

Chinese-Thai-English Translation Audible Electronic Dictionary Design and Implementation

R. LIN & J.M. WANG* & B.Z. LI & C.S. Yuan & F. LIU

School of Electrical Information Engineering, Yunnan MinZu University. Kunming, China, 650500

Yunnan Province colleges minority language information processing engineering research center, Yunnan MinZu University. Kunming, China

Corresponding author: J.M. WANG

ABSTRACT: This software is complete by Visual C++ 6.0 and QT, It designed in the Unicode character set patterns , Contribute to ASEAN's cooperation and exchanges, It's solve the problem that system use compatibility and character output garbled in current national language software development. This development model is simple use, stable operation, flexible interface, convenient in user for vocabulary and voice database unified processing (backup, print), at the same time also provides technical guidance to other national language text translation software development. Thai Wen Chinese Translation Electronic dictionary is an important innovation in the field of Dai information technology, providing convenience for "the Belt and Road Initiative" policy. It's the basic support of starting research about minority language cultural information element representation and extraction. And the main function is responsible for Thai queries, translation, reading, etc. Thai Wen – Chinese – English Translation Electronic Dictionary designed to achieve the common functions such as Thai-Chinese bilingual translation, Thai people reading and English display. It's also support the thesaurus to add, modify, delete custom actions, it implements the good human-computer interaction function.

KEYWORD: Thai Wen; Electronic Dictionary; Thai Chinese English translation; Unicode

1 INTRODUCTION

With the implementation of "The Belt and Road Initiative" Policy, Increased cross-border cooperation in the ASEAN region, the field of information processing is broadening and progressing, so that more people are beginning to learn and study Thai. Traditional Thai learning aids are mostly books, computers and internet, learning on the computer and network accessibility are rare. Thai is Thailand's native language, Thai is also used in a few areas of China, Thai and Zhuang language in China belong to the same language family, they also belong to the Sino-Tibetan, and they are widely used in all walks of life. With the rise of government online project implementation and electronic commerce, It is also a sharp increase in demand for Chinese and Thai translation. Due to the various aspects of information, New term appear constantly, the traditional human translation can not meet the majority of the people increasingly demand access to information, All levels of government departments, enterprises, institutions of the translator and the masses generally hope to develop an intelligent machine translation system or bilingual electronic dictionary can translate in real-time translation on the computer. Compared with the level of development of domestic electronic dictionary, Chinese-Thai translation electronic dictionary research

started relatively late. In recent years, Thai news website constantly available on the internet and Thai input method in the research and development, the new era of Thai informatization development. Due to the Thai machine translation research has just started, In a short period of time to develop in the field of different levels, different users have satisfactory full text translation software too difficult. But we need to collect large vocabulary, the online electronic dictionaries which have more specification translations can solve the masses of users to browse the web and online writing difficulty at last. No matter from the users or from the development present situation. The most realistic approach is needed to resolve translation software, this is Chinese-Thai translation software and auxiliary software.

Thai (ภาษาไทย) , also known as Dai language, belongs to the Sino-Tibetan. About 68 million people worldwide use Thai, mainly distributed in Thailand, Laos, Myanmar, Vietnam, northwest, northwest Cambodia, southwest China, northeast India and Dai. This topic is based on conducting research in Bangkok, Thailand. We analyzed the standard Thai, Thai information processing its technical difficulty is how to implement in computer, Thai query, translation and

pronunciation problems. The development and application of the electronic dictionary will be the key to solve the problem.

2 THEORETICAL VALUE

Yunnan is a province which is multi-ethnic and multilingual. Although the language software development is a case study, with the study of new discoveries and endangered to the comprehensive and in-depth. Its development value, the demonstration value, the theory of value is worth attention.

Thai informatization treatment scheme can be represented as other languages information theoretical guidance, especially for database build vocabulary and pronunciation. (Such as: the character encoding, code conversion) It is on the computer to collect, store, display and print to the language of basic work.

