

Reliable Improvement for Collective Intelligence on Thai Herbal Information

Verayuth Lertnattee¹, Sinthop Chomya¹, and Virach Sornlertlamvanich²

¹ Faculty of Pharmacy, Silpakorn University
Nakorn Pathom, 73000, Thailand
verayuths@hotmail.com, verayuth@email.pharm.su.ac.th,
sinthop@email.pharm.su.ac.th

² Thai Computational Linguistics Laboratory
NICT Asia Research Center, Pathumthani, 12000, Thailand
virach@tccllab.org

Abstract. Creating a system for collecting herbal information on the Internet, is not a trivial task. With the conventional techniques, it is hard to find the way which the experts can build a self-sustainable community for exchanging their knowledge. In this work, the Knowledge Unifying Initiator for Herbal Information (KUIHerb) is used as a platform for building a web community for collecting the intercultural herbal knowledge with the concept of a collective intelligence. With this system, images of herbs, herbal vocabulary and medicinal usages can be collected from this system. Due to the diversities of herbs, geographic distribution and their applications, one problem is the reliability of herbal information which is collected from this system. In this paper, three mechanisms are utilized for improving reliability of the system: (1) information for an herb is divided into several topics. Contributors could select some topics which they are expertise, (2) a voting system is applied and the standard source members (SSMs) are able to contribute their knowledge on text information, (3) a voting system, keywords and comments are implemented for controlling quality and reliable of images of an herb. With these mechanisms, herbal information on KUIHerb is more accurate and reliable.

1 Introduction

Origins of many traditional treatments in Thailand can be traced to India. The derivation has been diversified through out many cultures since then [1]. For example, herb names and their medicinal usages are gradually spread out into communities resulting in distinction from each other according to their cultural background. Some are named different and hardly found the relation between each other. Some are complimentary knowledge of their usages. These herb names and terminology on herbal medicine are useful for searching herbal information on the Internet. With the conventional techniques such as interview with traditional doctors, it is hard to find the way which the experts

can build a self-sustainable community for exchanging their information. The Internet is an excellent source for providing and sharing information. With Web 2.0 system, it provides an opportunity for sharing information from a group of members on a topic of interest. The Knowledge Unifying Initiator for Herbal Information (KUIHerb), a system for collective intelligence on herbal medicine based on Web 2.0 system, is used as a platform for building a web community for collecting the intercultural knowledge. However, mechanisms for controlling accurate and reliable herbal information contributed by members, are questionable. In this paper, three mechanisms are utilized for improving reliability of the system: (1) information for an herb is divided into several topics. Contributors could select some topics which they are expertise, (2) a voting system is applied and the standard source members (SSMs) are able to contribute their knowledge on text information, (3) a voting system, keywords and comments are implemented for controlling quality and reliable of images of an herb.

In the rest of this paper, the concept of collective intelligence with Web 2.0 and the future Web system is described in Section 2. Section 3 gives a detail of herbal information. Section 4 presents four components of a system for collecting herbal information, namely KUIHerb, with the aspects that we describes. Section 5 explains mechanisms for reliable improvement in KUIHerb. The experimental results are described in Section 6. A conclusion and future works is made in Section 7.

2 Collective Intelligence with Web 2.0 and the Future Web

In Web 2.0 era, the Internet users easily share their opinions and resources. Consequently, users can collectively contribute to the Web community and generate massive content behind their virtual collaboration [2]. For a system with collective intelligence, implementing scalability can indeed be challenging, but sensibility comes at variable sophistication levels. Several approaches are dealing with the sensibility e.g., user feedback, recommender systems, search engine, and mashups. As suggested by Gruber T., the true collective intelligence can be considered if the data collected from all those participants is aggregated and recombined to create new knowledge and new ways of learning that individual humans cannot do by themselves [3]. However, it provided only a little bit on control of information in Web 2.0.

