INSTITUT FEMTO-ST

Webservices Platform for Tourism (WICHAI)

Kitsiri Chochiang — Fouad Hanna — Marie-Laure Betbeder — Jean-Christophe Lapayre

Webservices Platform for Tourism (WICHAI)

Kitsiri Chochiang , Fouad Hanna , Marie-Laure Betbeder , Jean-Christophe Lapayre

Département DISC

CARTOON

Abstract: The "WICHAI" platform is a core of the search engine components. This new algorithm enhances the effectiveness of search engines in tourism domain. It can reduce the search time and provide more flexibility information for non-native travelers in Phuket,Thailand. In this technical report we present 3 parts : Language processing, Website and Ontologies. The results are calculated similarity using 5 algorithms.

Key-words: Thai Language, RTGS , Phuket Tourism Platform, Ontologies

Technical Report : Webservices Platform for Tourism (WICHAI)

Résumé : La plate-forme "WICHAI" est un noyau des composants du moteur de recherche. Le nouvel algorithme améliore l'efficacité des moteurs de recherche dans le domaine du tourisme. Il peut réduire le temps de recherche et de fournir davantage d'informations de la flexibilité pour les voyageurs non-indigènes en Phuket,Thaïlande. Dans ce rapport technique, nous présentons 3 parties: Traitement du langage, le site et les ontologies. Les résultats sont calculés à l'aide de 5 similarité algorithmes.

Mots-clés : Thai Language, RTGS, Phuket Tourisme Plateforme, Ontologies

FEMTO-ST Institute, DISC research department
UFR Sciences - Route de Gray - F-25030 BESANÇON cedex FRANCE
Tél : (33 3) 81 66 64 00 – Fax : (33 3) 81 66 64 23 – e-mail : brigitte.bataillard@univ-fcomte.fr

1 Language processing

With the increasing amount of information available on the internet one of the most challenging tasks is to provide searching that are easy to use without having to learn a specific syntax. In order to overcome this limitation, we proposed a system that applied between information and language processing based on 3 ontologies. This system support both English words and Transliterated words. In addition,it can search the appropriate words whereas the level of word-meaning understanding is still maintain.

1.1 Thai-English Transliteration Words

In this research presented operation of a semantic web based intelligent tour planning tool. This tool allows the user to input their travel requirement in English words or Transliterated words. Usually, Transliteration between Thai to English is very difficult due to phonetics distinstion. RTGS is the official system to processing but it is not suitable for tourists who do not have any knowledge of Thai linguistic. So, we applied the similar table to improve the search efficiency.

1.1.1 RTGS Table

The Royal Thai General System of Transcription is the official system for rendering Thai language words in the Latin alphabet. RTGS published by the Royal Institute of Thailand. It is used in road signs and government publications. Romanization of Thai words is based on sound transcriptionas shown in Fig.1. Both consonants and vowels must be spelled and sorted grammatically correctly.

The purpose of RTGS is to enable a user to read and write Thai words using the Roman alphabet that are the closet to the original Thai ones. Thai romanization could be performed on the basis of transcription, transliteration or both.

Available in English :
`http://www.arts.chula.ac.th/~ling/tts/principles_eng.pdf`

RTGS Table						
Consonant				Vowel		
No.	Thai Alphabet	Romanized Character		No.	Thai character	Romanized Character
		Initial consonant (sound)	Final consonant (sound)			
1	ก	k	k	1	ะ, า, ◌ั	a
	ข ฃ ค ฅ ฆ	kh		2	เ-ะ, เ-	e
2	ง	ng	ng	3	แ-ะ, แ-	ae
3	จ ฉ ช ฌ	ch	t	4	◌ิ	i
4	ซ ส ศ ษ [สร ทร]	s	t		◌ี	i
5	ญ	y	n	5	◌ึ , ◌ื	ue
	ย		-	6	◌ุ , ◌ู	u
6	ด ฎ ฑ	d	t	7	โ-ะ, โ-, เ-าะ, -อ	o
	ต ฏ	t		8	เ-อะ, เ-อ, เ-ิ	oe
	ฐ ฑ ฒ ท ธ ฑ	th		9	เ-ียะ , เ-ีย	ia
7	ณ น	n	n	10	เ-ือะ , เ-ือ	uea
8	บ	b	p	11	◌ัวะ, ◌ัว, -ว-	ua
9	ป	p	p	12	ไ-, ใ-, ไ-ย, ◌ัย, -าย	ai
	ผ พ ภ	ph		13	เ-า, -าว	ao
10	ฝ ฟ	f	p	14	โ-ย, -อย	oi
11	ม	m	m	15	-ิว	io
12	ร	r	n	16	เ-ว, เ-็ว	eo
	ล ฬ	l		17	-วย	uai
13	ว	w	-	18	เ-ย	oei
14	ห ฮ	h	-	19	เ-ือย	ueai
				20	แ-ว	aeo
				21	เ-ียว	iao

Figure 1: RTGS Table

1.1.2 Similar Table

This work reuses the results of a research paper written in Thai[1] .Then we apply to it phonetics mapping of consonants and vowels between Thai and English. We created the similar table (Fig.2) which organizes the groups of similar consonants and vowels together. Moreover, the number of similar characters is increased in order to gain more comprehensibility.

This table is not officially used as an alternative standard. However, it will be useful in reducing search limitations in practice.

Similar Table						
Consonant				Vowel		
No.	Thai Alphabet	Romanized Character		No.	Thai character	Romanized Character
		Initial consonant (sound)	Final consonant (sound)			
1	ก	ck, g, x, c, q	ck, g, x, c, q	1	ะ, า, ั	u, ar
	ข ฃ ค ฅ ฆ			2	เ-ะ, เ-	ay
2	ง	-	-	3	แ-ะ, แ-	air, are
3	จ ฉ ช ฌ	j, x	j, x	4	-ิ	-
4	ซ ส ศ ษ [สร ทร]	s, z	s, z		-ี	ee, ei, ea, ey
5	ญ	-	y	5	-ึ , -ื	eu, eo
	ย		-	6	-ุ , -ู	oo
6	ด ฎ ฑ	dh	d	7	โ-ะ, โ-, เ-าะ, -อ	-
	ต ฏ			8	เ-อะ, เ-อ, เ-ิ	er, ur
	ถ ฐ ฑ ฒ ธ ฒ			9	เ-ียะ , เ-ีย	ie, aiu, ear
7	ณ น	-	-	10	เ-ือะ , เ-ือ	ui
8	บ	-	b	11	-ัวะ, -ัว, -ว-	our
9	ป	bh	-	12	ไ-, ใ-, ไ-ย, -ัย, -าย	ie, uy
	ผ พ ภ			13	เ-า, -าว	ou, ow
10	ฝ ฟ	-	f	14	โ-ย, -อย	oy
11	ม	-	-	15	-ิว	iu, ew, eua
12	ร	-	l	16	เ-ว, เ-็ว	-
	ล ฬ			17	-าย	-
13	ว	v	-	18	เ-ย	-
14	ห ฮ	-	-	19	เ-ือย	-
				20	แ-ว	aew
				21	เ-ียว	ieo, eaw, iow, iau, iew, iaw

Figure 2: Simialr table

1.2   Website

To be continued.

1.3   Ontologies

To be continued.

References

[1] Somchai Prasitjutrakul, Tasanawan Soonklang and Boonserm Kijsirikul. 2000. Thai-english translit- erated word encoding cross-language retrieval using neural networks. In National Comp. Sc. and Engineering Conf.

http://femto-st.fr