Thai and English Electronic Dictionary for Android Mobile Phones Using Python Language

Chakkrit Saengkaew

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. Android 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 android 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 android mobile phones do not need to be accessed online because users can look up Thai and English words on their own android mobile phones.**

Index Terms—**Android mobile phone, Android operating system, electronic dictionary, mobile application, NECTEC Lexitron, Python Language, SQLite DBMS, thai language, SL4A.**

I. 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 an android mobile phone because android 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 android mobile phones are important for Thais and foreigners during communication.

This research aimed at designing and building dictionaries on android 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 (SQLite DBMS) and interface using python language on SL4A layer running on android mobile phones. The program to look up words interacts with users and is used to search for words in terms of SQL on android mobile phone display. This journal will explain the procedures and summarizes in the last section.

II. 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 android mobile phone users who cannot access the internet. This project was initiated to make LEXiTRON available for android 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.
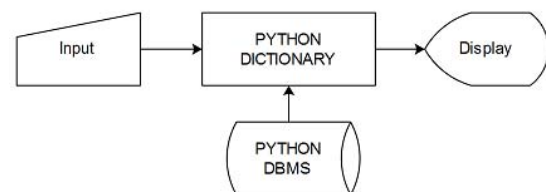

Fig. 1. Shows the overall idea of Thai and English electronic dictionary for android mobile phones

Today's smartphones support an abundance of different programming options. The researcher focused our work on the Android platform, as it is still the most prominent platform by far [6], with the largest number of available devices.

However, despite today's abundance of feature-rich mobile phone hardware and powerful software platforms, creating applications that leverage the platforms' potential is

still a time consuming process that challenges non-expert developers by requiring in-depth know-how. [7]

With the rise of the iPhone and the Android platform released by Google, the share of smart phones has been constantly growing[8]. A key advantage of smart phones is the ability to access the Internet while on the go [9]. Interaction wise however, smart phones still fight a number of teething problems, including the cumbersome entry of data.

While a number of scripting languages are available for mobile phones, there is usually a drawback involved. Either they are still in early development (Python[10], Perl[11], JRuby[12], Lua[13],php[14]).

III. 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 SQLite DBMS format. The developers used Python Language running on SL4A Layer of the android environment [10] because Python is stable and has a lot of libraries. It is also platform-independent and freeware. The procedures are shown in Figure 2.
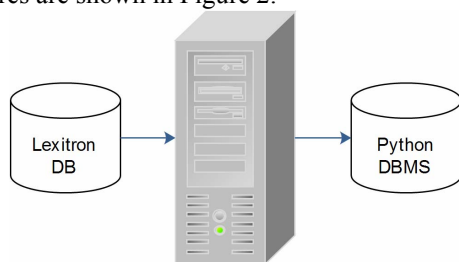

Fig. 2. Shows the procedure of converting database on microcomputer

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

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

The full source code for convert English Thai XML Format to SQLite DBMS which is specific like

```
>>> t =
open("/home/dictionary/etlex","r").read().split("""</Doc>
<Doc>""")
>>> fp = open("/home/dictionary/sqlite_et","w")
>>> for i in t[:]:
    a=i.split("\n")
    vocab = {}
    for j in a:
        tag = j.split(">")[0][1:]
        if tag != "Doc":
            data = j.replace("<"+tag+">", "").replace("</"+tag+">",
"")
            vocab[tag]=data
    k=vocab.keys()[1:]
    v=vocab.values()[1:]
    k = "\\"" + joinfields(k,"\\",\\"") + "\\""
    v = "\\"" + joinfields(v,"\\",\\"") + "\\""
    fp.write("cursor.execute(\"INSERT INTO etlex(%s) VALUES
(%s);\")\n"%(k,v))
>>> fp.close()
```