Nowadays, we are going to the new generation of Web technology i.e., Web 3.0 or the future Web. Although it has already received quite a number of definitions, some useful features of Web 3.0 are described as follow. It can be considered as “The data Web” instead of “The document Web” in Web 2.0. The control of sharing information is better. The decision for the opinions which are provided in Web 3.0, is more accurate. The intelligence Web is a new

important feature in Web 3.0 while in Web 2.0, it is only the social Web [4]. Unlike Web 2.0 which participants are usually general Internet users, wisdom of the expert is essential for constructing more knowledge that is valuable. From these features of Web 3.0, it should be a better collective intelligence system for building new knowledge by way of Information Technology (IT), especially medical knowledge, and herbal knowledge should be no exception.

3 Herbal Information

Herbal information is a special type of information dealing with medicinal herbs. Some topics such as name identification and medicinal uses which may be different among cultures, are still problems. For instance, the same species of an herb may be known by different names in different areas. On the other hand, a certain herbal name may mean one thing in one area but something completely different in another. The relationship between herbs and their names is Many-Many i.e., a plant may have several names while a name may be several plants. For example, *Dracaena loureiri* Gagnep. We use its hard wood for fever and call Chan dang. Some time we call this plant in other names up to the area of country, e.g., Chan pha (northern part), and Lakka chan (central part) [5]. Lack of information about native herbs has made them more difficult for applying. Herbal specialists usually seek herbal information in a standard monograph. The herbal monograph deals with information to determine the proper identity of a plant genus or genus and species, including part used, indication, method for preparation and so on. However, these sources of information are limited. In the case of the herb does not appear in the pharmacopoeia, it is hard to seek accurately information about the herb. This causes general users and herbal specialists find information of herbs and their products on the Internet. A set of images of an herb is excellent sources for sharing knowledge about herb identity. From the images, the users can discuss which species (including variety) it should be. The scientific name of an herb and its images are used for common understanding. Furthermore, the users can discuss about which herb should be the real herb that appears in herbal formulas [6].

4 KUIHerb: Collective Intelligence System for Herbal Information

KUIHerb has been implemented using all open source software components. The url is <http://inf.pharm.su.ac.th/~kuiherb> in Thai version. The scripting language is PHP. The data are collected in a database which is constructed with MySQL. With the concept of Web 2.0 and some features of the future Web, the system has been designed for general users and members who would like to participate. The four components in KUIHerb are described as follow.

4.1 Accessing Information

Information of an herb can be reached by two methods i.e., keyword search and directory search. KUIHerb provides the ability to keyword search by using a Thai common name, a Thai local name, an English name, a scientific name of an herb as well as a family name. Is also provides the ability to browse categories of part used and symptom. In the Figure 1, it shows the Web page for searching information by a directory and keyword.