Software development based on QT system to complete, In the Unicode character set mode to software design, this software can be easily transplanted to facilitate other words on the electronic dictionary, this design has increased portability. This electronic dictionary in the early development of other similar national language provides the technical guidance.

In the electronic dictionary thesaurus construction and no unified format requirements. XML, plain text and Access database is the commonly used format for electronic dictionary. This design uses the Access data table creation thesaurus, considers Unicode characters display problem, chooses Access2000 or above, solves the VC6.0 and Access database compatibility problems.

In VC ++ 6.0 development environment, the default character set is MBCS multibyte character set, Former Thai character input software was designed based on the Unicode character set, this will lead to the output of Thai characters can't display properly, In this way. In this way, this software in Unicode character mode to carry on the design, solved the software interface output Unicode characters appear garbled.

Word processing software now used for Thai translation system has yet to launch, So Chinese-Thai translation electronic dictionary which work under the Windows operating platform is an applied innovation. The results will promote the study of Thai Chinese machine translation system.

In the field, Thai coding standard has not explicitly formulated, by this software were first developed for Thai electronic dictionary of improving and optimizing provides the basis. How to make the Thai the technical index of the electronic dictionary has a certain reference value.

3 APPLICATION VALUE

For universities, research institutes teaching of language research and reference tools , for Thai Chinese

translation the translator and the related user provide a convenient electronic assistant. Such as teaching, language contact cases, the Thai branch of study, language comparative studies.

Promote the use of tourism for Thailand. Thailand is famous for "the kingdom of thousand buddhas". The capital Bangkok and Chiang mai is known as the "rose" to peaceful places worth a visit. The electronic dictionary provides convenience for Chinese tourists. Chinese-Thai- English electronic dictionary design, not only can bring fiscal revenue, also can attract more bilingual talents to join in language learning has the development potential of the nation.

It will help the news media promotion, there are a lot of cross-border ethnic groups in China, these peoples neighboring countries share the same root, Friendship from generation to generation, they have the same or similar cultural features and language. National language of news media is an important window they understand Chinese, Use of the unique advantages of Thai media spread national culture, this traditional Chinese seem more approachable

With the popularity of electronic products and multimedia teaching, similar to the Chinese electronic dictionary, Chinese-Thai translation electronic dictionary will be a daily use word processing software. This topic "Chinese-Thai-English translation electronic dictionary" software products can be published in Thai ancient books, literature, academic research, teaching material compilation and construction, as well as school bilingual education in the application. Still can make government departments at all levels and school names, addresses, government documents, envelopes tai han translation pairs of document writing, etc.

4 RESEARCH SUMMARY

4.1 Design goals

This software has Chinese-Thai control lexical retrieval functions. Chinese-Thai controlled vocabulary query and retrieve each other, Supporting Thai real pronunciation, automatic display Thai international phonetic alphabet, The supplementary function of text segmentation, to fully support Unicode Thai system, It can look at any word of this code and thesaurus to add, modify, delete operations and other functions. This software system to follow the humanized design, face reality, close to the user, beautiful and easy, many functions, stable performance, cost-effective. Development promotion of user-oriented to learn Thai students, Learning Chinese ethnic minorities, ethnic language workers, han Thai English translator, Thai learners of computer typing, national workers in text information processing and other services.

4.2 Comparing

At present, the domestic research and development of Thai Chinese translation electronic dictionary was still in the stage of application innovation, the related products are not listed or promotion. To facilitate the realization of the function of the analysis of this software. Our team can only select two software functions such as compared to the reference, and choose the Google translate for validation.

The first software for "Yi -Chinese controlled electronic dictionary of primary and secondary schools", the product is based on - software achievements of key project in Sichuan province department of education. By southwest national university institute of national language information processing research and development in 2005. Main function has a query (precise query) and vocabulary changes. Second software as "Dai-Chinese Translation Audible Electronic Dictionary".

The software by the colleges and universities in Yunnan minority language information processing engineering research center in 2015.