The full source code for convert Thai English XML Format to SQLite DBMS which is specific like

```
>>> t =
open("/home/dictionary/telex","r").read().split("""</Doc>
<Doc>""")
>>> fp = open("/home/directionary/sqlite_te","w")
>>> for i in t[:]:
    a=i.split("\n")
    vocab = {}
    for j in a:
        tag = j.split(">")[0][1:]
        if tag != "Doc":
            data = j.replace("<"+tag+">", "").replace("</"+tag+">",
"")
            vocab[tag]=data
    k=vocab.keys()[1:]
    v=vocab.values()[1:]
    k = "\\"" + joinfields(k,"\\",\\"") + "\\""
    v = "\\"" + joinfields(v,"\\",\\"") + "\\""
    fp.write("cursor.execute(\"INSERT INTO telex(%s) VALUES
(%s);\")\n"%(k,v))
>>> fp.close()
```

IV. CONVERTING VOCABULARY DATABASE

The split function is very important for conversion. However, XML format needs XML element, therefore, this kind of data was also stored in the SQLite database file. The developer decided to code a program to extract the all parts of the data in the xml database to build a new database (SQLite DBMS). The procedure is shown in Figure 3.

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 SQLite as shown in Figure 4.
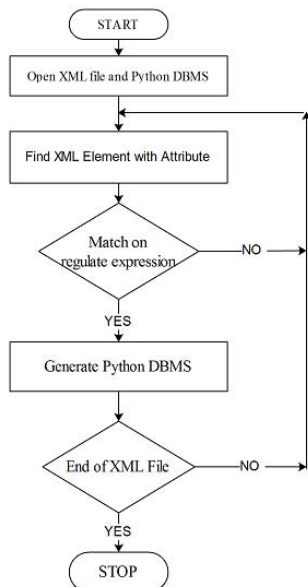
Fig. 3. Shows the procedure of converting XML DB to SQLite DBMS

Sample source code for Insert English Thai XML Format to SQLite DBMS which is specific like

```
# -*- encoding: utf-8 -*-
import sqlite3

DB_NAME = '/mnt/sdcard/etlex.db'

database = sqlite3.connect(DB_NAME) # Create a database file

cursor = database.cursor() # Create a cursor

cursor.execute("CREATE TABLE IF NOT EXISTS etlex(id
VARCHAR(20),esearch VARCHAR(255),eentry
VARCHAR(255),tentry VARCHAR(255),ecat VARCHAR(255),ethai
VARCHAR(255),esyn VARCHAR(255),eant VARCHAR(255))")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"tentry\",\"id\",\"ecat\") VALUES
(\"a\",\"a\",\"หนึ่ง (คำนำหน้าคำนามเพื่อแสดงว่าคำนามนั้นๆ
ไม่ชี้เฉพาะ)\",\"0\",\"DET\");")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"tentry\",\"id\",\"ecat\") VALUES
(\"A\",\"A\",\"อักษรตัวแรกในภาษาอังกฤษ\",\"1\",\"N\");")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"tentry\",\"id\",\"ecat\") VALUES
(\"a\",\"a\",\"อักษรตัวแรกในภาษาอังกฤษ\",\"2\",\"N\");")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"tentry\",\"id\",\"ecat\") VALUES (\"a
posteriori\",\"a posteriori\",\"จากผลไปสู่เหตุ\",\"3\",\"ADJ\");")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"ethai\",\"tentry\",\"id\",\"ecat\")
VALUES (\"A.B.\",\"A.B.\",\"ศิลปศาสตรบัณฑิต\",\"อักษรศาสตรบัณฑิต (คำย่อของ
Artium Baccfalaureus เท่ากับ Bachelor of Arts)\",\"4\",\"ABBR\");")
```

```
cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"ethai\",\"tentry\",\"id\",\"ecat\")
VALUES (\"A.D.\",\"A.D.\",\"ค.ศ.\",\"ปีคริสต์ศักราช (คำย่อของ Anno
Domini)\",\"5\",\"ABBR\");")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"tentry\",\"id\",\"ecat\") VALUES
(\"zygotic\",\"zygotic\",\"เกี่ยวกับ zygote\",\"83230\",\"ADJ\");")

cursor.execute("INSERT INTO
etlex(\"esearch\",\"eentry\",\"ethai\",\"tentry\",\"id\",\"ecat\")
VALUES
(\"zymurgy\",\"zymurgy\",\"การกลั่นสุรา\",\"การหมักสุรา\",\"83231\",\"N\")
;")

database.commit()
database.close()
print "สมบูรณ์"
```
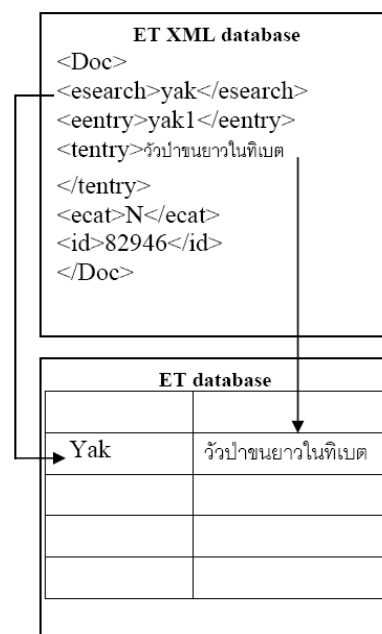

