World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:7, No:10, 2013

Using Multi-Linguistic Techniques for Thailand Herb and Traditional Medicine Registration Systems

Thanapol Wisuttikul, Choochart Haruechaiyasak, Santipong Thaiprayoon

Open Science Index, Biomedical and Biological Engineering Vol:7, No:10, 2013 publications.waset.org/17085/pdf

***Abstract*—**Thailand has evolved many unique culture and knowledge, and the leading is the Thai traditional medicine (TTM). Recently, a number of researchers have tried to save this indigenous knowledge. However, the system to do so has still been scant. To preserve this ancient knowledge, we therefore invented and integrated multi-linguistic techniques to create the system of the collected all of recipes. This application extracted the medical recipes from antique scriptures then normalized antiquarian words, primitive grammar and antiquated measurement of them to the modern ones. Then, we applied ingredient-duplication-calculation, proportion-similarity-calculation and score-ranking to examine duplicate recipes. We collected the questionnaires from registrants and people to investigate the users' satisfaction. The satisfactory results were found. This application assists not only registrants to validating the copyright violation in TTM registration process but also people to cure their illness that aids both Thai people and all mankind to fight for intractable diseases.

***Keywords*—**Medicine Registration, Search Engine, Text Approximation, Traditional Medicine.

I. Introduction

AVICENNA, Persian physician and philosopher, once said there are no incurable diseases - only lack of will, there are no worthless herbs - only lack of knowledge.

As a result, more people are getting sick every day in the world, and they need to take medicines to treat illnesses. From the past, human made drugs from herbs, animals and minerals, called the traditional or folk medicine. In Thailand, the government has realized that many decadent Thai Traditional Medicine (TTM) scriptures were stolen, lost and gradually disappear from written records. Institute of Thai Traditional Medicine has consequently been established to seek solution to such problems. To date there have been many Thai herbal search engines in Thailand [1]-[3]. However, few studies have used them to store tradition medicine recipes.

In the same situation, there are many studies on Traditional Chinese Medicine (TCM) digital libraries, semantic infrastructure and knowledge discovery systems. The TCM digital library keeps the scanned Chinese medicine books, converts these ones to digital character format with optical character recognition (OCR), and stores them in the database [4]. However, it can't adapt with antique, fragile and blurred traditional Thai scriptures. The researcher from Zhejiang University developed a semantic search based on TCM ontology [5]. KISTCM: knowledge discovery system for TCM uses many algorithms: data mining, neural network and gene expression programming to discovery knowledge from TCM [6]. It is the good concept, but this research's consequence is still doubtful.

As a result, the institute committees decided to use the information technology (IT) to handle these problems. From the cooperation with National Electronics and Computer Technology Center (NECTEC), the committee had an agreement to create the web application to preserving, managing and registration information of Thai herb, traditional doctor and traditional medicine recipes. However, the researchers discovered some duplicated recipes in many scriptures, so they needed to make the algorithm to detect them. This web application will open to Thai traditional doctors to register their own medicine recipes as the patent registration system; consequently, it needs to check if this one does not duplicate with national, generic, and other registered ones.

In this paper, we address the problems regarding traditional recipes of TTM and their solutions through the use of the duplication checking module. We discovered three majority problems -- the meanings similar to but different from those of local herb names, complex structures of ancient Thai language and a variety of herbal weight measurements in traditional medicine recipe. Therefore, we applied many techniques to convert them to the scientific names, common production methods and metric weight units.

II. TTM Duplication Checking Module

The stone inscription of TTM usage was written almost eight hundred years ago [7]. As a result, these recipes were written by ancient Thai language. To check the duplication of recipes, we needed to transform them to modern Thai language. We developed two important modules. First, Thai medicine duplication checking module contains ingredient duplication calculation, proportion similarity calculation and threshold ranking. Second, Thai medicine search module contains full-text, soundex and query approximation search engine. Then we integrated these modules with the medicine registration module to assist officers to detect the copyright

T. Wisuttikul is with Software Engineering Laboratory, Thailand National Electronics and Computer Technology Center, Pathumthani, 12120Thailand (phone: +662-564-6900 ext. 2490; fax: 662-564-6761; e-mail: thanapol.wisuttikul@nectec.or.th).

