

Development of Thai-Japanese e-Dictionary

Krisada Visavateeranon^{#1}, Surapan Meknavin^{#2}, Datchakorn Tancharoen^{#3},
Wanwimon Rungtheera^{#4}, Shuichi SANO^{#5}, Thitisorn Sang-urai^{#6}, Somjai Sae-ung^{#7},
Bandhit Rojarayanont^{#8}, Natthaphon Jarat^{#9}, Saleela Boontrakulsuk^{#10}, Bancha Borikub^{#11}

[#]Thai-Nichi Institute of Technology,

1771/1 Pattanakarn Rd., Suanluang, Bangkok 10250 Thailand

¹Krisada.v@tni.ac.th

Research Institute for Languages and Cultures of Asia and Africa,

Tokyo University of Foreign Studies

3-11-1 Asahi-cho, Fuchu-shi, Tokyo, 183-8534, Japan

and Makoto Minegishi

Abstract-The purpose of this project is to create a Thai-Japanese dictionary for students at Thai-Nichi Institute of Technology (TNI) who are studying Japanese language and technical Japanese words. The dictionary's database initially contains about 50,000 Japanese words which were created from Tokyo University of Foreign Studies (TUFS)'s XML database of Japanese-Thai dictionary, the Technology Promotion Association (Thailand-Japan) (TPA)'s Japanese-Thai Technical words dictionary and Thai-Japanese Tourism dictionary. The structure of the database is divided to multiple fields including Romaji, Kana, Kanji, Thai, English, synonym, sample sentence, parts, field, etymology and style. Our e-dictionary uses the TNI's website for testing the system while at the same time provide service to the students and general Thai and Japanese users. Accessing of Japanese word can be done by entering Romaji, Kana or Kanji. Accessing of Thai words can be done by entering Thai characters. The development of new Thai and Japanese words is done by a group of researchers, TNI-TUFS staffs and students who are involved in proposing, editing and approving of new words. More than 3,000 technical words in Engineering, Information Technology and Business Administration field are being added to the database via TNI web service.

Keywords- e-Dictionary, Japanese-Thai dictionary, online dictionary, Thai-Japanese dictionary

I. INTRODUCTION

The development of a Thai-Japanese electronic dictionary is a joint research between Thai-Nichi Institute of Technology (TNI) and Tokyo University of Foreign Studies (TUFS). TUFS's research program on "Corpus-based Linguistics and Language Education" which is one of the Global Center of Excellent (COE) programs conducted under the auspices of the Japanese Ministry of Education Culture, Sports, Science and Technology (MEXT). The objective of this research project is to create a Thai-Japanese and Japanese-Thai electronic online dictionary for educational purpose and give service to TNI students and TPA's member who are studying Japanese and Japanese people who are learning Thai language. In this paper, we first describe on the digitization of Thai-Japanese dictionary, the detail of e-dictionary database. We will then discuss about the further development of e-dictionary.

II. THAI-JAPANESE ONLINE DICTIONARY

The advent of the internet and the increasing popularity of the web have altered many aspect of language usage. At present, an online dictionary plays an essential role as a tool of understanding and studying foreign language. Many Thai are now learning Japanese and at the same time some Japanese are studying Thai language. The online Thai-Japanese dictionary is helpful for learning of both languages. Some online Thai-Japanese dictionaries are in services, for example Longdo [1], JTDIC [2], TUFS's Thai-Japanese and Japanese-Thai dictionary [3], and Nagaoka University of Technology's Multilingual Terminology Dictionary (NUT-MTD) [4] which is a web-based dictionary for students and engineers.

TUFS has completed creating the XML database of Tomita (2003) [5]. The number of Thai and Japanese headwords is approximately 20,000 each, which should suffice for foreign language learners of both the languages. The Japanese-Thai and Thai-Japanese dictionary is now accessible to the public via the following URLs [6].

<http://cblle.tufts.ac.jp/dic/th/jath/>
<http://cblle.tufts.ac.jp/dic/th/thja/>

As the original dictionaries are designed to help Thai people to learn Japanese, which is possible for Thai people to search for Japanese words using Japanese Hiragana, Kanji, or Romaji. On the other hand, Japanese learners can only search for Thai words in the Thai-Japanese part of the dictionary using Thai characters.