Fig. 4. Shows example table in SQLite DBMS

Sample source code for Insert Thai English XML Format to SQLite DBMS which is specific like

Thai English Vocaburary for Sqlite DBMS
```
# -*- encoding: utf-8 -*-

import sqlite3

DB_NAME = '/mnt/sdcard/telex.db'

database = sqlite3.connect(DB_NAME) # Create a database file

cursor = database.cursor() # Create a cursor

cursor.execute("CREATE TABLE IF NOT EXISTS
   telex(tsearch VARCHAR(255),
   tentry VARCHAR(255),
   eentry VARCHAR(255),
```

```
tcat VARCHAR(255),
tsyn VARCHAR(255),
tsample VARCHAR(255),
id  VARCHAR(20),
tant VARCHAR(255),
tdef VARCHAR(255),
tenglish VARCHAR(255),
tnum VARCHAR(255),
notes VARCHAR(255))")
```

```
cursor.execute("INSERT INTO
telex(\"tcat\",\"tsample\",\"tentry\",\"tsyn\",\"tsearch\",\"id\",\"eent
ry\") VALUES (\"PRON\",\"หน่วยงานของเราสามารถรับบทบาทได้เป็นอย่างดี
ตามสภาพความพร้อมด้านต่างๆ
ดังกล่าวข้างต้น\",\"ดังกล่าวข้างต้น\",\"ดังกล่าว\",\"ดังกล่าวข้างต้น\",\"0\",\"abovemen
tioned\");")
```

```
cursor.execute("INSERT INTO
telex(\"tenglish\",\"eentry\",\"tdef\",\"tentry\",\"tsearch\",\"tcat\",\"
id\",\"tsyn\",\"tsample\") VALUES (\"according to the law;
legally\",\"according to
law\",\"ถูกต้องตามกฎหมาย\",\"ตามกฎหมาย\",\"ตามกฎหมาย\",\"ADV\",\"2\",\"โด
ยชอบด้วยกฎหมาย\",\"หนังสือลาออกของนายพรพจน์ยัง ไม่ทันมีผลตามกฎหมาย
เรื่องราวก็จบลงเสียก่อน\");")
```

```
cursor.execute("INSERT INTO
telex(\"tnum\",\"tcat\",\"eentry\",\"tentry\",\"tsyn\",\"tsearch\",\"ta
nt\",\"id\",\"tsample\") VALUES (\"ชนิด\",\"N\",\"advanced
technologies\",\"เทคโนโลยีสมัยใหม่\",\"เทคโนโลยีทันสมัย\",\"เทคโนโลยีสมัยใหม่\",\
"เทคโนโลยีรุ่นเก่า\",\"3\",\"คอมพิวเตอร์เป็นปัจจัยสำคัญในการพัฒนาเทคโนโลยีสมัยใหม่เกือบทุ
กชนิด\");")
```

```
cursor.execute("INSERT INTO
telex(\"tnum\",\"tcat\",\"eentry\",\"tdef\",\"tentry\",\"tsyn\",\"tsearc
h\",\"id\",\"tsample\") VALUES
(\"วาระ\",\"N\",\"agenda\",\"ลำดับรายการการประชุมที่กำหนดไว้\",\"ระเบียบวาระการ
ประชุม\",\"วาระการประชุม,
ระเบียบวาระ\",\"ระเบียบวาระการประชุม\",\"4\",\"ระเบียบวาระการประชุมนี้จะผูกพันให้ทั้งสอง
ฝ่ายต้องเจรจา เฉพาะในเรื่องที่อยู่ในระเบียบวาระการประชุมนั้นเท่านั้น\");")
```

```
database.commit()
database.close()
print "สมบูรณ์"
```

