"தமிழ்பட்டை"- Multi function Unicode Search Bar using Transliterated Input

Nithianandam Sathish Kumar[*]**, Sudhakar Balakrishnan**
New York, USA

INTRODUCTION

This paper addresses the needs of Tamil users to have a single tool to perform Unicode search with Transliterated Input and convert TSCII pages online to Unicode.

"தமிழ்பட்டை" (tamizpaTTai) - provides simple packaged solution for Tamil users with a Multi-functional toolbar. It seamlessly merges with IE, the most used browser. It provides user with many helpful tools, which helps surfing Tamil pages with less overheads. They do not need a separate transliteration tool to perform their search.

FEATURES

- ❖ Perform Quick Unicode based Tamil searches
- ❖ Handy Transliteration Learning tool with instant feedback
- ❖ Transliteration Scheme Table is provide as help to guide new users
- ❖ Changes the last input character to its alternate character, when control key is pressed
- ❖ Support multiple transliteration input schemes to support existing users
- ❖ Preset with Major Unicode based search engines
- ❖ Easy installable tool for Tamil users
- ❖ Online access method for "உருமாற்றி" (urumARRi) Script

TECHNOLOGY

" தமிழ்பட்டை " (tamizpaTTai) is developed using Microsoft Visual C++ 6.0 with ATL & WTL. It runs on Internet Explorer v5.5 and above.

TOOL USAGE

"tamizpaTTai"- accepts Romanized inputs in its first box, transliterates it to Tamil Unicode (UTF-8) character and shows the result dynamically in the second box. When clicked the "தேடு" button, it uses the second box Unicode text to perform search on the user's selected search engine.

"tamizpaTTai"- has a third non-editable box, which shows how the input characters are transliterated into the Tamil Unicode characters. Again it shows results dynamically when the user input the text in the first input box. It greatly helps users who are new to Tamil transliteration. It also guides them how it works internally to learn the scheme faster.

In the Configuration Setting, the user is given option to select the transliteration scheme which the tool bar should use to transliterate the input ASCII codes to Unicode. Also users can set the URL path for the custom search 1 & 2 search engine options.

Transliteration Scheme used by "tamizpaTTai" has been represented in **five tables**. When the user inputs the character, its default value is shown in the output box. If he intends to change the character to its related character, he needs to press the [Ctrl], to get its Alternate Characters 1,2,3,etc., from the table. User gets all the available alternate character in cycle from the table.   This time saving option would be useful as it minimizes the need of backspace.