This software is" Chinese-Thai-English Translation Audible Electronic Dictionary", Yunnan province colleges minority language information processing engineering research center performed it. Main function words query editor (precise query) and vocabulary. Query results also support the Thai the phonetic symbol, Thai real reading, Thailand's tourism strategy and Thai learning function. On the whole, the software development of the basic meet

the requirements of daily use, the interface is shown in figure 1.

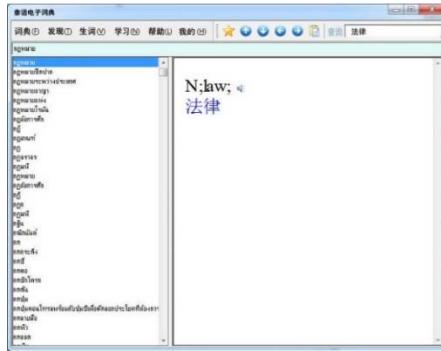


Figure 1 Chinese-Thai-English Translation Audible Electronic Dictionary

4.3 Query to retrieve

General translation dictionary developers can design for the user to provide two types of lexical retrieval in the software, the first one is precise query another is fuzzy query. But the current collection Thai vocabulary and Thai the tape library also lack of perfect, only supports 29945 Thai words commonly used in daily life to retrieve. It joins in the design of the fuzzy query that has little significance and will make inconvenience for the user. After we expand the thesaurus quantity, we will add them. For example, input "law", the software operation interface as shown in figure 2.

Table 1 Three query software function parameters

Query the dictionary types		Yi-Chinese controlled electronic dictionary query interface of primary and secondary schools	Dai-Chinese Translation Audible Electronic Dictionary	Chinese-Thai-English Translation Audible Electronic Dictionary
Software functions	FX	Precise query	Precise query	Precise query
	Query category	Chinese ↔ Yi	Chines↔Dai	Chinese↔Thai↔ English
	Vocabulary editor	Add	CRUD	CRUD
	Real people reading	No	Pronunciation	Pronunciation
	Phonetic annotations	No	Soundmark	Soundmark
	Vocabulary data	No	No	No
	Code conversion	No	Unicode code	Unicode code
	Interface customization	No	No	Color, font
design	Input module	No	Using	Using
	Development of language	Visual C	Visual C++	Visual C++, PHP
	Data format	TXT	Access	Access
	Database	Access2000	Access2003	Access2003

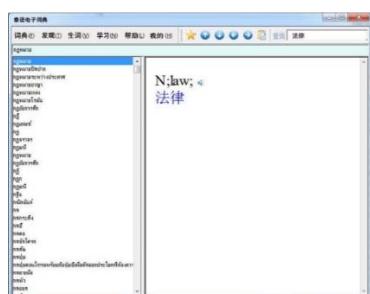


Figure 2 Chinese-Thai-English electronic dictionary operation interface

4.4 Phonetic annotations

The software can be used to query the Thai corresponding phonetic symbol, Phonetic annotation using a common international standard, International phonetic alphabet annotation and human assistance, this greatly solve the Thai beginners which is understanding but not read in the plight of the learning process, when users use the process, phonetic annotations to manage with a single thread class. Therefore, the query of data displayed on the interface rather than

waited for words show that after the completion of the phonetic symbol to display, as shown in figure 3.

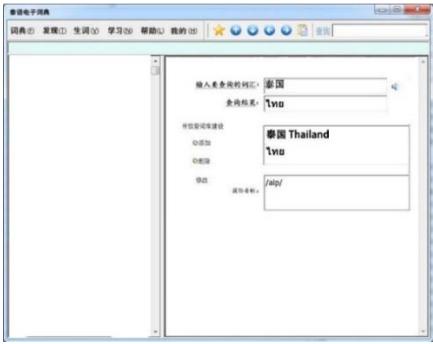


Figure 3 Chinese-Thai-English electronic dictionary phonetic demonstration

4.5 Vocabulary update

Vocabulary CRUD

Acquisition thesaurus and voice is a huge and complicated engineering, this software can provide limited vocabulary, it covers basic Thai commonly words, It meets the demands of general study and work activities, In view of this, the software based on open design, Allows the user increase and supplement their vocabulary, allows the Thai professionals proof-read it and allows the Thai professionals delete errors or repeat words, as shown in figure 4,