C. Haruechaiyasak is with Speech and Audio Technology Laboratory, Thailand National Electronics and Computer Technology Center, Pathumthani, 12120 Thailand (phone: +662-564-6900 ext. 2240; e-mail: choochart.haruechaiyasak@nectec.or.th).

S. Thaiprayoon is with Speech and Audio Technology Laboratory, Thailand National Electronics and Computer Technology Center, Pathumthani, 12120 Thailand (phone: +662-564-6900 ext. 2281; e-mail: santipong.thaiprayoon@nectec.or.th).

Open Science Index, Biomedical and Biological Engineering Vol:7, No:10, 2013 publications.waset.org/17085/pdf

infringement recipe and to help people find the right traditional medicine recipe.

We designed Thai medicine duplication checking module in three parts: ingredient duplication calculation, calculation of proportion similarity and score ranking as shown in Fig. 1.

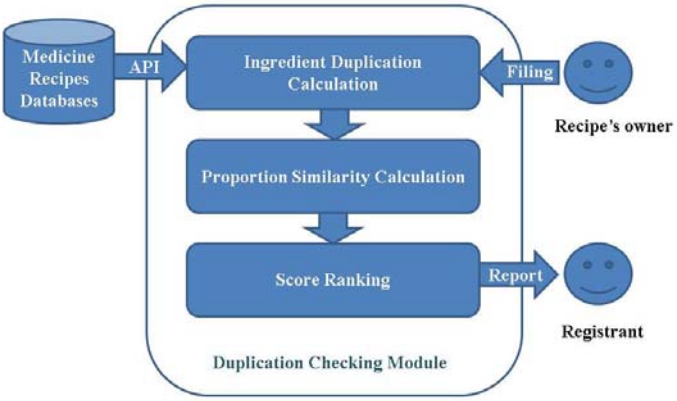


Fig. 1 Architecture of Duplication checking module

A. Ingredient Duplication Calculation

This section reveals the ingredient of medicine recipes and duplication recipe judgment determined from the following:

- Recipe Name
- Ingredients
- Ingredients' proportion or ratio

From our finding, there are three cases of similarity shown in Fig. 2 and Table I.

- Partial match
- Exact match
- Not match

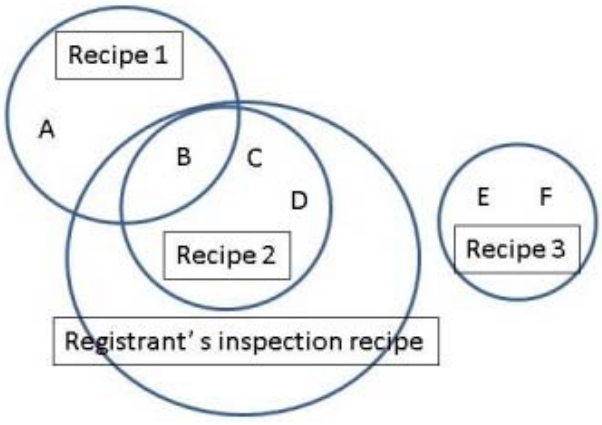


Fig. 2 The Venn diagram of medicine recipes

TABLE I
THE INGREDIENTS' TYPE FROM VENN DIAGRAM

Recipe	Ingredient Match Type
Recipe 1	Partial Match
Recipe 2	Exact Match
Recipe 3	Not Match

However, some Thai herbs have many local names or similar name but different meanings, so we have to normalize local names to scientific names. The examples are shown in Table II, according to official data from the Thai Science Biodiversity Database (http://www.thaiscibiodiversity.org).

TABLE II
EXAMPLE OF THAI HERB' S NAME

Thai Common's Name	English Common's Name	Scientific Name	Thai Local's Name
ดีปลี	Long Pepper	Piper chaba Vahl	- พริก - ประดงข้อ - ปานนุ
ขมิ้นอ้อย	Zedoary, Luya-Luyahan	Curcuma Zedoaria Rose	- ว่านเหลือง - ขมิ้นขึ้น
ฟ้าทะลายโจร	Kariyat, The Creat	Andrographis paniculata (Burm.f.) Wall. ex Nees	- หญ้ากันงู - น้ำลายพังพอน - ฟ้าละลายโจร - ฟ้าสาง - เขยตายยายคลุม - สามสิบดี - เมฆทะลาย - ฟ้าสะท้าน

Moreover, Thais has many unit measures. We needed to convert old traditional local measurement to metric standard units. There are more than one hundred unit measurements in Thailand, so that it takes a lot of time to discover and transform most of them to the standard. The Thailand traditional measurements examples are shown in Table III.

