ThaiHerbMiner: A Thai Herbal Medicine Mining and Visualizing Tool

Choochart Haruechaiyasak† **Jaruwat Pailai**‡ **Wasna Viratyosin**∗ **Rachada Kongkachandra**‡

†Human Language Technology Laboratory (HLT),
National Electronics and Computer Technology Center (NECTEC), Thailand 12120

‡Department of Computer Science, Faculty of Science and Technology
Thammasat University, Thailand 12121

∗BIOTEC Central Research Unit,
National Center for Genetic Engineering and Biotechnology, Thailand 12120

Abstract

Thai Traditional Medicine (TTM) has a long history in Thailand and is nowadays considered an effective alternative approach to the modern medicine. One of the main knowledge in Thai traditional medicine is the use of various types of herbs to form medicines. Our main goal is to bridge the gap between the traditional knowledge and the modern biomedical knowledge. Using text mining and visualization techniques, some implicit relations from one source could be used to verify and enhance the knowledge discovery in another source. In this paper, we present our ongoing work, *ThaiHerbMiner*, a Thai herbal medicine mining and visualizing tool. ThaiHerbMiner applies text mining to extract some salient relations from a collection of PubMed articles related to Thai herbs. The extracted relations can be browsed and viewed using information visualization. Our proposed tool can also recommend a list of herbs which have similar medical properties.

1 Introduction

In 1993, the Royal Thai Government instituted the National Institute of Thai Traditional Medicine, under the supervision of the Ministry of Public Health. The goal of the institute is to systematize and standardize the body of Thai Traditional Medicine (TTM) knowledge. The main task is to gather, revise, verify, classify, and explain the TTM knowledge. There are many ongoing project collaboration to digitize the TTM knowledge, many of which are documented on palm leaves. The digitized contents contain information on Thai medical herbal formulations with the healing properties. A medical herbal formulation could contain more than one herb and combined with others for better effect.

Apart from the traditional knowledge, today biomedical research has advanced into the genetic level. Many researchers have performed in-depth studies of herbs' medical properties on disease treatment. The main goal of our research is to combine the knowledge from traditional and modern biomedical research. Using knowledge from one source could support the knowledge discovery in another source. To assist the researchers in Thai herbal medicine, we propose *ThaiHerbMiner*, a text mining and visualizing platform. ThaiHerbMiner's main task is to extract and visualize relations among herbs, properties and other entities. Our work is similar to the current ongoing research in mining Traditional Chinese Medicine (TCM) which has gained increasing attention in recent years (He et al., 2011; Lukman et al., 2007).

2 Design and implementation

Text mining has become a widely applied technique for analyzing biomedical texts (Cohen and Hersh, 2005). The proposed *ThaiHerbMiner* is designed with the standard text mining process. We started by collecting PubMed articles by using herb names as keywords. Currently, we have obtained approximately 18,000 articles related to Thai herbs such as garlic, curcuma and ginger.

Figure 1 shows the text mining process of extracting relations from given input texts. The process includes sentence segmentation, tokenization, POS

Proceedings of the 2011 Workshop on Biomedical Natural Language Processing, ACL-HLT 2011, pages 186–187,
Portland, Oregon, USA, June 23-24, 2011.

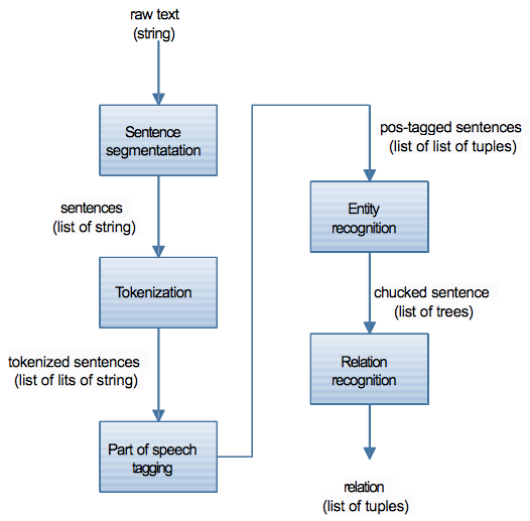

Table 1: The text mining process for extracting relations from input texts.