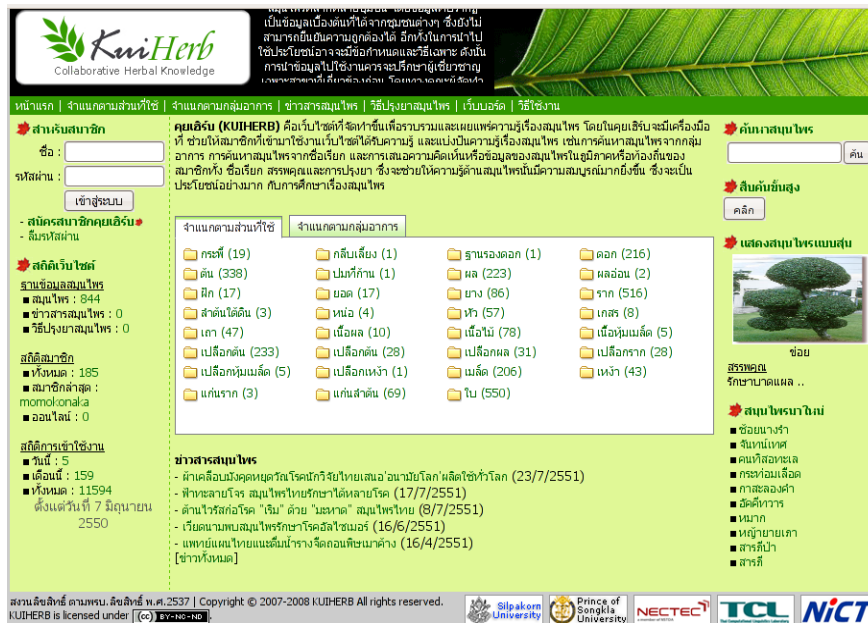


Fig. 1. The Home Page of the KUIHerb

The scientific name of an herb and its images are used for common understanding. In this platform, not only the text content can be shared among members but also images of an herb which can be uploaded to the system. This is very important for herbs whose part used rarely appear.

4.2 Sharing Information

For the first version of KUIHerb, six topics are taken into account i.e., general characteristics, pictures, local name, medicinal usages (i.e., part used with their indications and methods for preparation), toxicity, and additional information. Among these topics, a poll-based system is implemented on local names and medicinal usages.

4.3 Providing Information

Two approaches are constructed for providing herbal information. The first approach is the current news about herbal information by Web links. The administrator of the KUIHerb usually added news about herbs and it is easy to link to the source of information. The other approach, information of an herb is randomly selected from KUIHerb database when users visit the homepage of the Web site. It also provides a list of new herbs added to the database.

4.4 Web Site Statistics

In KUIHerb, hit counters roughly indicate Web sites' relative popularity and users' activity. Three set of counters are created for these proposes. The first set is for herbal database activity. The volume of user information contributes to KUIHerb can be used as an indicator for the level of user participation. For this set, three counters are used i.e., the number of herbs, news and topics in Web board. The second set is for describing the members of the community i.e., the number of member, the newest member, the number of active members of that time. This set indicates the popularity of the system. The last statistical set reports the total activities in a period of a day, a month and a year.

5 Reliable Improvement in KUIHerb

Several mechanisms are used for improving the reliability of the system. The detail of these mechanisms is described as follow.

5.1 Subdividing All Information into a Set of Topics

In fact, information of an herb is huge. It is hard for a person to recognize everything. For example, some photographers would like to take pictures of plants. They may not know the local names and usefulness of an herb. However, they can contribute the pictures of whole plants, flower, fruit, etc. A native people of the area can contribute the local names of an herb. Pharmacists may suggest its indication and toxicity but they may not know the local names of an herb. With this concept, information of an herb should be subdividing into several topics. Members should contribute information on topics which they have some experience, Moreover, some reliable or standard references will be added for further finding information.

5.2 Applying Voting System

The voting mechanism is widely used to improve accuracy of the system such as in [7]. For more accurate information, only members (and the administrator) in the system are able to contribute and modify their information.

With this system, a contributor may choose to work individually by posting his/her opinions into the topics. Any opinions or suggestions are committed to voting. While opinions may be different, majority votes determine the view of the communities. These features naturally realize the online collaborative works to create the knowledge communities. The weighting system for each opinion can be calculated by the formula

$$Wsum_{ik} = \sum_j w_{ijk}$$

Here, $Wsum_{ik}$ is defined as the total weight of the i^{th} opinion of the k^{th} topic. The w_{ijk} is the weight of the i^{th} opinion which is given by the j^{th} member who would like to vote in this opinion for the k^{th} topic. The value of w depends on the priority and agreeeness of the member. The weight from the member, who contributes more accurate information for a long period, should be higher than the new one. The w is needed to update from a period of time. Furthermore, if the member agrees with the opinion, the value is positive and vice versa. A set of higher weight opinions for each topic, tends to be more believable.

To increase more accuracy, a concept of standard source members (SSM) is proposed. For herbal information, standard sources of information may standard textbook about herbal medicine which written by experts in this area. Each standard source can be represented a member in the system. The list of these standard references is included on the topic of references. The administrator creates SSMs and input opinions from the SSMs to system (only text information). This process is transparent to the real members. With this method, not only information with voting system is more accurate but also gives information which is used for calculating w for each member (a real member and a SSM) in the future. The voting system can be applied in both text and image information.

