DATABASE FOR CRUDE DRUGS AND KAMPO MEDICINE

MASANORI ARITA\textsuperscript{1,2,*}  MIWA YOSHIMOTO\textsuperscript{3}  KAZUHIRO SUWA\textsuperscript{4}
arita@bi.s.u-tokyo.ac.jp  yoshimotom@jaist.ac.jp  suwa@liorect.jp

AKI HIRAI\textsuperscript{5}  SHIGEHIKO KANAYA\textsuperscript{2,5}  NAOTOSHI SHIBAHARA\textsuperscript{6}  KEN TANAKA\textsuperscript{7}
{hirai,skanaya}@gtc.naist.jp  {shiba1,ktanaka}@inm.u-toyama.ac.jp

\textsuperscript{1} Department of Biophysics and Biochemistry, Graduate School of Science
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

\textsuperscript{2} Metabolome Informatics Unit, RIKEN Plant Science Center
1-7-2 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan

\textsuperscript{3} School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Iwate, 923-1292, Japan

\textsuperscript{4} The Liorect, Co., Ltd.  5-10-5 Shimochiai, Chuoku, Saitama 338-0002, Japan

\textsuperscript{5} Department of Bioinformatics and Genomics, Graduate School of Information Science
Nara Institute of Science and Technology, 8916-5 Takayama, Ibaraki, Nara 630-0192, Japan

\textsuperscript{6} Research Center for Ethnomedicine, Institute of Natural Medicine, University of Toyama
2630 Sugitani, Toyama 930-0194, Japan

\textsuperscript{7} Department of Medicinal Resources, Institute of Natural Medicine, University of Toyama
2630 Sugitani, Toyama 930-0194, Japan

A wiki-based repository for crude drugs and Kampo medicine is introduced. It provides taxonomic and chemical information for 158 crude drugs and 348 prescriptions of the traditional Kampo medicine in Japan, which is a variation of ancient Chinese medicine. The system is built on MediaWiki with extensions for inline page search and for sending user-input elements to the server. These functions together realize implementation of word checks and data integration at the user-level. In this scheme, any user can participate in creating an integrated database with controlled vocabularies on the wiki system. Our implementation and data are accessible at http://metabolomics.jp/wiki/.

Keywords: Kampo, Chinese medicine, database, wiki

1. Introduction

1.1. What is Kampo medicine?

Kampo medicine (KM) is different from the traditional Chinese medicine (TCM). Japanese Pharmacopoeia (JP or “Nihon Yakkyoku Ho“) [1] contains around 150 crude drugs, and they are prescribed in combination with Western drugs. The combined

\textsuperscript{*} Part of this work first appeared as “Ontology checking and integration of crude-drug and kampo information” in the Proceedings of the 2nd International Conference of Biomedical Engineering and Informatics (BMEI2009, Tianjin China, 3:1304-7) and as “Database for Crude Drugs and Kampo Medicines” in the Proceedings of the 2010 Annual Conference of the Japanese Society of Bioinformatics (JSBi2010, Fukuoka Japan, T03).
prescription is unparalleled in any traditional medicine, including traditional Korean medicine and TCM. Japan is a unique country where crude drugs have been administered in Western standards with records for around 40 years. In this sense, KM is neither complementary nor alternative medicine. This distinction from Chinese or other traditional medicines is sometimes emphasized in Japan. (In China, on the other hand, all Asian medicines are integrated into TCM.) Since traditional medicines, contrary to patented Western medicine, cannot escape from the debate of intellectual property rights, accurate understanding of their origins and history is important. In the World Health Organization (WHO) terminology, KM is defined as “the medicine traditionally practiced in Japan, based on ancient Chinese medicine” [2].

1.2. Kampo formula

In KM, prescribed formula is a weighted combination of multiple crude drugs, and is specified by a list of drug names and their amounts. As an example, “Kakkonto” (pueraria decoction) formula is shown in Table 1.