Figure1: A snapshot of our implementation: User's browser with "தமிழ்பட்டை" (tamizpaTTai)*, performing a search on "tamiz theedal" using Google as search engine*
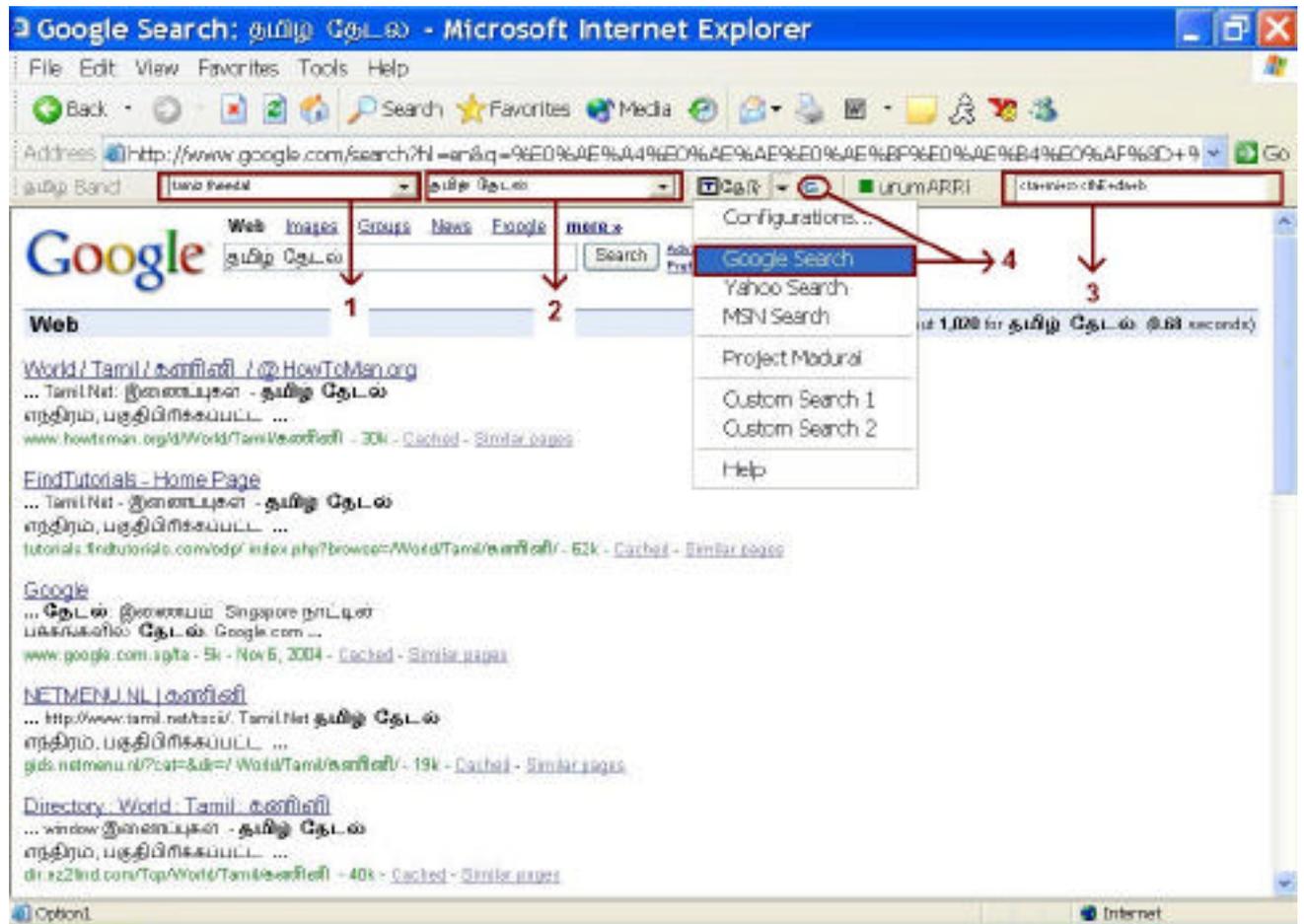
1. " தமிழ்பட்டை " (tamizpaTTai) - ASCII / English Input Box
2. " தமிழ்பட்டை " (tamizpaTTai) - Unicode / Tamil Input/Output Box
3. " தமிழ்பட்டை " (tamizpaTTai) - Feedback Output Box with Normalized Transliterated Scheme Info
4. Google Search Engine, selection shown in the toolbar.