Since the original dictionary is designed to help Thai people to learn Japanese, thus it is not sufficient for Japanese who wish to learn Thai. In order to apply the data to both Thai and Japanese language learning, TUFS plan to add information that is useful for Japanese people, such as pronunciation and grammatical categories of Thai words. By adding information on etymology or loan words from Pali-Sanskrit, Khmer, Chinese, English, etc., is important not only for learning the language, but also for aiding further understanding of the cultural background of the society where the language is spoken.

III. DICTIONARY DATABASE

Creation of TNI dictionary database started from converting TUFS's dictionary which was available in form of a XML data to a database form. As the result from the conversion, two dictionary's databases were created. They are Thai-Japanese dictionary database for Japanese user and Japanese-Thai database for Thai user. The field of database consists of Romaji, Kana, and Kanji to describe Japanese words, Thai and phonetic for Thai words, English meaning of the words to understand Katakana words. Other fields which are essential for the language learner are also included in the database such as synonym, etymology, field, part of speech, style, and sample sentence. The database structure of both database is shown in Fig. 1

More data was added to the two databases by digitizing two TPA dictionaries. They are Japanese-Thai dictionary for technician [7] and Thai-Japanese dictionary for traveling [8]. The electronics data are then added to the dictionary database. The number of words in each dictionary is approximately 25,000 words.

IV. WEB SERVICE OF E-DICTIONARY

TNI e-Dictionary system is developed to be web service of the Thai-Japanese dictionary database for TNI students, TPA members and general users including Japanese users in Japan. Thai-Japanese dictionary is now accessible to the public via the following URLs [9].

Structure of Japanese-Thai dictionary

dic_item ID	Romaji	Kana	Kanji	Thai	English	Synonym	Etymology	Field	Parts	See	Sample sentence
รหัสค่าศัพท์	โรมาจิ	ชิรางานะ	คันจิ	ไทย	อังกฤษ	คำเหมือน	ที่มาของคำ	สาขา	ชนิด	อ้างอิง	ตัวอย่างประโยค
tmjt00049400	anzen	あんぜん	安全	ปลอดภัย	safety			ind	n		安全第一 ปลอดภัยไว้ก่อน
tmjt00049400_00300	anzen-chitai	あんぜん ちたい	安全地帯	เขต ปลอดภัย	safety zone			ind	n		

Structure of Thai-Japanese dictionary

dic_item ID	Thai	Phonetic	Romaji	Kanji (Hiragana)	English	Synonym	Etymology	Field	Parts	Style
รหัสค่าศัพท์	ไทย	การออกเสียง	โรมาจิ	คันจิ	อังกฤษ	คำเหมือน		สาขา	ชนิด	
tmtj00108100	ข้าว		indo-ine	インドイネ	rice			pln	n	
tmtj00108100_00300	ข้าวของ		nimotsu	荷物 (にもつ)	thing	สัมภาระ		comm.	n	

Fig 1. The structure of Thai-Japanese and Japanese-Thai dictionaries.

<http://dict.tni.ac.th>

The user of dictionary can access the database via the internet. The system help Thai student to search for a Japanese word in the dictionary to get the meaning in Thai and English and Japanese users can search to find the meanings of Thai words. The user of the system will be divided into 4 categories: general user, member, approver and system administrator.

General User: user who searches for the meaning of a word.

Member: user who is registered to system and can propose new words or edit the existing words in the dictionary database.

Approver: linguistics specialist who is assigned by the system to approve new words or edited words proposed by member.

System Administrator: person who does administration work, assigns new member and approver, updates the dictionary database, gets all statistic data of the systems usage and improve the system performance.