5.3 Reliable for Images of an Herb

Due to synonyms and homonyms of the local names of an herb, the scientific name of an herb and its images are used for common understanding. In this system, images of an herb can be uploaded to the system. The images described should relate to the whole plant. The parts which have medicinal usage such as leaves, roots, flowers, seeds, resin, root bark, inner bark (cambium), berries and sometimes the pericarp or other portions, should be included. This is very useful to visitors who would like to see parts used of an herb. They should be in both fresh and dry forms. To make these images more reliable, three mechanisms are provided as follow.

- The keywords and contributors' names can be given to the system: keywords suggest visitors about the focus point on the image. Contributors' names guarantee visitors for quality of their images.

- The voting system: this mechanism may summarize the popularity and quality of the images. The basic idea is that images which high quality and/or useful for treatments should be more popular.
- A comment : in case of an image has some problems; e.g., incorrect picture, image is not clear. This comment can be used as a tool to inform visitors. The administrator may have a decision to the image.

6 Experimental Results

When the KUIHerb has been collected information from members for one year. Several mechanisms has been proposed for improving reliability of the systems. In order to represent the concept for improving reliability of the system, six herbs are randomly selected as a set of sample. The results are described as follow.

6.1 Main Topics

With careful design, information of an herb can be divided into seven topics; i.e., general characteristics, pictures, local name, medicinal usages (i.e., parts used with their indications and methods for preparation), toxicity, additional information and references. The detail description for each topic is shown in Table 1. Notice that some information; i.e., locations and part used, can be selected from a list. This controls typo errors from contributors.

6.2 Voting System

In this version, a majority voting score is applied on the topics of local names and medicinal usages. For local names, we can summarize that the herb name is used in the levels of a city, province or larger area. Due to the first period of use, all members are given equal weight. A member can suggest a new opinion which the score of one is initiated. When other members agree with the opinion, a simple click on the button “Vote” will increase the score by one. Each member has only one vote for an opinion. The opinion with the higher score will be moved to the upper part of the window. In cases of multiple opinions, the popular vote will select more preferred opinions used in the community. The Figure 2 represents a list of local names for an herb with their scores. Moreover, a hierarchy of locations is provided in the list. It can be selected by members. The hierarchy of locations controls accuracy of inputs from users and uses for summarization. If a local name is called in several provinces of a region, we can conclude that the local name is belonged to the region. For medicinal usages, information about part used, indication, method for preparation can be contributed.

In order to represent the effectiveness of the SSM, 27 standard Thai textbooks for herbal medicine are added into the KUIHerb as 27 SSMs. We

Table 1. Seven topics for herbal information

Topic	Explanation	Contributor	Type
General Characteristics	General characteristics of an herb such as description about leaves, flowers, fruit, etc., location, culture are added into the system	Admin (to give basic information, not for sharing)	Text
Images	Images of a whole plant and parts which have some medicinal usages	Member (general member, photographer)	Images, Text
Local Names	Local names of herbs are suggested by members. Multi-lingual names can also be applied. Lists of location are provided in the level of city, province, region, etc.	Member (local area people, herbal specialist)	Text, List
Medicinal Usages	A list of predefined parts which may be used for treatments is provided. A member may select the part and suggest its indications. The method for preparation can be suggested. In the case of a part with several indications and several methods for preparation, the opinion should separate the indications and methods of preparation for each part used.	Member (traditional doctor, herbal specialist, pharmacist)	Text, List
Precaution and Toxicity	Any suggestions about precaution and toxicity will be kept for warning when someone would like to use the herb	Member (traditional doctor, herbal specialist, pharmacist)	Text
Additional Information	Other valuable information such as cultivation may also be given.	Member (general member)	Text
References	This area can be applied for suggesting references for an opinion in order to make the opinion more reliable.	Admin	Text

randomly select six herbs. Each herb has at least one image. The numbers of local names as well as medicinal usages for each herb before SSM (no SSM) and after 27 SSMs (with SSM), are shown in Table 2. The maximum scores for both topics from each herb are presented in parenthesis.