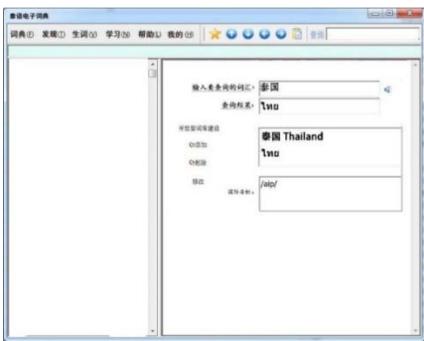


Figure 4 Words to modify schematic diagram
Vocabulary data

In order to facilitate the needs of management and print the query data, the software provides a query data printing function and can put the text data into a TXT, the query results can be printed and convenient user. As shown in figure 5.

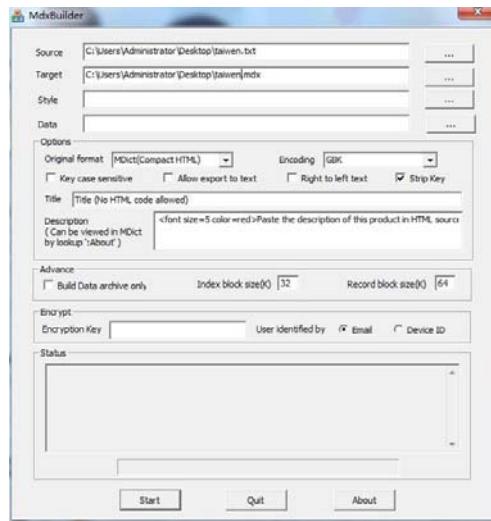


Figure 5 Data into TXT schematic diagram

4.6 Other features show

Discovery

With the progress and development of the economy, Thailand tourism became a big topic of hot debate, more and more Chinese go to Thailand. With the movie " Lost In Thailand " and other films aggressively, " Thailand hot" has become essential for travelers.

The dictionary adds "discovery" feature, we click the "discovery" to see some of the tour. For the majority of Thai tourism enthusiasts to provide convenient. As shown in figure 6

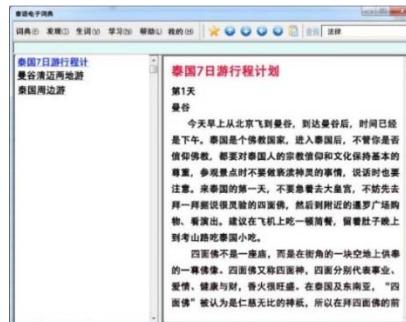


Figure 6 Found the function diagram

Learning function

Learning function is also a feature of the software, learning function in order to facilitate learning Thai and Chinese pronunciation. Learning function is under the order of Chinese Pinyin, arranged thesaurus of words in order to facilitate the use of Chinese learning and Chinese - Thai - English translation. As shown in figure7. Start interface shown in Figure 8.

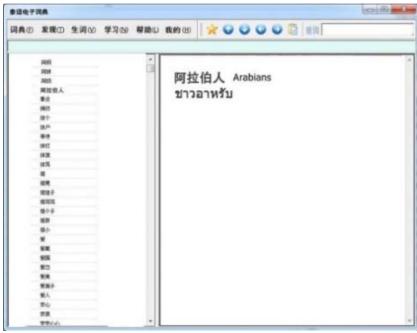


Figure 7 Learning function schematic



Figure 8 Start interface

5 VOICE LIBRARY CONSTRUCTION

Speech Database is an important part of pronunciation dictionary auxiliary module. Its database quality will directly affect the real effect of the speech aid.

