2011 International Conference on Network and Electronics Engineering
IPCSIT vol.11 (2011) © (2011) IACSIT Press, Singapore

Thai and English Electronic Dictionary for Mobile Phones

Chakrit Saengkaew
Mahasarakham University
Mahasarakham, Thailand
chakrit.s@msu.ac.th

Abstract. The world today can communicate in a wide dimension. There is a connection among groups of people of different races. One obstacle of the communication is the problem about language. Even though there are many courses in English language in Thailand, there are still a lot of Thai people who cannot communicate in English. The developer recognized this importance and decided to develop a tool which can help people (Thais and foreigners) connect. Mobile phones are widely-used devices and they have a lot of potentials in becoming an electronic dictionary (both Thai-English and English-Thai) so that mobile phone users can look for words and translations. This way, Thais and foreigners can study, look up, understand, widen opportunities and reduce the gaps during communication. Programs and databases on mobile phones do not need to be accessed online because users can look up Thai and English words on their own mobile phones.

Keywords: Electronic dictionary, Thai language, mobile phone, Python DBMS, mobile application

1. Introduction

English is considered the universal language for communication between people in the world who use different languages. In Thailand, there is a communication between Thais and foreigners all the time. During the moment when the word is not recognized, we tend to use a dictionary [1] to look up the definitions or the foreign words. It takes time to look up the words online [2] because there is a restriction about the computer size and the internet connectivity. There are electronic dictionaries available but they are quite expensive. The developer decided to build a dictionary on a mobile phone because mobile phones are devices widely used. The dictionary which the researcher developed can translate Thai into English and English into Thai. Thai is a unique language in that there are 44 consonant letter and 32 vowels along with 4 tone marks. Some consonants can work as final consonants and vowels. Some vowel forms require other vowel forms so that they become words. Therefore, dictionaries on mobile phones are important for Thais and foreigners during communication.

This research aimed at designing and building dictionaries on mobile phones. The program has 2 main components that are database for vocabulary and program to look up words. The database consisted of vocabulary, translations. The data were stored in the form of Table (Python DBMS). The program to look up words interacts with users and is used to search for words in terms of SQL on mobile phone display. This article will explain the procedures and summarizes in the last section.

2. Literature Review

Yuen Poovarawan [3] did research on analysis of Thai words and built a dictionary by choosing randomly word and sentence examples extracted from books, newspaper, journals, magazines, letters, official letters and general reading books, excluding books about literature and academic texts translated from other languages. Therefore, words from foreign countries are not included in LEXiTRON [4], an online dictionary developed by Nation Electronics and Computer Technology Center or NECTEC, and Nation

Science and Technology Development Agency or NSTDA. This dictionary contains frequently used words in documents. The current database consists of 79,000 English-Thai translations and 51,000 Thai-English translations [5]. However, LEXiTRON is limited to those who use can access the internet, resulting in limitations for mobile phone users who cannot access the internet. This project was initiated to make LEXiTRON available for mobile phones. At the same time, the accuracy of the data is still the same as the database. The contents are the same. Only the format is changed. The overall picture of the research is shown in Figure 1.

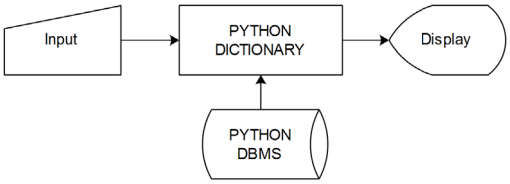

Figure 1 shows the overall idea of Thai and English electronic dictionary for mobile phones

3. Converting Vocabulary Database

Converting Thai and English vocabulary database required programming and working on microcomputers via Python. The database of LEXiTRON [4] which is called LEXiTRON DB was distributed in the form of XML and then converted to Python DBMS format. The advantage of this was that the file size was smaller because only the necessary information was extracted for development. The developers used Python (PYS60 V1.4) [6] because Python is stable and has a lot of libraries. It is also platform-independent and freeware (Python, RDBMS). The procedures are shown in Figure 2.