Table 2. Comparing voting results from two patterns of membership

Scientific Name	English Name	Local Name		Medicinal Usage	
		no SSM	with SSM	no SSM	with SSM
<i>Alpinia galanga</i> (L.) Willd.	Galangal	11(1)	13(5)	10(2)	12(5)
<i>Senna alata</i> (L.) Roxb.	Ringworm bush	17(7)	22(10)	7(7)	12(9)
<i>Millingtonia hortensis</i> L.f.	Cork tree	5(7)	12(8)	4(7)	5(11)
<i>Oroxylum indicum</i> (L.) Kurz	Indian trumpet flower	14(2)	24(9)	8(2)	19(5)
<i>Piper betle</i> L.	Betel pepper	10(5)	11(6)	7(6)	9(8)
<i>Solanum indicum</i> L.	Sparrow 's Brinjal	15(4)	25(7)	3(5)	8(10)



Fig. 2. Majority Voting System for the Thai Local Name

From the result, some observations are made; 1) the more different local names and medicinal usages are obtained with SSM, 2) the maximum scores are increased on both topics for the system with SSM, 3) for some herbs, the large gap of the differences can be found. We can conclude that the SSM is useful for the system. The opinions on each topic are more diverse, more reliable and more interesting.

6.3 Reliable for Images of an Herb

In this system, a voting system, keywords and a comment are applied on contribution of images. Keywords provide clear text information to images about an herb name and the name of part used. This is useful for users who do not familiar with herbal information. With voting system, an image that is frequently clicked by visitors will obtain the higher score and this image will be shifted to the upper part of the window. For example, images of *Alpinia galanga* (L.) Willd., is shown in Figure 3. Five images of the herb has been collected i.e., stem, flower1, fruit, rhizome and flower2. At this time, the numbers of hits for each image are 5, 3, 3, 3 and 1, respectively. Keywords below an image help us for identifying it. The Table 3, represents parts of an herb given by members on the image topic and medicinal usages topic. The first part used for each cell (in bold) is the part which gained the highest score. Moreover, the comment is valuable for controlling quality of an image.

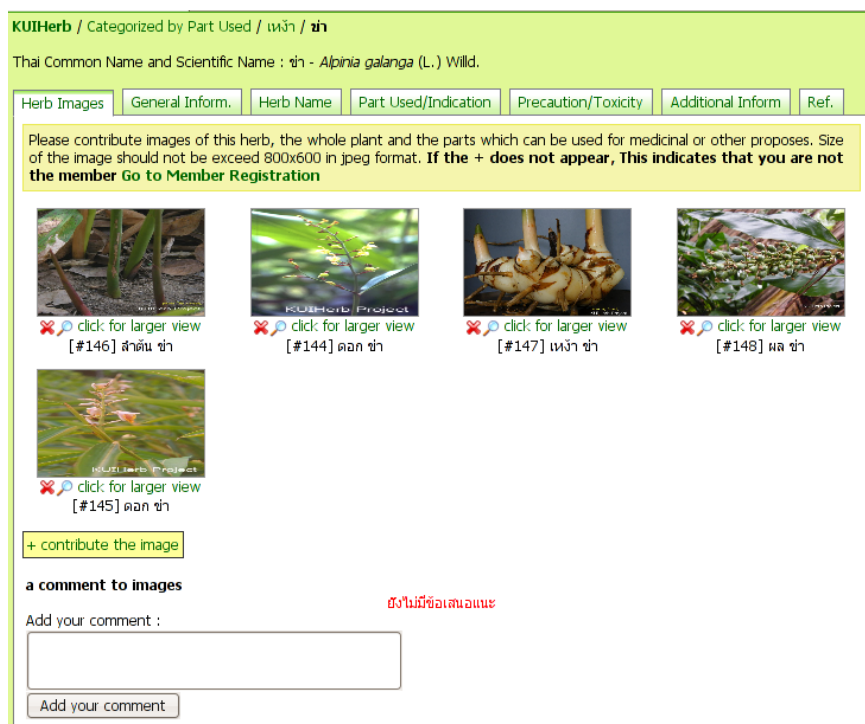


Fig. 3. Images of *Alpina galanga* (L.) Willd. are ranked according to popular view