5.1 Recording system

Thai is Thailand's mother tongue, also used in the Dai language. The world has a population of around 76 million use of Thai. We use Thai Chinese words and the international phonetic alphabet as the main basis.



Figure 9 Thai recording system interface

5.2 Original sound recording

After collecting data, the next step is to the original Thai sound recording. The purpose of the original sound recording is to get a scientific, accurate, and reliable of the original audio files. For recording personnel selection and the choice of recording environment put forward higher requirements. That is to say, we need to professional personnel for Thai, the original sound recording, quiet studio and playable store tools.

5.3 Recording personnel

The first we should consider the problem of accuracy and reliability. The second is the issue of generality. With the support of Jiamei Wang teacher. We invite international students to the undergraduate is Thai students to the recording. Seven students come from Bangkok and grandparents are Thai nationality. It can guaranteed the authentic of Thai. At present, all input is completed by it. Our recordings with the promotion of pronunciation to reach consensus. In the process of the establishment of the library into the tunnel, pure, thick Thai, speech and language and culture.

5.4 Recording characteristics

Finally through dialogue and qualitative screening, left a complete Thai language audio files and store it for suffix. WAV format. In the Windows platform, WAV is by far the best audio software support audio formats. So far, the group has established the Thai library. This will prepare for Chinese-Thai-English electronic dictionary.

5.5 Achievements exhibition

For convenience of recording and query analysis of labor division, we will refer to Thai vocabulary Thai the tape library astronomy, geography, Time, location, animal, plant, parts of the body, houses, buildings, Commodity, Clothing, fabric; Tools, supplies entertainment; Diet ,Actions, behavior, Nature, state, Characters and title. A total of 29946 words.

6 DATABASE CONSTRUCTION

6.1 Database Principle

In generally, database is the place to store data. Database is a new technology for data management. It is an important branch of computer science, and it is one of the fastest growing disciplines in computer field. Database is technical support for the national language research information. Based on this consideration, our team is constructing large corpus and speech database basing on database information system.

6.2 Database selection

Relational database is one of the most important and popular database. It applied mathematics method to handle the database data, and database is a set which can be stored on the computer, shared and organized. Database can be text, graphics, and sound. This system adopts the Access database and XML to organize and store data information.

6.3 Vocabulary database

This software, respectively, adopts the JDBC and JDOM to store data. There are two data tables. The first table is the vocabulary table, which includes 29976 Thai words. This table can be modified, added and deleted vocabularies. Query of vocabulary (precise search) is retrieved from the table.

The second table is the Thai soundmark table, using XML to describe and store them. It stores 29976 soundmarks. The file is read in the program initialization and encapsulated into a hash table data structure to resident in memory. In the retrieval of data, soundmarks use a thread to read the hash table and on the interface according to the phonetic symbol. Thread can reduce system waiting time, and improve query speed. Such as the soundmarks and Thai text are stored by the XML format, as shown below:

```
<?xml version="1.0" encoding="UTF-8"?>
<ThaiFontTB>
<id>1</id>
<Thai_font> </Thai_font>
<pronunciation>bit</pronunciation>
</ThaiFontTB>
<ThaiFontTB>
<id>2</id>
<Thai_font> </Thai_font>
<pronunciation>bix</pronunciation>
</ThaiFontTB>
<ThaiFontTB>
```

...