TABLE III
EXAMPLE OF THAILAND TRADITIONAL MEASUREMENTS

Thai traditional measurement unit	Metric measurement unit
1 บาท(baht)	15 กรัม(15 gram)
1 ตำลึง (tamleung)	4 บาท(60 gram)
1 ชั่ง (chang)	20 ตำลึง (1.2 kilogram)
1 หาบ (harp)	50 ชั่ง (60 kilogram)

For now, we normalized recipe's old unit measures to SI ones and various localized herb names to general scientific ones. Next, we sent these values to Proportion Similarity Calculation Module to finding redundant thresholds.

B. Proportion Similarity Calculation

From A., some of the recipes have the same ingredients but different proportions that have difference treatment properties. Therefore, we needed to verify not only recipe's ingredients but also recipe's proportion similarity in the determining process too. We categorized the similarity in two values, redundancy and neighborhood, with three cases.

- Partial match
- Exact match
- Not match

Then, we used these formulas to judge them.

The redundancy is shown in the following formula:

$$A_{(1,2)} = \left(\frac{N_r}{N_u}\right) \times 100\% \quad (1)$$

Open Science Index, Biomedical and Biological Engineering Vol:7, No:10, 2013 publications.waset.org/17085/pdf

where $A_{(1,2)}$ is the redundancy of recipe 1 and recipe 2, N_r is Familiar ingredients of recipe 1 and 2 ingredients and N_u is Sum of all ingredients in recipe 1 and 2.

For example:

We want to verify the recipe 1 with recipe x, where recipe 1's ingredients are A, B and recipe x contains B, C and D. We calculate the redundancy from formula (1).

$$A = \frac{1}{4} \times 100\% = 25\%$$

The redundancy of recipe x with recipe 1 is 25%

The neighborhood is

$$S_{(1,2)} = \frac{C - \sum_{i=1}^{n} (|i_{1,i} - i_{2,i}|)}{C} \times 100\% \quad (2)$$

where $S_{(1,2)}$ is The neighborhood of recipe 1 and recipe 2, C is Sum of max ingredients' proportion,

$i_{(1,i)}$ = Ingredient's proportion of recipe 1 and $i_{(2,i)}$ = Ingredient's proportion of recipe 2.

Note: $i_{(1,i)}$ and $i_{(2,i)}$ are the same ingredient.

If the neighborhood (S) is more than a length constant (C), we give the neighborhood equal a length constant.

For example:

We want to verify the recipe x with recipe 1, where recipe x has an ingredient proportion B 2, C 1 and D 1. Recipe 1 contains A 1 and B 1. We calculate the neighborhood from formula (2).

$$S = \frac{4 - [(|2 - 1|)]}{4} \times 100\% = 75\%$$

The neighborhood of recipe x with recipe 1 is 75%.

In summary, values of the redundancy and the neighborhood between recipe x and recipe 1 are 25% and 75%, respectively. The system sends this value to the Score Ranking module to evaluate them later.

C. Score Ranking

Connected to the Ingredient Duplication Calculation and Proportion Similarity Calculation mentioned earlier is Score Ranking. We determined copyright infringement recipes by redundancy and neighborhood value, so we rearranged them by the threshold ranking method. Therefore, the registrant can set these thresholds to detect suspected recipes. As a result, the module reported all of recipes which redundancy and neighborhood values exceeded thresholds to consider whatever this recipe was duplicated. To avoid legal issues, The Protection and Promotion of Thai Traditional Medicine Wisdom Act B.E. 2542 (1999) [8] has a section to establish the TTM expert committee to prudently review the accused recipe and make decision from the support data discussed above and relative contexts. This system makes all stakeholders judge these issues confidently.

From the examples in section A and B, we have received redundancy and neighborhood values with recipe 1 are 25% and 75%. Because we set redundancy and neighborhood threshold at 20% and 50%, the module listed recipe 1 may be duplicated with this recipe.

III. TTM Search Module

Many functions on TTM duplication checking module need the precise search function. There are many libraries of open-source search engine in the Internet. It is quiet easy to choose one of them to be used with the data in English. On the other hand, the structure of Thai is more complicated than the English language, as the intricate thing is not only grammar, but also spelling too. The researchers then used three methods to search for the precise result, as shown in Fig. 3.