V. 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 android 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 android 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 SQLite. The SQLite Dbms interface with Python programming language because this dbms 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 character set on android mobile phones.
Sample code to store data on DBMS database

```
import sqlite3
DB_NAME = '/mnt/sdcard/telex.db'
cursor = database.cursor() # Create a cursor
cursor.execute(INSERT_COMMAND)
database.commit()
database.close()
```

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

```
# -*- encoding: utf-8 -*-
"""โปรแกรมนี้พัฒนาโดย อ.จักรกฤษณ์ แสงแก้ว คณะวิทยาการสารสนเทศ
มหาวิทยาลัยมหาสารคาม เขียนด้วยภาษาไพธอน (ไม่เข้ารหัสไบต์โค๊ด)
เพื่อให้ผู้ศึกษาทีหลังสามารถเรียนรู้ได้อย่างรวดเร็ว"""
import android,time
import sqlite3
droid = android.Android()
droid.makeToast('English Thai Electronic Dictionary!')
droid.vibrate(300)
VocabList = []

DB_NAME = '/mnt/sdcard/etlex.db'
database = sqlite3.connect(DB_NAME) # Open the database file
cursor = database.cursor() # Create a cursor
# =================================
# id ,esearch ,eentry ,tentry ,ecat ,ethai ,esyn ,eant
# attributes = ['id' ,'esearch' ,'eentry' ,'tentry' ,'ecat
# ,'ethai' ,'esyn' ,'eant']
# =================================
vocab = droid.dialogGetInput("พจนานุกรมไทยอังกฤษ-อังกฤษไทย:",
"กรุณาป้อนคำศัพท์ : ").result
droid.dialogDismiss()
cursor.execute("SELECT * FROM etlex WHERE esearch like '%" +
vocab + "%' LIMIT 0,20") # Select everyone in the table
results = cursor.fetchall()
result=""
for entry in results:
    result = entry[1].encode('utf-8') + " " + entry[3].encode('utf-8')
    if (type(entry[3]) == type("str")):
        result += ("(th) " + entry[3].encode('utf-8'))
    if (type(entry[4]) == type("str")):
        result += ("(cat) " + entry[4].encode('utf-8'))
    if (type(entry[5]) == type("str")):
        result += ("(ethai) " + entry[6].encode('utf-8'))
    if (type(entry[6]) == type("str")):
        result += ("(syn) " + entry[6].encode('utf-8'))
    if (type(entry[7]) == type("str")):
        result += ("(ant) " + entry[7].encode('utf-8'))
    VocabList.append(result)
```

VI. TESTING RESULTS

In the test, the word "com" was to be looked up on the

android mobile phone display. After "enter" was pressed, the program would show the list of vocabulary on display and user can click on the checkbox list for the translation on the screen as shown in Figures 6,7,8 and 9.
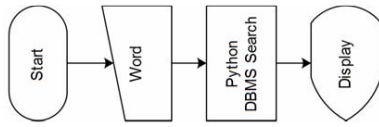

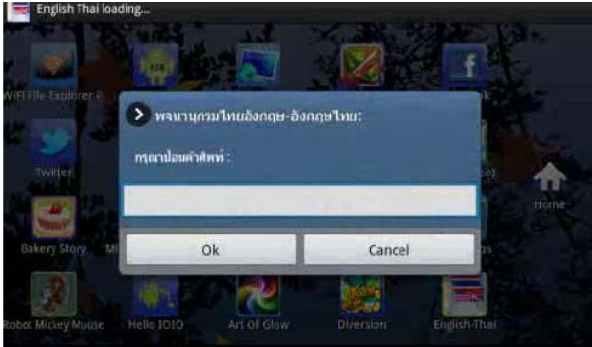Fig. 5. Procedure of Word Searching Program

Fig. 6. Shows the display waiting for word input.