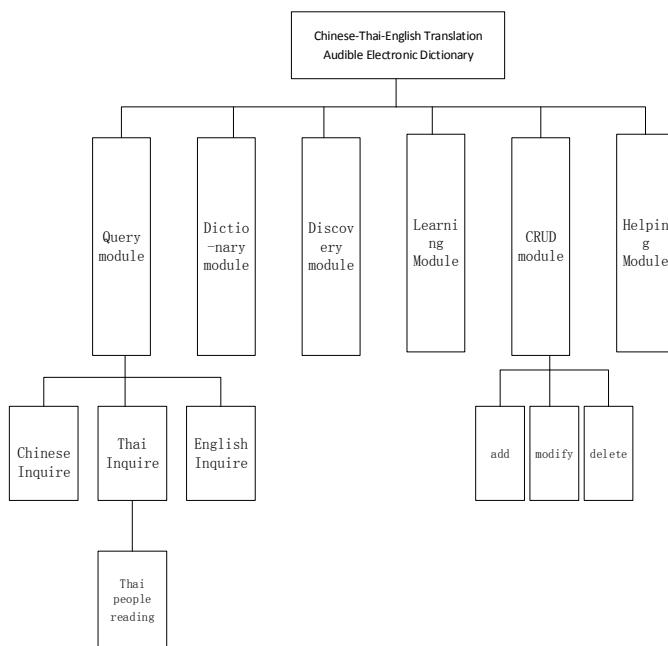


Figure 10 functional modules

7 MODULE DESIGN

The goal of this study is realized electronic dictionary that can achieve Chinese-Thai-English mutual query

function under Windows environment. Dictionary includes vocabulary queries, dictionary, discovery, learning, and instant translation dictionary function. The key of design is words and speech database. The back-end database for an electronic dictionary is necessary. Therefore, the core functionality of the software module is a query module, mainly relying on the Chinese-Thai-English database to complete this goal.

The main function of the electronic dictionary contains vocabulary search, auxiliary query, thesaurus maintenance. Functional modules as shown in Figure 10,

Query Module: query module is an important part of an electronic dictionary. Chinese and Thai which are input will be precisely found by electronic dictionary.

Dictionary module: operating system need to support code, otherwise developers need to make code word stock, and import into the system from outside thesaurus (C:\Windows\Fonts).

Add, delete, modify module: due to input Chinese-Thai-English by manual work, they may have some errors. So it is necessary for an electronic dictionary to design a thesaurus maintenance function.

Found module: it has some funny functions, such as Thailand tourism and some study information about Thai.

8 CONCLUSION

Now, Thai word processing software which is used for translation system has not introduced. Therefore Chinese-Thai-English electronic dictionary that is based on windows operating platform is an innovative work.

Our team design Chinese-Thai-English electronic dictionary which basically achieve lexical search function, vocabulary custom, human pronunciation, international use, accompanying text segmentation function, data backup, and output dictionary database. This software system follows the humanized design. And it faces to reality, close to the user and stabilize performance.

ACKNOWLEDGEMENTS

The regional project of National Natural Science Fund (61363085); The commission project National Language Committee Research Fund (WT125-61); The major project of Yunnan Provincial Department of Education Research Fund (ZD2013013); The construction of high level university for Nationalities project of Yunnan MinZu University Research Fund (ZZZC1501-JF12002) The general project of Yunnan Minzu University Postgraduate Innovation Fund.

REFERENCES

- Cai,zangtai. Hua,quecairang. 2005. Establish and Design the Tibetan Segmentation Dictionary Using for Corpus of Tibetan Processing. China Academic Journal Electronic Publishing House[A].2005:
- Feng,Yuntian. Novel conditional random field model extended by tensor and its application in natural language processing tasks [J].2015,09
- Guliwuya,abudoujili. Design and Implementation of Chinese-Kazakh Diglossia Electronic Dictionary [D].2013:
- Hu,gang.2014.Daile Wen - Chinese Translation Audible Electronic Dictionary Design and Implementation. Computer Systems Applications[J].2014:
- LI,Cuixia. Modern Computer Intelligent Recognition Technology of Natural Language Processing Research Progress and Application. Science Technology and Engineering[J] 2012, 12:
- Nimazhaxi.The Research and Implementation of Chinese-Tibetan Online Translation of Multimedia Electronic Dictionary [D].2008:
- SUO, DongmeiA. Study of Natural Language Understanding. Journal of Chang Chun Teachers College [J]2005.03
- Zha,lima. The Design and Realization of English to Chinese and Mongolian Electronic Dictionary Computer Inquiring Software [D].2006: