on multilingual texts. The holy grail M17N C library is the flag ship for multilingual computing. The link http://unicode.org/resources/libraries.html contains links for various multilingual libraries.

5. Applications

The end user experience on multilingual computing based on the Applications. The simple commodity applications like Internet surfing, Word Processing and e-mail etc. can be done using
existing applications like Mozilla Firefox, ThunderBird, OpenOffice, Libre office etc. The user has to only choose the appropriate Input Method (so that he can type in Tamil). Applications like Yahoo Messenger is multilingual and once can chat in clear Tamil (instead of transliteration) Empathy which is an GNOME application can be used for Instant Messaging and its multilingual. Simple editor like gedit can be used to create multilingual files.

Web Static Pages can be created by including a http header Content Type. This can be done by adding a meta tag between <head> and </head> as follows.

```html
<head>
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
</head>
```

In the body section, the content can be multilingual. The multilingual content can be created even with simple editor like gedit. Of course the browser and the system in which it runs must have UTF-8 support with appropriate language fonts installed. Enterprise and backend applications use Database Management Systems and in Open Source Community one of the most popular DBMS is Postgres. The current version of Postgres supports UTF-8 encoding by default on Posix systems. This facilitates the storage of data in multilingual form.

6. Requirements and Suggestions

Although it is possible to carry out most of the multilingual computing, the following are the suggestion by the authors for further enhanced functionality.

1. The simpler keyboards can be made to meet the various requirements of the end user. For example to use TamilNet99 keyboard requires practice and considerable understanding of the scripts, the people who just knows only to read or write the scripts could not easily use these keyboard layouts or definitely not as trivial as someone who only knows English and use English keyboard. A much simpler keyboard may be prototyped after doing sufficient study on the above aspects. As the input methods are implemented in a software, these initiatives can be done without changes in underlying software.

2. As of now Input Methods are implemented in software and they run in main system/CPU. It is possible to design and implement an Intelligent keyboard, which directly sends the UTF-8 sequences instead of raw keyboard scan codes. As Unicode integrates across language/scripts, culture etc, the Universal Intelligent Keyboard is appropriate. The keyboard can be made programmable such that the end user can configure the keyboards to function either as raw scancode mode, or direct UTF-8 mode. If the same kind of keyboard is mechanically designed foldable/wrappable, it can be used in mobile devices such as smart phones, tablets universally (which can bring down the costs).

3. The authors could see that more works to be done on Optical Character Recognition, Speech to text conversion, Accessibility in the context of Multilingual computing.

4. Better Tamil outline fonts needs to be forged for High Resolution equipments.
7. Conclusion

Multilingual computing in Open Source Systems like Linux is matured. Works like M17N C libraries, UTF-8 encodings really makes Multilingualization possible in Linux like systems which carefully eliminates need for any special magic numbers. Due to glibc and m17n libraries, large number of applications are already multilingual enabled.

8. Acknowledgements

I would like to thank Mr.T.N.C. VenkataRangan of Vishwak Solutions Ltd, for spending hours in answering my novice questions when I started to learn Multilingual Computing.

My special thanks to CDAC Chennai officials (Mr.C.Sudhakar, Mr.Nagesh, Mr.Gowri Ganesh) for assisting me in learning some of the concepts in this area.

My thanks to my seniors (Mr.E.I.Nehru, Mr.JD Prince) for encouraging as well answering my doubts on the subject. My thanks to NIC’s FMG Engineers (Mr.Emthias and others) for their Technical Support.

I thank NIC for providing enough resources.

Thanks to IITM Students for providing help in preparing the document in LaTeX.

Not but not least, my thanks to my parents and my brother who patiently accepts my late arrivals and patiently serve late dinner as I had to spend late hours while preparing this article.

9. Few References

- http://www.cl.cam.ac.uk/~mgk25/unicode.html
- http://code.google.com/p/ibus/
- http://www.m17n.org
- http://www.postgresql.org/docs/9.0/static/sql-createdatabase.html
- Ofcourse, Wiki and Google pages.