<table>
<thead>
<tr>
<th>English Name</th>
<th>Latin Name of Original Plants</th>
<th>Amount (g/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pueraria Root</td>
<td>Pueraria lobata</td>
<td>4</td>
</tr>
<tr>
<td>Jujube</td>
<td>Zizyphus jujuba</td>
<td>3</td>
</tr>
<tr>
<td>Ephedra Herb</td>
<td>Ephedra sinica</td>
<td>3</td>
</tr>
<tr>
<td>Glycyrrhiza</td>
<td>Glycyrrhiza uralensis</td>
<td>2</td>
</tr>
<tr>
<td>Cinnamon Bark</td>
<td>Cinnamomum cassia</td>
<td>2</td>
</tr>
<tr>
<td>Peony Root</td>
<td>Paeonia lactiflora</td>
<td>2</td>
</tr>
<tr>
<td>Ginger</td>
<td>Zingiber officinale</td>
<td>2</td>
</tr>
</tbody>
</table>

In Japan, 236 official Kampo formulas are approved by the Ministry of Health, Labour and Welfare (MHLW) and available in an official literature source [3]. The internationally official name for each Kampo prescription is written in Romanized Japanese kana script [4]. (Therefore, Kakkonto in English is “Kakkonto”, not “Pueraria decoction”.) Nevertheless, standardization of KM is not straightforward mainly for two reasons. First, quality of herbal products may vary depending on their source materials, place of cultivation, and their production process. Second, each formula may be personalized according to the health condition of patients. To internationalize KM and to utilize accumulated data for knowledge extraction, medical records need to be digitized and organized systematically.

1.3. Necessity for a public information repository

In the genome-sequencing era, the wealth of clinical information hitherto accumulated will inevitably face the matching amount of genome information such as the recently launched 1000 plants (1KP) or other genome projects [5,6]. An immediate task will be the resolution of discrepancy between biochemical knowledge such as molecular interaction/function and medical knowledge such as hypo/hypertensive activity. In
addition, linking information on active compounds with their biosynthetic genes remains to be another important task [7]. To achieve this goal, the following three steps are necessary [8]:
1. Data accumulation in which quantity is more emphasized than quality;
2. Data curation/organization, i.e., creating ontology by a community effort; and
3. Knowledge extraction from the data for unbiased analysis.
We consider that a wiki-based repository can achieve these steps. The advantage of using wiki is the user-friendliness and flexibility to accommodate various data types such as pictures and chromatograms. It also reduces the cost for data revision and updates. Here we introduce our initial implementation, which also includes information from various collaborators including the members of Kanaya Laboratory at Nara Institute of Science and Technology; Tochimoto Tenkaido Co Ltd; Alps Pharmaceutical Ind. Co Ltd; and NTS Publishing Inc.

2. Methods

2.1. Wiki-based vocabulary control: HTML Forms
The notable difference between the relational database system and wiki systems is the notion of data integrity. Standard wiki is merely a collection of web-pages, in which users can edit contents arbitrarily. To standardize and control user inputs, we need a mechanism to guide users to follow certain ontology. The easiest solution is to support a choice list, i.e., the function to import certified terms from other information sources in order to control vocabulary. In more practical examples, a wiki user should use interactive contents such as drop-down lists and check boxes. Still, the mechanism for such interactive choices should not be predefined, and the word choices be only suggested, not enforced.
Many dynamic web pages utilize the ‘HTML Forms’ to create interactive contents. The W3C recommendation of the ‘Forms’ describes components such as drop-down box, check box, or (radio) button, and they are supported by all web browsers. Obviously, the above mentioned solution can be achieved by integrating the Forms into wiki, and it is also the simpler way to implement interactive components than to use, say, JavaScript.
We implemented the core set of the Forms into our MediaWiki system: specifically, \(<form>\) and \(<input>\) HTML tags for achieving single/multiple choices were allowed inside wiki. However, our goal is not just writing such tags inside the wiki. In their original usage, selection items and tag commands are written within the same HTML page. In our wiki implementation, on the other hand, items to be chosen are imported from another page on demand so that possible selection are externally given and can be shared among multiple pages.
To realize word selection written in external pages, a mechanism to duplicate information between pages is required. The mechanism cannot be achieved by describing a reference function (e.g. by using programs) inside wiki pages, because selected words must be instantiated as a page source instead of being just displayed as a
result of computation. For example, let us imagine implementation of a click button to edit certain data field in a wiki page (Fig. 1).

![Flow of the word choosing process](image)

Figure 1. Flow of the word choosing process. The user invokes a transition by pressing the edit/finish button. Executed programs invoked by HTML Forms must be editable on the user side. Resources on the wiki server are inaccessible and data are copied only between client pages.