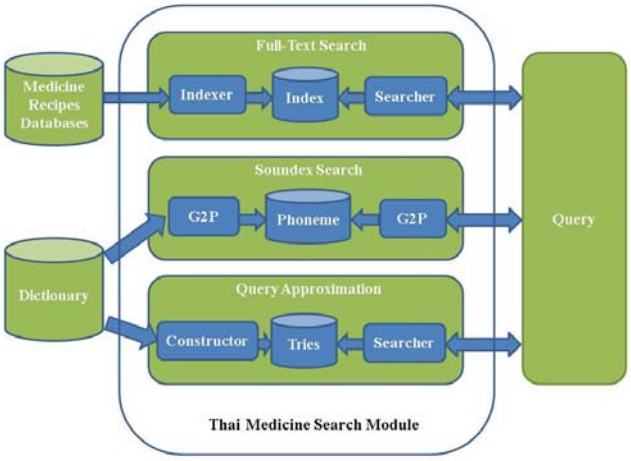


Fig. 3 Thai Medicine Search Module

A. Full-Text Search

We consider the Sansan API [9], the Thai information retrieval system, which consists of two major modules: ApacheLucene, a high-performance text index and search engine library to provide full-text indexing records across Thai medicine recipe and Thai Herbal databases. In addition with ThaiAnalyzer package, its work is tokenizing Thai texts via the dictionary based-on algorithm. Fig. 4 shows the flow of Thai analyzer package. Thai Tokenizer tokenizes the Thai texts, we uses the tries data structure to increasing the speed of dictionary look-up. Then Thai Filter filters all non-Thai texts and passes them to Lucene Analysis to tokenize later.

Open Science Index, Biomedical and Biological Engineering Vol:7, No:10, 2013 publications.waset.org/17085/pdf

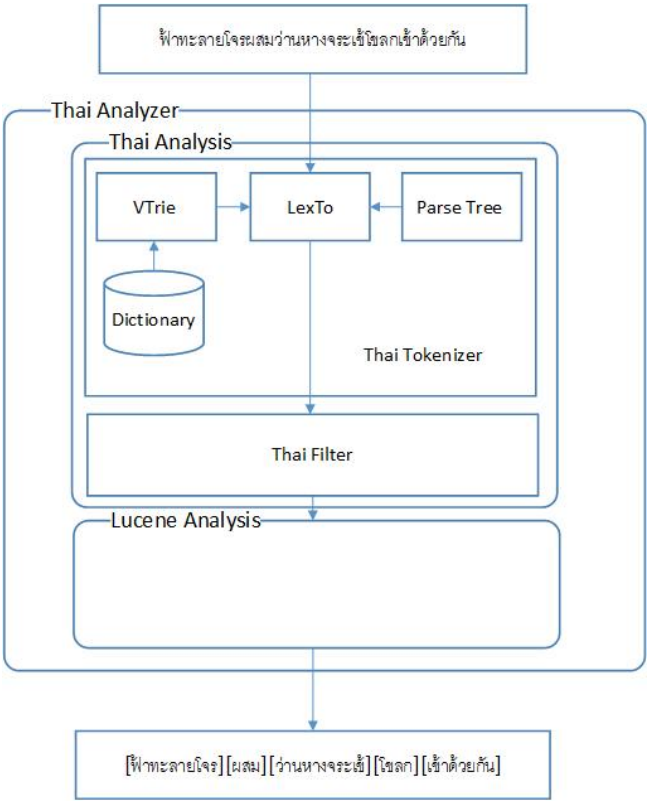


Fig. 4 The Thai Analyzerpackage architecture

Next, the searcher examined a record containing a particular keyword using the created index. A search result, therefore, returns to the user. This application supports the users to search from these criteria:

- The name of Thai herbs, TTM' recipes and Thai Traditional physicians
- Illness Symptoms that Thai herbs and TTM' recipes can be cured.

B. Soundex Search

In this module, the user's word was sent to Grapheme-to-Phoneme (G2P) engine, which was transferred from the graphemes of written language to the phonemes of spoken language, and was handed to Thai Phoneme database to search in Thai dictionary. Therefore, the match possible Thai word returns to a user.

C. Query Suggestion & Approximation

This function is based on the edit-distance calculation between the word input and the dictionary's word. If one matches the other exactly, then it is used directly. Otherwise, the nearest word is chosen as the corrected word instead [10].