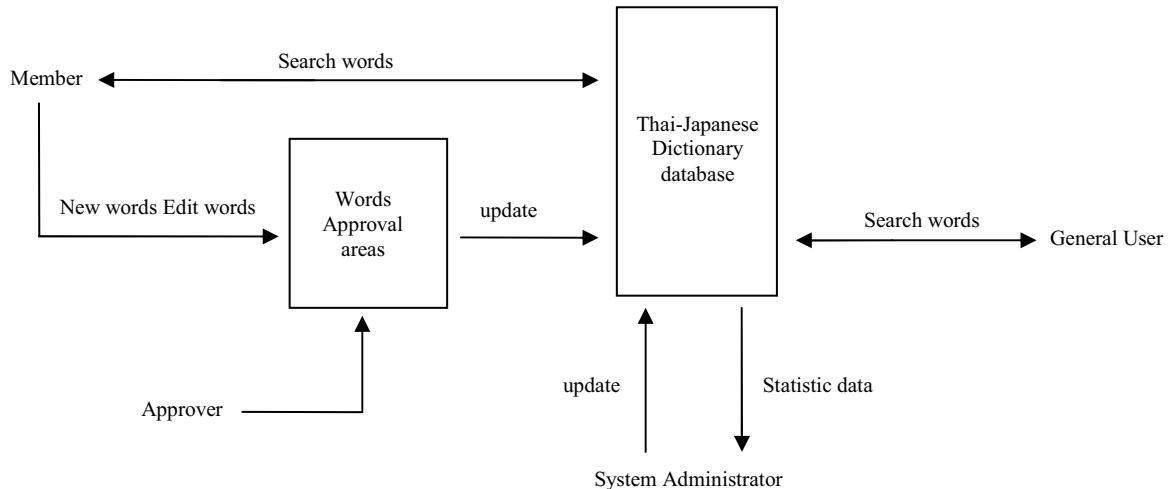


Fig. 2 Users of the TNI's e-Dictionary system

To search for a Japanese word, the word entry can be either Romaji, Hiragana, Katakana or Kanji. Single word and compound word can be searched by a single word entry, all the word will be displayed as the result. Compound words can be searched by using single word entry and select prefix or suffix search. Found words will be displayed on the screen in Romaji, Kana, Kanji, Thai meaning of the words and English meaning. More information of each word can be displayed by simply clicking at the selected word on the screen, the information are field, part of speech, antonym, and sample sentence.

To help the Japanese learner, when entry search in Romaji, all combination of short and long pronunciation will be displayed, for example 'gakuko' searching will

result in 'gakuko' and gakukoo'. Some of the word has same pronunciation but different Romaji writing such as 'ou', 'oo' and 'ee' and 'ei', the e-Dictionary will search and display all cases for the user

New words can be added to dictionary database by member proposing new words or editing existing words in the database. All new words will be recorded in the words approve area waiting to get approval from the approver who are linguistic specialist. The approver may edit the data to get more accuracy of the meaning. When a word is approved, it will be written into the dictionary database for further access by general user. Fig.4 and Fig.5 show the sample screen of adding new word and approve of word by approver.



The screenshot shows the TNI e-Dictionary beta version homepage. The top navigation bar includes links for HOME, ABOUT, CONTACT US, NEWS, LOGIN, LINK, and HELP. The main content area features a search bar with the placeholder 'Enter Japanese or Thai word' and a 'Lookup' button. Below the search bar, a message indicates 14 entries found for 'nichi'. The search results table has columns for Romaji, Kana, Kanji, Thai, and English. The first entry for 'nichi' is shown with its details: Romaji 'nichi', Kana 'にち', Kanji '日', Thai 'นีชี', and English ''. A sidebar on the right provides options for user interaction, such as offering comments on vocabulary, suggesting a translation, and searching other online dictionaries.

Romaji	Kana	Kanji	Thai	English
nichi,Nichi	にち	日	นีชี	-
nichi,Nichi	にち	日	นีชี	-
Vocabulary : nichi,Nichi Etymology : - Fields : - Styles : - Proverb : - Antonyms : - Synonyms : - See also : - Example sentence : -				
Nichi-Bel にちべい 日米 นีชี-บี-เมีย ^ก Nichi-Ei にちえい 日英 นีชี-อี-เก็ต nichibotsu にちぼつ 日没 กลางวัน-นีชี-บ็อกส์ nichigen にちげん 日限 กำหนดเวลา นีชี-เก็น nichiji にちじ 日時 วันเวลา นีชี-จิ nichijo-seikatsu にちじょううせいかつ 日常生活 ชีวิตประจำวัน นีชี-จอยู-เซิกคัทสุ nichijono shigoto にちじょうのしごと 日常の仕事 ภารกิจประจำวัน นีชี-จอยู-ชิกก็อต nichinichiso ニチニチソウ ニチニチソウ นีชี-นีชี-โซว nichivva にちふわ 日暮 ช่วงเวลาที่แสงอาทิตย์匿かれて นีชี-ฟู-วา				