When a user clicks this button, another page (or window) for word selection appears, and a user chooses any set of words by using check boxes. After selection, the user returns to the original page by pressing a ‘finish’ button, and the selected words are copied into the original page. After the choice, the selected words can be edited arbitrarily. There are three required parameters in this workflow: (1) location of the page under edit (i.e. where the choice button is); (2) location of the selection list and the finish button; and (3) the list of words currently chosen. In the implementation, two MediaWiki commands were introduced:

```mediawiki
{{#formtag: type | parameter}} and {{#get: parameter}},
```

where `type` is either ‘form’ or ‘input’, and `parameter` is a string directly passed to the HTML tag. Specifically, the `{{#formtag:form}}` command is translated into:

```
<form action="cgi-bin/foo.cgi" method="post">parameter</form>,
```

and the `{{#formtag:input}}` command is translated into `<input parameter>` by the wiki engine.

The `{{#get}}` command is used to obtain parameter values which are posted to the server. Through these commands, the cgi script (indicated as foo.cgi in the above sample code) receives the three parameters: source page, selection page, and the current word choice. When a ‘submit’ option is invoked on the source page, the user will be forwarded to the selection page. The selection page is preprocessed to display interactive components such as check boxes. This preprocessing will be explained in the next subsection.

When the user finishes the selection of words, another ‘submit’ option is invoked. This time, the user will be forwarded back to the source page, and the page must be updated to reflect the new selection. To designate the page contents to be replaced, the source page has a special marker `++variable name++` into which the new selection is inserted. With these settings, it is not hard to implement the required workflow in the cgi program.
2.2. Preprocessing a selection list using the Lua programming language

If the preprocessing part were fixed, i.e. how and where to place check boxes and drop down lists were PHP-coded as in standard ‘special pages’ of MediaWiki, then not much flexibility would be left for end users. Since the selection page is better to be freely designed like normal wiki pages, the layout of form elements such as check boxes and buttons is controlled by the Lua programming language [9]. Lua is a lightweight functional language and is designed as an embeddable scripting engine [10]. In our implementation, a Lua process is called by a special command

```
{{#lua: source code | standard input}}
```

The standard output of the Lua process will be embedded where the command is placed. For security reasons, file-accessing libraries of Lua are nullified by overriding its primitives, and the running time is limited by an external daemon process. Since Lua is a full-fledged programming language, it adds universal computation power to all end users. The only limitations are the running time limit and closed I/O libraries. (Note that the I/O is closed only for the Lua language. The I/O of MediaWiki is left untouched.) In a typical implementation style, a selection mechanism is implemented as a Lua program in one page, and the selection data are located in other pages. This strategy separates edit buttons and related functions from user-defined data, and improves reusability of Lua programs.

2.3. Data collection

Information of crude drugs, their key identifiers (Latin or scientific names), and indications were manually collected from the JP 15th Edition [1] (The most recent version is the 16th Edition, released in April 2011). Kampo formulas, especially the variation in their compositions, were taken from other literature sources [3,11,12]. For the database, we have accumulated information of 158 crude drugs and about 348 Kampo formulas. The full lists of crude drugs and formulas are available on the page “http://metabolomics.jp/wiki/Persist:CrudeDrugList” and “Persist:KampoList” †, respectively. The total number of literature sources exceeded 60, and 7677 academic references were also integrated. Each reference item occupies one wiki page, and is associated with Kampo formulas to which it describes. From every page of Kampo formulas, associated references are linked automatically.

3. Results

3.1. Overview of the Kampo data repository

In Fig. 2, the overview of the Kampo data repository is depicted. Users are expected to visit the top page at “Category:CD”, which is linked to many index pages. Index pages

---

† Hereafter, we indicate only page titles of our wiki site, and the common server address (http://metabolomics.jp/wiki/) will be omitted.
are automatically generated pages by searching keywords in each data group such as names of crude drugs or Kampo formulas. The primary keys in our repository are Latin names for crude drugs and Romanized names for Kampo formulas. Such keys are used to represent explicit links between data pages. When no primary key is shared between data pages, related links are generated by on-demand searches. For example, each Kampo formula is a combination of crude drugs; each formula page can have explicit links to crude drugs. On the other hand, each crude-drug page does not keep track of all information on Kampo formulas, whose contents may change in the repository. Therefore, the crude-drug name in Latin is searched against all formula pages at the time of page rendering; hit pages are listed as links which look like just the same as explicit links for end users. Since the system is wiki-based, each page is freely editable through a browser.