Fig. 7. Shows the display waiting for word input with the large size of screen display.
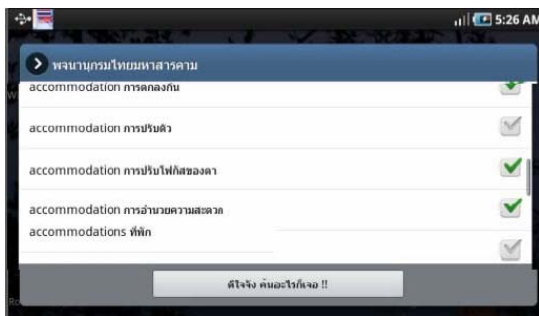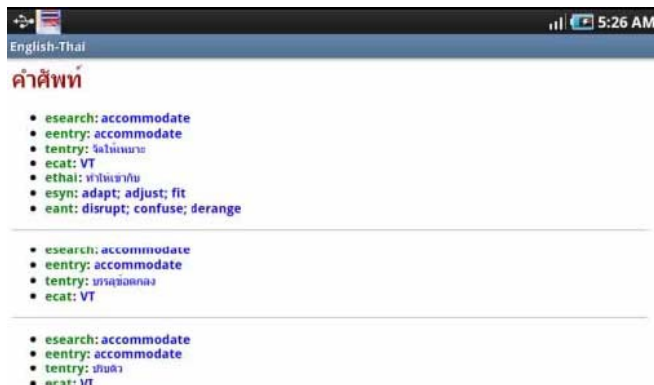

Fig. 8. Shows the display list of the words.

Fig. 9 Shows the display with word searching results

VII. 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 [15] 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 [16] 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.

VIII. REFERENCE

[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] Gartner Report on Smartphone Sales, www.gartner.com/it/page.jsp?id=1622614
[7] Huebscher, M., et al.: Issues in Developing Ubicomp Applications on Symbian Phones. In: Proc. FUMCA'06. *IEEE Computer Society*, 51-56 (2006)
[8] R. Want, "When cell phones become computers." *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 2-5, April-June 2009.
[9] R. Ballagas, J. Borchers, M. Rohs, and J. Sheridan (2006): The smart phone: a ubiquitous input device, *IEEE Pervasive Computing*, vol. 5, no. 1, pp. 70–77, January-March 2006.
[10] Python for Android Devices, www.code.google.com/p/python-for-android
[11] Perl for Android Devices, www.code.google.com/p/perldroid/
[12] JRuby for Android Devices, www.code.google.com/p/jruby-for-android/
[13] Lua for Android Devices, www.code.google.com/p/android-lua/
[14] php for android devices, www.code.google.com/p/php-for-android/
[15] Yuen Poovaravan and Vivat Im-aram. 1986. Dividing Thai Syllables by Dictionary. *Proceedings on the Ninth Electrical Engineering Conference*, Khon Kaen.
[16] Kooptiwoot C. 1999. Segmentation of Ambiguous Thai Words by Inductive Logic Programming.Bangkok : Chulalongkorn University.

Chakkrit Saengkaew is a lecturer at department of Information Science, Faculty of Informatics, Mahasarakham University Thailand. He received the B.Sc. degree in Information Technology from Faculty of Science, King Mongkut's University of Technology Thonburi in 1997. He obtained his M.S. Ind. Ed. degree in Computer and Information Technology from Faculty of Technical Education, King Mongkut's University of Technology Thonburi (KMUTT) Thailand in 2004. During 1997-2004, he works in the field of Computer and Information Technology. He has experience in many positions such as Programmer, System Engineer and Project Manager. He is an author of the book "Python Programming Language by yourself" – best seller computer programming book in Thailand (year 2005).