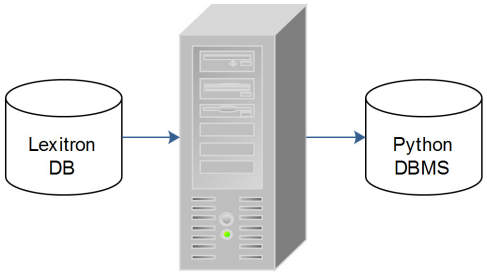

Figure 2 shows the procedure of converting database on microcomputer

The vocabulary database was converted from XML format which is specific like

<Doc><esearch>yak</esearch><eentry>yak1</eentry><tentry>ว้วป่าขนยาวในทิเบต</tentry><ecat>N</ecat>
<id>82946</id></Doc>

4. Converting Vocabulary Database

The XML path (XPath) is very important for conversion. However, XML format needs XML element, therefore, this kind of data was also stored in the file, resulting in larger file size. This is the limitation of mobile phones in terms of memory. The developer decided to code a program to extract only the important parts of the data in the database to build a new database (Python DBMS), collecting only vocabulary and translations to reduce the size of the dictionary. The procedure is shown in Figure 3.

The principle of program is to Match on regulate expression by instruction like

*Select * from et where esearch like "doc";*

In order to compare the words in the XML-formatted database. The example looked up the word "doc" which divides the data into groups and is used along with other keywords. The important data are organized in the table of database in the format of Python as shown in Figure 4..

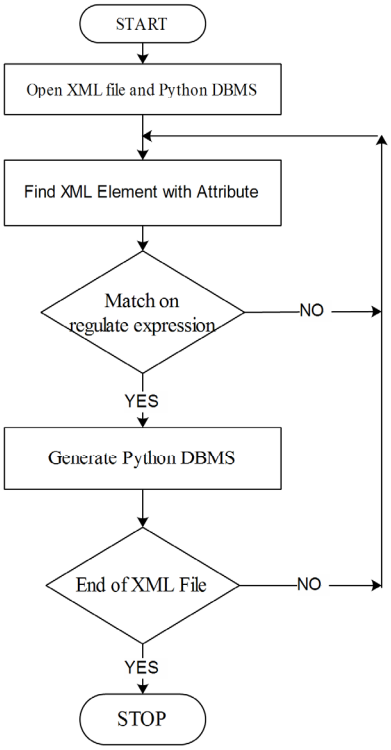

Figure 3 shows the procedure of converting XML DB to Python DBMS

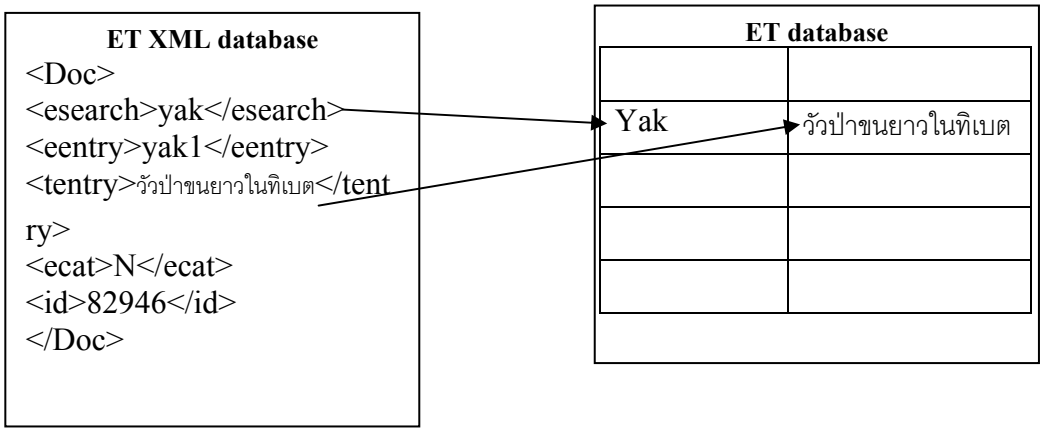

Figure 4 shows example table in Python DBMS

A sample dictionary database after converting XML DB to Python DBMS.