Figure 2. Overview of the Kampo repository. Each square box indicates one wiki page, and overlaid boxes indicate page groups. Solid lines show explicit links. Dashed lines show automatic links generated by page searches. Page titles are in bold face, page contents in standard face, and automatically generated contents in italic face.

### 3.2. Large variation in Kampo formula

Not a few differences in ingredients and their names were found in Kampo formulas from various literature sources. Of total 348 Kampo formulas from over 60 literatures, 30% (108 formulas) were described in a single literature only. Among the rest, only 27% (65) had identical drug ingredients in multiple literature sources. When drug amounts were also considered, only 2% (8 formulas: Dokkatsuto, Jashoshito, Jingyobofuto, Jingyokyokatsuto, Jogan’ippo, Oshosan, Satotsuko, and Shobaito) had exactly the same doses in multiple references. Most varied was Shakanzoto, which had
15 dose variations. In our site, each formula can be accessed by name, for example, as “Kampo:Shakanzoto” on the wiki platform. This result does not mean, however, that KM is unorganized. First, Kampo formula is originally designed to be prescribed for each patient by considering symptoms and conditions. Second, some crude drugs can be considered almost identical. For example, (Kan)Shokyo (dried ginger) and Kankyo (steamed and dried ginger) are made from the same ginger root and only the production process is different (steaming or not). The same is true for Kippi and Chimpi (both from sour orange peel) or Keihi and Keishi (both from cinnamon bark). When these names were considered identical, the number of inconsistent formulas was reduced to less than 50. Typical variation was addition of optional ingredients. Even after this merging process, we could find several unique formulas in the book titled “Natural Medical Resource 2nd Edition” (in Japanese). To give some examples, Choijokito prescribed for constipation contains Uncaria hook, which is used as vasodilator. Choreito for kidney disorder contains Angelica root which is used for feminine ailment. It is noteworthy that we could locate these putative errors through integration and comparison with other literature sources.

3.3. Extraction of modern Kampo usage

For each Kampo formula, related keywords were extracted from associated reference titles. Frequent co-occurrence of Japanese keywords, for example more than 4 in reference titles, already produced observations such that “Anchusan” is used as parasiticide against anisakis, “Kakkonto” against influenza, “Tokishakuyakusan” for dementia, or “Chotosan” for hypertension. Currently each formula page in our database displays automatically extracted keywords from reference titles. Such words are not described in official indications but reflect modern trends of Kampo usage. There is a severe limitation for on-demand computation inside wiki pages, but if such relationships are once output (e.g. downloaded from the wiki site), they can be further used for network analysis.

3.4. Linking with plant taxa and English information

Each crude drug is linked with information of original plants (currently no information for animals or minerals). Also available is statistical information at the plant family level. Although details of crude drugs have been available only in Japanese by tradition, the database contains English information for 148 formulas taken from the special edition of Kampo, Acupuncture and Integrated Medicine “Current Kampo Medicine” (2006) [13] and 119 English package inserts for ethical use compiled by Japan Pharmaceutical Reference ‡. Page titles are in Latin names and additional herbal information is also available in English. Bioactivities and reference information are

‡ Its full list is shown on the “Index:JPR” page.
organized through the mechanisms in Section 2, and registered users can edit both data processing programs in Lua as well as data themselves.

It is worth mentioning that the drug- and Kampo-list pages on our server are not always created and curated manually: data redundancy can be minimized as in the relational database system (RDB). Except for key identifiers corresponding to primary keys in RDB, the data are only referenced inside multiple pages through the extensive use of key identifiers and inline page search commands [9]. For example, pages for crude drugs contain information of related Kampo formulas, and these links are automatically generated by page searches. Similarly, composition of Kampo formulas is automatically linked to each crude-drug page. Therefore, wiki editors do not need to care about manual addition of error-prone hyperlinks. Since all functions are implemented at the user level, a user can verify details of the computation flow by browsing page sources and can modify or copy them if necessary.