Figure 2: Preview of Configuration Panel
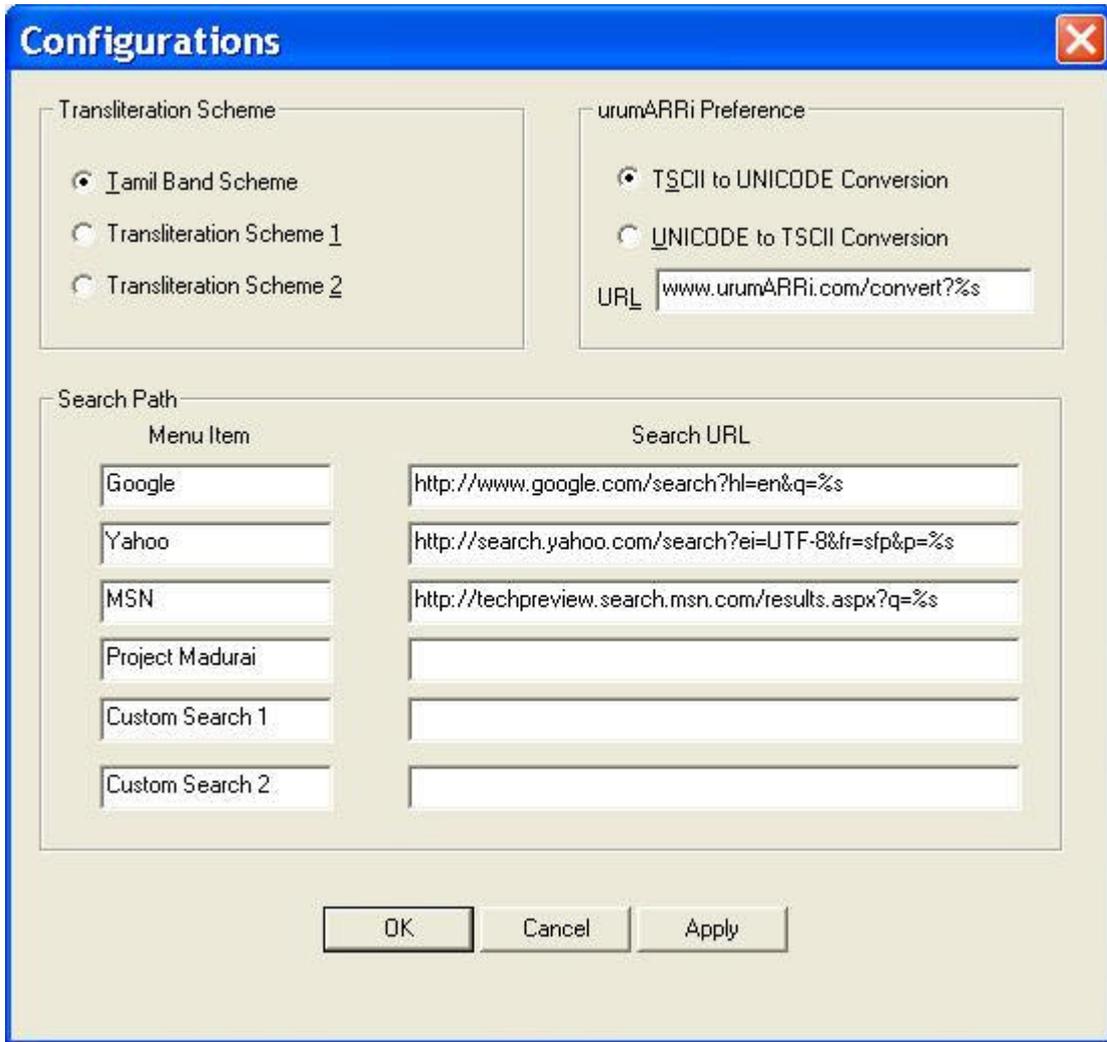
Section 1: Transliteration Scheme   Section 2: urumARRi Preference   Section 3: Search Path

tamizpaTTai's TRANSLITERATION SCHEME TABLE

Tamil Vowels	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ
தமிழ்பட்டை Default	a	A	i	I	u	U	e	E	ai	o	O	au	q
Alternate 1	A	aa	I	ii	U	uu	E	ee	i	O	oo	ow	h
Alternate 2		a	e	i	o	u	a	e	ae	u	o	ou	k
Alternate 3			e		o		a	ea		u			

Tamil Consonants	க்	ங்	ச்	ஞ்	ட்	ண்	த்	ந்	ப்
தமிழ்பட்டை Default	k	ng	c	nj	d	N	th	n-	p
Alternate 1	q		s	ng	t	n	dh	w	b
Alternate 2	g		ch	gn		n-	t	N	f
Alternate 3				j			d	n-	

Tamil Consonants	ம்	ய்	ர்	ல்	வ்	ழ்	ள்	ற்	ன்
தமிழ்பட்டை Default	m	y	r	l	v	z	L	R	n
Alternate 1			R	L	f		l	r	N
Alternate 2									n-
Alternate 3									

Grantha (Not fully in Unicode)	ஷ	ஸ	ஜ	ஹ	க்ஷ	ஸ்ரீ
தமிழ்பட்டை Default	sh	S	j	h	ksh	sri
Alternate 1		s	g		x	sre

Lower case	a	b	c	d	e	f	g	h	i	j	k	l	m
தமிழ்பட்டை default	அ		ச்	ட்	எ			ஹ்	இ	ஜ்	க்	ல்	ம்
Alternate 1	ஆ	ப்		த்	ஏ	ப்	ஐ		ஈ	ஞ்		ள்	
Alternate 2	ஏ				இ	வ்	க்		ஐ				
Alternate 3					ஈ								
Upper case	A	B	C	D	E	F	G	H	I	J	K	L	M
தமிழ்பட்டை default	ஆ				ஏ				ஈ			ள்	
Alternate 1	அ				எ				இ			ல்	
Alternate 2													