5. Program to look up words

The program will show the display and wait for users to put the words to look up. Users can type via keyboard. They can choose between Thai and English or they can use standard keys on mobile phones. Users press “enter” and the program will look up the words then show the data about the words on the screen of the mobile phones.

The dictionary program depends on the program to look up words in the database which was converted from XML. The word searching feature is based on SQL which uses e32db.Dbms module in Python programming language because this module is suitable for the database which need little revision.

The module will look up words from the TE table when users want to translate from Thai into English and will look up words from the ET table when users want to translate from English into Thai.

The developer use Thai Unicode encoding so that it can be used with the Thai font on mobile phones.

Sample code to store data on DBMS database

```
import e32db,time
dbms = e32db.Dbms()
dbms.open(u'E:\\sample.db')
query = u"CREATE TABLE et(ESEARCH VARCHAR, EENTRY VARCHAR, TENTRY VARCHAR, ECAT VARCHAR,ID INTEGER, ETHAI VARCHAR, ESYN VARCHAR, EANT VARCHAR)"
rtn = dbms.execute(query)
print rtn
insert = u"INSERT INTO et VALUES ('a','a','%s','DET',0,",",")"%(thai.thaicode("หนึ่ง (คำนำหน้าคำนามเพื่อแสดงว่าคำนามนั้นๆ ไม่ชี้เฉพาะ)"))
dbms.begin()
count = dbms.execute(insert)
dbms.commit()
```

Sample code to retrieve data from DBMS database to display on the screen

```
dbv = e32db.Db_view()
dbv.prepare(dbms, u"SELECT TENTRY FROM et")
dbv.first_line()
dbv.get_line()
vocab = dbv.col(1)
appuifw.note(vocab)
dbms.close()
```

6. Test results

In the test, the word “a” was to be looked up on the mobile phone display. After “enter” was pressed, the program would show the translations on display that is “หนึ่ง (คำนำหน้าคำนามเพื่อแสดงว่าคำนามนั้นๆ ไม่ชี้เฉพาะ” as shown in Figures 6 and 7.

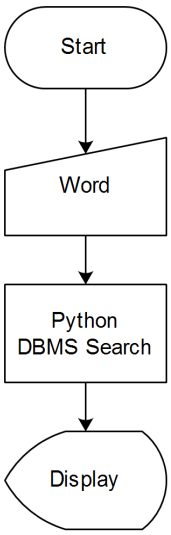

Figure 5 Procedure of Word Searching Program

Figure 6,7 shows the display waiting for word input and shows the display with word searching results

7. Suggestions for further development

The words which were not available in the dictionary were normally proper nouns, for example, names of persons or places, new words and words from foreign countries. In the future, the program will be developed to include new words. The translation for sentence can be done if the sentence undergoes Thai word segmentation by using dictionary [7] because this dictionary is to look up only words in the dictionary. In the future, there might be a program which can perform Thai word segmentation by rules [8] to check the rules about Thai language in terms of mixing letters, spacing, and beginning a new paragraph. This research can be further developed by giving sounds for the word which is looked up as well.

8. References

[1] Non-Royal Institute Thai Dictionary. 2001. Bangkok: Matichon Press.

[2] Longdo, http://www.longdo.com

[3] Yuen Poovaravan. 1985. An Analysis of Thai Word Data. Microcomputer Lab. Department of Electrical Engineering, Faculty of Engineering, Kasetsart University.

[4] LEXiTRON, http://lexitron.nectec.or.th

[5] Kanokorn Trakultaweekoon, Peerachet Porkaew and Thepchai Supnithi. 2007. LEXiTRON Vocabulary Suggestion System with Recommendation and Vote Mechanism, SNLP2007 The Seventh International Symposium on Natural Language Processing, Thailand, Dec 2007, pp. 43-48.

[6] Python, www.python.org

[7] Yuen Poovaravan and Vivat Im-aram. 1986. Dividing Thai Syllables by Dictionary. Proceedings on the Ninth Electrical Engineering Conference, Khon Kaen.

[8] Kooptiwoot C. 1999. Segmentation of Ambiguous Thai Words by Inductive Logic Programming.Bangkok : Chulalongkorn University.