Fig. 3 Example of Japanese word search result

Fig. 4 Example of adding new words by a member

Currently the beta version of TNI e-Dictionary system is on service to general users since mid 2009. The number of search is roughly 1,000 searches per day. Statistic data on the usage of the system will be simultaneously reported when requested by the member. The reported data are including the latest searches, popular words, unfound words and members giving most contribution. By investigating these data, new words will be added to the system and the new service for the user and member will be improved and developed.

V. FUTHER DEVELOPMENT.

In this paper, we have described the efforts at the initial stage to develop a Thai-Japanese e-Dictionary and put in service to test the performance of the system and develop the method to input more words to the database. The purpose of the project is to create a Thai - Japanese and Japanese-Thai electronic online dictionary for educational purpose and give service to TNI students and TPA's member who are studying Japanese and Japanese people who are learning Thai language. Further development is to add more words especially technical

No.	Romaji	Kana	Kanji	Thai	English	Date	Created by	Select All	Edit
1	kaizen-suru	かいぜんする	改善する	ปรับปรุง, แก้ไข	improvement	2009-09-30 11:11:02	Krisada	<input type="checkbox"/>	
2	shitsuke-suru	しつけする	しつけす	กิริ埃, สำอางค์	discipline	2009-09-30 11:09:56	Krisada	<input type="checkbox"/>	
3	seiso-suru	せいそうする	清掃す	ทำความสะอาด	cleaning	2009-09-30 11:07:58	Krisada	<input type="checkbox"/>	
4	seiri-suru	せいりする	整理す	ส่วนตัว, จัดให้เรียบร้อย	keep things in order	2009-09-30 11:06:56	Krisada	<input type="checkbox"/>	
5	seiton-suru	せいとんする	差鋪す	จัดให้เรียบร้อย, จัดให้秩序整頓	tidying up	2009-09-30 11:03:33	Krisada	<input type="checkbox"/>	
6	seketsuna	せいけつな	清潔な	สะอาด, กระถางต้นไม้	keep things clean	2009-09-30 11:01:44	Krisada	<input type="checkbox"/>	
7	monozukuri	ものづくり	物作り	วิถีกรรมการผลิตแบบญี่ปุ่น, การสร้างสรรค์สิ่งของ	Japanese art of manufacturing	2009-09-30 10:53:04	Krisada	<input type="checkbox"/>	

Fig. 5 Example of approving new words or edited words

words in the field of Engineering, Information Technology and Business Administration which are the major of students at TNI. The ultimate objective of the project is to promote language education and linguistic analysis in both Thailand and Japan. Since the demands of linguists, language educators, and language learners are going to continue increasing, more cooperation between Thai and Japanese Universities in making the Thai-Japanese and Japanese-Thai dictionary is necessary.

REFERENCES

- [1] Londo, <http://www.longdo.com>
- [2] JTDIC, <http://www.jtdic.com/jtdic.aspx>
- [3] TUFS, <http://www.cblle.tufs.ac.jp/dic/th/jath/>
- [4] NUT-MTD, <http://gii.nagaokaut.ac.jp/mtd/>
- [5] Tomita Takejirō (2003): Japanese-Thai and Thai-Japanese Dictionary for Thai People, Revised Edition, 367pp+270pp, Bangkok: Association for Thai Education and Cultural Promotion.
- [6] Minegishi Makoto and Akagi Osamu (2009): Development of and Electronic Dictionary for Analyzing Linguistic Data.. Proceedings of Chula-Japan Linguistics Conference 2008, pp.11-16, Tokyo University of Foreign Studies, Tokyo
- [7] Klanhan Voraputiporn, Phulporn Saenbanpra (1990): Japanese-Thai Technical words dictionary. Technology Promotion Association (Thai-Japan) press.
- [8] Pranee Chongsucharitkul (2006): Thai_Japanese Tourism words dictionary. Technology Promotion Association (Thai-Japan) press.
- [9] TNI e-Dictionary: <http://dict.thainichi.net>