When members find that an image is incorrect and/or low quality, they can write a message to the system for warning other users.

From the result, almost all parts used on image topic can be found on medicinal usages topic. The most popular part used is quite similar. In addition, information on medicinal topic also suggests that which images of an

Table 3. Information about part used on image topic and medicinal usage topic

Scientific Name	Image	Part Used for Medicinal Usages
<i>Alpina galanga</i> (L.) Willd.	stem, flowers, fruit, rhizome	rhizome, leaves, fruit, stem, flowers, root
<i>Senna alata</i> (L.) Roxb.	flowers, stem	flowers, seed, leaves, pod, stem, root
<i>Millingtonia hortensis</i> L.f.	flowers, leaves	flowers, root, heartwood
<i>Oroxylum indicum</i> (L.) Kurz	flowers, pod, leaves, seed	bark, root, pod, seed, root bark, leaves, stem
<i>Piper betle</i> L.	leaves	leaves, root
<i>Solanum indicum</i> L.	fruit, flower	fruit, root, leaves, stem

herb should be included into the system. We can summarize that keywords and voting system is valuable for increasing reliability of images topic.

7 Conclusion and Future Works

In this work, the KUIHerb was used as a platform for building a web community for collecting the intercultural herbal knowledge based on the concepts of Web 2.0 and some features of the future Web. Due to the diversities of herbs, geographic distribution and their applications, one problem was the reliability of herbal information which was collected from this system. Three mechanisms were utilized for improving reliable on herbal information. First, information for an herb was divided into several topics. Contributors could select some topics which they are expertise. Second, a voting system was applied and the standard source members (SSMs) could contribute their knowledge. Finally, a voting system, keywords and comments were implemented for controlling quality and reliable of images. With these mechanisms, herbal information on KUIHerb was more accurate and reliable. It could be applied for medical and pharmaceutical usages with confidence.

For the first version of KUIHerb, majority voting with equal weight from the members were used for selecting a set of accepted opinions. However, the member who has made contribution that is more valuable to the system should be given more weight. Furthermore, applying data mining to the collected data will be useful. These issues are left for our future works.

Acknowledgments

This work has been supported by Thailand Research Fund and Commission on Higher Education (CHE) under project number MRG5080125 as well as the National Electronics and Computer Technology Center (NECTEC) via research grant NT-B-22-MA-17-50-14.

References

1. Lovell-Smith, H.D.: In defence of ayurvedic medicine. *The New Zealand Medical Journal* 119, 1–3 (2006)
2. Lin, K.J.: Building web 2.0. *IEEE Computer* 40, 101–102 (2007)
3. Gruber, T.: Collective knowledge systems: Where the solcial web meets the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web* 6, 4–13 (2007)
4. Glustini, D.: Web 3.0 and medicine. *British Medical Journal* 335, 1273–1274 (2007)
5. Smitinand, T.: Thai Plant Names. In: *The forest Herbarium Royal Forest Department, Thailand* (2001)

6. Wu, K., Farrelly, J., Upton, R., Chen, J.: Complexities of the herbal nomenclature system in traditional chinese medicine (tcm): lessons learned from the misuse of aristolochia-related species and the importance of the pharmaceutical name during botanical drug product development. *Phytomedicine* 14, 273–279 (2007)
7. Ko, Y., Park, J., Seo, J.: Using the feature projection technique based on a normalized voting method for text classification. *Information Processing and Management* 40, 191–208 (2004)