4. Discussion

4.1. Related databases in Asian countries

After much discussion on biodiversity and environment protection, national institutions in major East Asian countries commenced collection of information on traditional medicines. In Japan, Research Center for Medicinal Plant Resources (National Institute of Biomedical Innovation) launched a database project on chemical constituents, biosynthetic genes, biological activities of 75 medicinal plants in collaboration with several Japanese universities in 2010 [14]. The center also released a pilot database for medicinal plants early in 2010 [15]. Korea Institute of Oriental Medicine designed a portal site OASIS (Oriental Medicine Advanced Searching Integrated System) for establishing an alliance among Korean research institutions on medicinal plants [16]. China Academy of Traditional Chinese Medicine provides Traditional Chinese Medicine Database System, which includes literature information (more than 120,000 references from 800 journals) , materia medica (herbal medicine), medical formula, clinical medicine, and so on [17]. The often mentioned drawback of these websites is their exclusive nature. Data access is usually restricted to collaborators. It is very hard for foreign users to navigate these sites. We have contacted Research Center for Medicinal Plant Resources in Japan and agreed to collaborate for creating a more open, user-friendly database system. Our wiki design will therefore contribute to the Japan-wide consortium of the medicinal plant resources. Indeed, the crude drug information will become a part of new database system at Institute of Natural Medicine, University of Toyama under the authorization of the Medical and Pharmaceutical Society for WAKAN-YAKU in Japan.
4.2. Relationship with semantic webs and XML

Semantic web is the next-generation web standard and its quintessence are hyperlinks with predicates: each link is assigned a predicate so that the relationship of links is machine-interpretable. Inline page searches are also supported in several implementations (for example, see the “SemanticWiki” page in English Wikipedia). Although there is no doubt about its superiority over current monotone hyperlinks, ontology of predicates themselves becomes an important issue for discussion. Since each data provider can add custom-made predicates, streamlining and understanding predicates is even more difficult than to standardize original data. Interested readers are recommended to visit websites for predicate ontology such as the Open Biological and Biomedical Ontologies (OBO) [18] or Vocabulary of Interlinked Datasets (VOID) [19]. The disadvantage of XML is similar. It bloats up the data amount and the custom set of tags or predicates is sometimes complicated. An easy-to-understand mechanism to manage predicates will be needed and our proposal fits into this niche. Our idea does not conflict with Semantic Wiki or XML. It can be utilized not only for standard wiki-based sites but also for such semantic-wiki sites.

Not a few biological databases have been converted to the wiki style [20], but an approach to manage ontology through inline page search or dynamic page copying is unheard of. By extending data contents to metabolome and genome information, we plan to develop interdisciplinary information portal using this approach.

4.3. Limitations of the wiki-based approach

The proposed approach uses full-text search extensively. The search is performed on demand to collect necessary information (e.g. statistics) and to create links to other pages. The strategy is therefore not scalable. (In MediaWiki, page contents are recommended to be less than 30 Kbytes in each page by default.) Users do not suffer from navigation delay only when the number of pages remains less than tens of thousands; pre-computation such as using a suffix array will be necessary for a larger scale. The pre-computation is not straightforward, however, because the update needs to be on-demand and dynamic. The development of dynamic data structures is a major research topic to improve scalability. The wiki-based approach also does not fit with large-scale data processing using application program interface (API) due to its flexible format. Although users can download full page contents using the original interface of MediaWiki (see its user manual for details), much efforts will be needed to integrate the system into conventional style of data federation using fully defined formats and protocols. Since MediaWiki uses SQL database in its backend, using the wiki-based interface only for human-computer interaction is the likely solution.

5. Conclusion

We created a database for crude drugs and Kampo medicine using a wiki system. In biology, there are numerous wiki-based database projects. In a standard setting, each
page is edited independently and detection of data duplicates or installation of ontology is difficult. In such situations, our approach would be an easy choice than to introduce a whole new system and to force all users to reformat their data accordingly.

Acknowledgments

This work is partially supported by JST Institute for Bioinformatics Research and Development and Grant-in-Aid for Publication of Scientific Research Results & Scientific Research on Priority Areas “Systems Genomics” from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

http://www.mhlw.go.jp/topics/bukyoku/iyaku/yakkyoku/english.html
Available online at http://www.jsom.or.jp/html/magazine.html
[5] PlantGDB by the National Science Foundation, USA. http://www.plantgdb.org/
http://www.yakuji.co.jp/entry20132.html (last accessed Jan 21, 2011)
    http://cowork.cintcm.com/engine/wdbintro1.jsp
[18] The Open Biological and Biomedical Ontology: http://www.obofoundry.org/