IV. System Architectural Design

A concept design showing many components integrated in our system is shown in Fig. 5. It illustrates main functional engines of system that consists of three parts: registration engine, support engine and web interface.

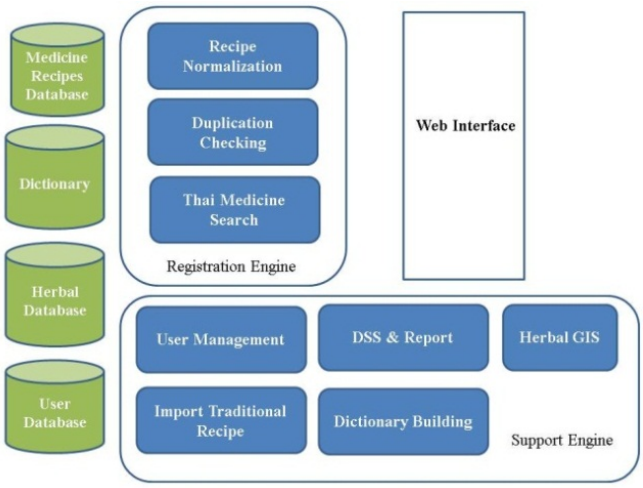


Fig. 5 The system architecture

First, the registration engine consisted of three modules: recipe normalization (transforming local herb names and ancient weight measurements to scientific names and metric weight measurements), duplication checking and Thai medical search. Second, the support engine contained five functions: decision support system and report, dictionary building, import traditional recipe, herbal GIS and user management. Finally, the web interface was used to communicate with users as shown in Fig. 6.

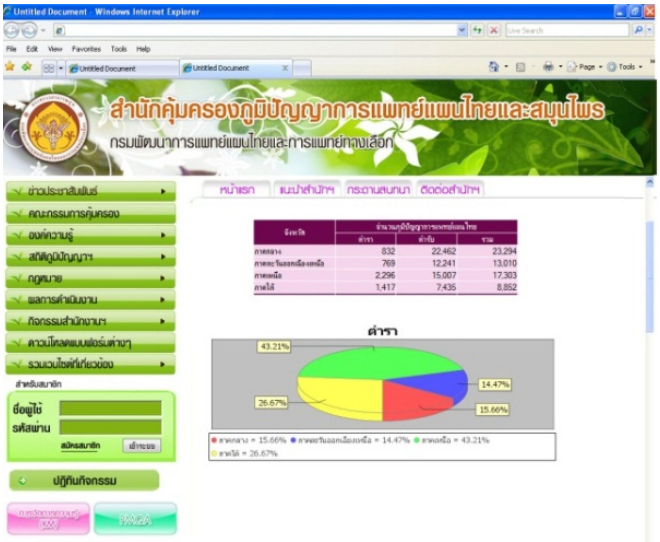


Fig. 6 User interface of this web application

V. Preliminary Results

This web application was approved and used by Institute of Thai Traditional Medicine to register Thai traditional medicines. Now, a number of recipes were submitting to the application and waiting for expert committee to judgments. We measured the user's satisfaction by the Usefulness, Satisfaction, and Ease of use (USE) questionnaire and the Computer System Usability Questionnaire (CSUQ) [11], [12]. On the other hand, these questions are often written in positive

Open Science Index, Biomedical and Biological Engineering Vol:7, No:10, 2013 publications.waset.org/17085/pdf

phrases, which mean that the results from them are biased towards positive responding. Thus, we carefully adapted these questions not to be more positive ones. We collected questionnaires from 150 staffs and 300 users randomly selected from seventy-seven provinces of Thailand. As a result of the survey, this application helped them become more productive, some of whom were satisfied with it and learned to use it quickly. Moreover, this web application helped them to accomplish their work easily, saved more time when they used it and required the fewest steps possible to accomplish any tasks.

VI. Conclusions

The goal of this paper is to adapt and integrate many software technologies that are used by this application to handle the problems with Thai herb names and traditional recipes. It helps Thai Traditional doctors to protect their valuable TTM recipes, assists medical recipe registrants to work for any tasks more easily and helps people to find what they want more easily and conveniently.

We overcomes major problems of Thai-style scriptures: similar to but different from herb name meanings, sophisticated structures of antiquated Thai language and various weight measurements of ingredients in TTM recipe by apply various linguistic algorithms.