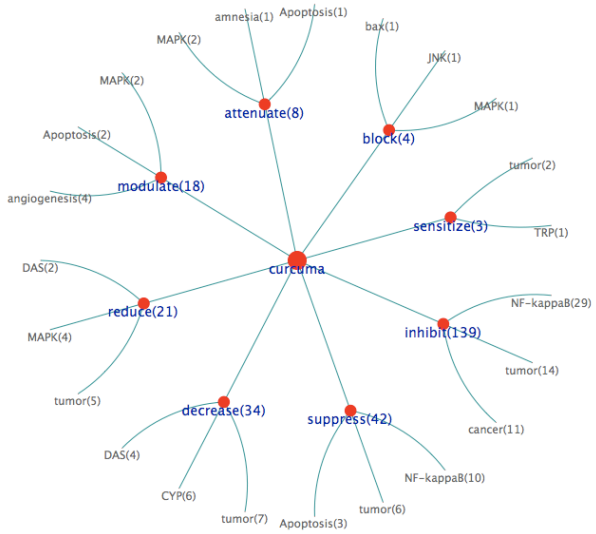

Table 2: An example of relations between curcuma and other relevant entities.

tagging and entity & relation recognition. We used *OpenNLP*[1] to perform all text processing tasks. For relation recognition based on syntactic structure, we focus on a group of causal verbs such as *activate*, *induce*, *inhibit*, *prevent*, *regulate* and *suppress*. Then the information visualization based on JavaScript [2] is applied to represent the extracted relations.

Figure 2 shows an example of a hyperbolic tree visualizing relations between curcuma and other entities. For example, *curcuma* has the property of *inhibit* with *NF-kappaB*, *tumor* and *cancer*. Figure 3 shows an example of a force-directed graph visualizing similar herbs sharing two entities, *cancer* and *NF-kappaB*. The visualizing result is useful to researchers for finding herbs which share similar medical properties.

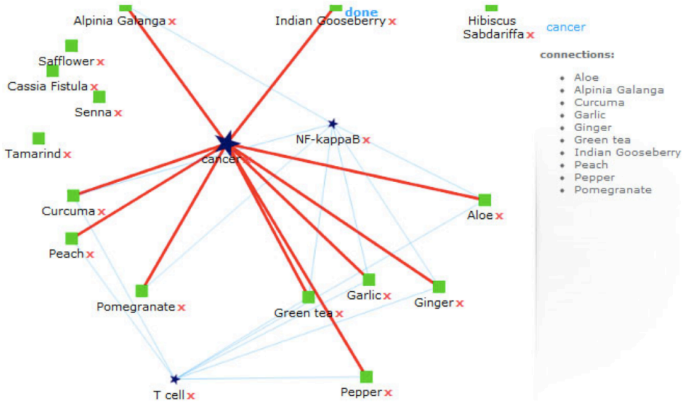

Table 3: An example of relations among different herbs sharing the same entities.

3 Conclusion and future work

The results of literature mining can be potentially useful in revealing implicit relations underlying the knowledge in herbal medicine. In particular, the results can be used in screening the research in Thai herbal medicine to form a novel hypothesis. Our next step is to perform comparative analysis on the knowledge from Thai traditional medicine and the knowledge extracted from the modern research publications.

[1]The OpenNLP Homepage, *http://opennlp.sourceforge.net*
[2]The JavaScript InfoVis Toolkit, *http://thejit.org*

References

Cohen, Aaron M. and William R. Hersh. 2005. A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6(1):57-71.

He, Ping, Ke Deng, Zhihai Liu, Delin Liu, Jun S Liu and Zhi Geng. 2011. Discovering herbal functional groups of traditional Chinese medicine. *Statistics in medicine*, March 17, 2011.

Lukman, Suryani, Yulan He, and Siu-Cheung Hui. 2007. Computational methods for Traditional Chinese Medicine: A survey. *Comput. Methods Prog. Biomed.*, 88(3): 283–294.