Lower case	n	o	p	q	r	s	t	u	v	w	x	y	z
தமிழ்பட்டை default	ன்	ஒ	ப்	.	ர்			உ	வ்			ய்	ழ்
Alternate 1	ண்	ஓ		க்	ற்	ச்	ட்	ஊ		ந்	கூ		
Alternate 2	ந்	உ				ஸ்	த்	ஒ	ஐ				
Alternate 3		ஊ						ஒ					
Upper case	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
தமிழ்பட்டை default	ண்	ஓ			ற்	ஸ்		ஊ					
Alternate 1	ன்	ஒ			ர்			உ					
Alternate 2	ந்												

Multi Char	aa	ii	uu	ai	oo	au	ow	ng	ch	nj	ng	th
தமிழ்பட்டை default				ஐ		ஒள		ங்		ஞ்		த்
Alternate 1	ஆ	ஈ	ஊ		ஓ		ஒள		ச்		ஞ்	

Multi Char	dh	n-	sh	ksh	sri	sre	ae	ea	ou	ch	gn
தமிழ்பட்டை default			ஷ்		ஸ்ரீ						
Alternate 1	த்	ந்		கூ	ஸ்ரீ	ஐ	ஐ		ஒள	ச்	ஞ்

KEY ASPECTS

❖ Provide opportunity for users to learn transliteration scheme with visual aid feedback, given by three input boxes.

❖ Custom select input transliteration scheme

• User can select their preferred Transliteration scheme from the provided list.

❖ **Implementation of many well-known transliteration schemes for ease of use.**

• Current Tamil Software Schemes like e-Kalappai / Anjal.

• Literature & Phonetic Schemes like University of Madras / University of Cologne.

• Best selected Local Scheme based on new ideas from public.

❖ urumARRi' solution is implemented to facilitate user to view TSCII page in Unicode platform and vice versa dynamically.

• On a button click, URL from input box or the current page will be converted to Unicode / TSCII based on preference set in the Configuration Panel.

❖ Customizable Search Engine for searching user's favorite search engine.

❖ User can select one of the available search engines from a list.

- ❖ Major Unicode (UTF-8) based search engines are preset. It includes Google, Yahoo, MSN (implementing their Unicode based index solution Q4 2004).
- ❖ Search Project Madurai Unicode releases on www.infitt.org indexed by Google.
- ❖ Personalization of Search Engine is provided.
- • User can add in new Search Engine to the menu, by providing the Search URL like " http://www.thedu.com/q?%s ", %s in the URL will be replaced by Unicode value from the input box.
- • Once the user configures his new Search Engine, it's available from that point for his usage.

INSTALLATION

Installation and Uninstallation of "tamizpaTTai" is simplified with batch files. Close all your Internet Explorer sessions and run the batch file "Tamilband_Install.bat", to install the band on Internet Explorer and register in windows registry. Run the "Tamilband_Uninstall.bat", which deletes the installed files and un-registers itself from windows.

FUTURE SCOPE
- ➢ Online dictionary searches like the following can be provided –
  - o University of Cologne's Online Tamil Lexikon
  - o Umarthambi's TSCII interface Tamil dictionary on brinkster.com
- ➢ Yahoo groups and forum searches can be included
- ➢ Other Language users to convert their character sets to Unicode can use this approach. This can be implemented by extending the Transliteration scheme for their character sets.

LIMITATIONS
- ➢ Works only on Microsoft Internet Explorer specified versions and Unicode enabled Operating Systems.
- ➢ Only maximum of 2 additional personalized Search Engines can be added.
- ➢ Rendering of Unicode pages is slower as Unicode uses more space and memory.

REFERENCES
1. Internet Explorer Toolbar (Deskband) Tutorial   -   By Erik Thompson
      http://www.codeproject.com/atl/ietoolbartutorial.asp
2. Toolband (Toolbar for IE) sample using WTL   -   By Rashid Thadha
      http://www.codeproject.com/wtl/toolband.asp
3. Multiple references for transliteration schemes from web