In the future, we will make this system even more complete, for this reason, having many jobs to done. First, we plan to focus on constructing TTM's ontology based on history literatures thus it will help users to find what-they-want recipes easier. The next step, we will develop to a large-scale domain ontology TTM discipline hence assisting them to discover relationship between TTM recipes. We, finally, will expand our project to the TTM knowledge discovery and decision support system, consequently, encouraging administrators to make rapid and precise decision [13]. All these things that we've done will be more useful for all of stakeholders.

We hope that the proposed application may be potentially helpful for Thai people to search for useful Thai traditional medicine's recipes, traditional doctors to register their own valuable recipes, and Thai traditional medicine' registrants to determine redundant recipes. With the results, we believe this application can help not only Thai people but also all of the mankind to protect their ancestors' valuable heritages, improve these recipes and educate them to use these recipes to cure themselves and their family.

Acknowledgment

This research presented here was financially supported by Department for Development of Thai Traditional and Alternative Medicine (DTAM), Ministry of Public Health, Thailand. We were really grateful toAssistant Professor Dr. Saneh Thongrin, because all good advices she gave us. At the same time, we very appreciated DTAM central, provincial officers and many people to suggest us valuable feedbacks. We welcomed their comments and improved our system.

References

[1] V. Lertnattee, S. Chomya, A. Apirakaramwong, TH Search: Thai-Enable Herbal Search Engine. In *ISCIT'05. The fifth IEEE International Symposium on Communications and Information Technologies.* IEEE,2005.

[2] V. Lertnattee, S. Chomya, T. Theeramunkong, V. Sornlertlamvanich, Applying Collective Intelligence for Search Improvement on Thai Herbal Information. In *CIT'09, the ninth IEEE International Conference on Computer and Information Technology*. IEEE, 2009.

[3] C. Haruechaiyasak, J. Pailai, W. Viratyosin, R. Kongkachandra, ThaiHerbMiner: A Thai Herbal Medicine Mining and Visualizing Tool. In *ACL-HLT'11, the 2011 Workshop on Biomedical Natural Language Processing.* ACL,2011.

[4] J.Yuan, B. Wei, L. Wang, W. Lu, Y. Zhuang, CMSSOF: a structured data organization framework for scanned Chinese medicine books in digital libraries. Journal of Zhejiang University-SCIENCE C. p.882-892. 2010.

[5] H. Chen, National Semantic infrastructure for traditional Chinese medicine. Journal of Zhejiang University-SCIENCE C. p.311-314. 2012.

[6] S. Qiao, C. Tang, H. Jin, J. Peng, D. Davis, N. Han, KISTCM: knowledge discovery system for traditional Chinese medicine. Applied Intelligence, Volume 32, Number 3, June 2010, pp. 346-363(18).

[7] V. Chokevivat, A. Chuthaputti, The Role of Thai Traditional Medicine in Health Promotion. In *GCHP '05, the 6th Global Conference on Health Promotion (GCHP)*. WHO,2005.

[8] The Protection and Promotion of Thai Traditional Medicine Wisdom Act B.E. 2542 (1999).

[9] C. Haruechaiyasak, C. Damrongrat, C. Sangkeettrakarn, S. Kongyoung, N.Angkawattanawit, Sansarn Look!: A Platform for Developing Thai-Languageinformation Retrieval Systems. In ITC-CSCC2006, the 21st International Technical Conference on Circuits/Systems, Computers and Communications. p. 85–88. 2006.

[10] N. Angkawattanawit, C. Haruechaiyasak, S. Marukatat, Thai Q-Cor: Integrating Word Approximation and Soundex for Thai Query Correction. In *ECTI-CON'08, the fifth annual international conference organized by Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Association*. ECTI, 2008.

[11] J. Lewis, IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. In *International Journal of Human-Computer Interaction v.7 n.1 p.57-78.* IBM, 1995.

[12] A. Lund, Measuring Usability with the USE Questionnaire. In *STC Usability SIG Newsletter October 2001 issue Vol 8, No. 2*. STC,2001.

[13] X. Zhou, S. Chen, B. Liu, R. Zhang, Y. Wang, P. Li, Y. Guo, H. Zhang, Z. Gao, X. Yan, Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artificial Intelligence in Medicine, Volume 48, Issues 2–3, February–March 2010, pp. 139–152.