Mapping the Terms of Medicinal Material and Formula Classification to International Standard Terminology

Jin-Hyun Kim, Chul Kim, Sang-Jun Yea, Hyun-Chul Jang, Sang-Kyun Kim, Young-Eun Kim, Chang-Seok Kim, Mi-Young Song*

Information Research Center, Korea Institute of Oriental Medicine, Yuseong-gu Daejeon, 305-811, Korea

ABSTRACT

The current study aims to analyze the acceptance of International Standard Terminology (IST) related to herbs and formulas used in Korea. It also intends to examine limitations of each term source by linking texts for herbal medicine research and formula research used in schools of oriental medicine with medicinal substance-formula classification names within the IST framework. This study examined 64 medicinal classification names of IST, including synonyms, 41 formula classification names, 65 classification names of “Herbal Medicine Study,” 89 medicinal classification names of “Shin’s Clinical Herbal Medicine Study,” and lastly 83 formula classification names of “Formula Study.” Data on their chief virtue, efficacy and characteristics as medicinal substances were extracted from their definitions, and such data were used to perform Chinese character-English mapping using the IST. The outcomes of the mapping were then analyzed in terms of both lexical matching and semantic matching. In terms of classification names for medicinal substances, “Herbal Medicine Study” had 60.0% lexical matching, whereas “Shin’s Clinical Herbal Medicine Study” had 48.3% lexical matching. When semantic matching was also applied, “Herbal Medicine Study” showed a value of 87.7% and “Shin’s Clinical Herbal Medicine Study” 74.2%. In terms of formula classification names, lexical matching was 28.9% of 83 subjects, and when semantic matching was also considered, the value was 30.1%. When the conceptual elements of this study were applied, some IST terms that are classified with other codes were found to be conceptually consistent, and some terms were not accepted due to different depths in the classification systems of each source.

Keywords: Medicinal substance-formula classification names, International standard terminology, Oriental medicine

1. INTRODUCTION

The Western Pacific Regional Office (WPRO) of the World Health Organization (WHO) is implementing a series of projects in various areas in order to standardize traditional medicine (TM), and WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region (IST) [1] is an outcome within the standardization framework, produced using an evidence-based approach. The IST initiative aims to unify various English expressions of TM and use standard terminology in education, training, practice and research. The IST was officially published in 2007 after three consultation meetings in 2004 and 2005 involving experts of traditional Korean (TKM), Chinese (TCM), Japanese (Kampo) and Vietnamese medicine (TVM), followed by an additional two years of editorial work [2]-[4].

The IST has a total of 3,259 terms from basic theories, diagnostics, various therapeutics including acupuncture-moxibustion, and classic titles [2],[3]. On the other hand, acupuncture point names and locations were excluded from the IST because there are relevant existing WHO standards [5]. Medicinal materials were excluded from the IST as well, because each country’s domestic standards on pharmacopoeia and provisions [6-8] differ from one another, and formula titles of herbal medicinal prescriptions follow country-specific rules, which presents limitations for a conference of academic bodies to deal with [3],[4]. A separate initiative to standardize nomenclature, standards and experimentation is being pursued by the Forum on Harmonization of Herbal Medicine (FHH), a separate entity under WHO/WPRO [9]. Therefore, while medicinal substance and formula nomenclature are not included in the IST, it still contains 207 terms for medicinal substances and formulas, laying the foundation to introduce standard terms, albeit limited to certain areas of herbal medicine.

An IST such as this serves as a basis for establishing and amending domestic and group standards. For instance, a Korean version [10] of standard acupuncture point locations was published for domestic use, and texts [11],[12] used in colleges of oriental medicine have been amended to be actively utilized in education and research in the fields of point locations and acupuncture in Korea [13]. However, because only a small number of medicinal material- and formula-related terms are listed with the IST and their areas are limited, few terms have been linked to the terms used domestically or reflected in texts for education or research.

Therefore, this study aims to extract classification names...
from medicinal substance-formula terms found in “Herbal Medicine Study” [14] and “Shin’s Clinical Herbal Medicine Study” [15], which are texts for herbal medicine used at domestic schools of oriental medicine, and “Formula Study” [16], which is also a text for formula study. This study goes on to link these terms with those listed with the IST in order to analyze how many IST terms are included in the actual education terms for classification names used in the oriental medicine field in Korea and to examine limitations of each source. Ultimately, we hope to devise ways to ensure wider use of herbal medicine-formula texts as part of the IST.

2. MATERIALS & METHODS

2.1. Collection of IST medicinal-formula classification names

IST consists of a code, an English term, a Chinese term and definitions and descriptions in English. The terms are distinguished based on a concept, and Chinese terms are used to identify the English terms. The codes used in the IST follow a three-depth classification system, and the first depth is comprised of eight categories - General, Basic Theories, Diagnostics, Disease, Therapeutics, Acupuncture & Moxibustion, Medicinal Treatment, and Classics [2]-[4]. Among these, medicinal substance-formula terms are 207 terms belonging to the Medicinal Treatment category, and 112 of these are medicinal terms, whereas the remaining 95 are formula terms [2],[3]. Among the terms that belong to the third depth (sections) such as medicinal principles, processing and preparation form, 64 medicinal classification names and 41 formula classification names including synonyms were collected and each term was assigned a unique ID to which an IST code was applied as in [IST_ist code_H#] or [IST_ist code_F#]. From definitions described in English, chief value, efficacy and other information (nature-flavor of a medicine, medicinal ingredients included in the formula) were extracted (Table 1).

<table>
<thead>
<tr>
<th>ID</th>
<th>Term</th>
<th>Chinese</th>
<th>Chief value</th>
<th>Efficacy</th>
<th>Nature of medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>IST_6.1.49_H1</td>
<td>exterior-releasing medicinal</td>
<td>解表藥</td>
<td>external pathogen</td>
<td>the effect of dispelling external pathogen sweating</td>
<td></td>
</tr>
<tr>
<td>IST_6.1.50_H2</td>
<td>wind-cold-dispersing medicinal</td>
<td>散風寒藥</td>
<td>exterior pattern/syndrome</td>
<td>the effect of dispersing wind and cold</td>
<td></td>
</tr>
<tr>
<td>IST_6.1.51_H3</td>
<td>pungent-warm exterior-releasing medicinal</td>
<td>辛溫解表藥</td>
<td>wind-cold exterior pattern/syndrome</td>
<td>treating a wind-cold exterior pattern/syndrome</td>
<td>exterior-releasing medicinal pungent in flavor and warm in property</td>
</tr>
<tr>
<td>IST_6.1.52_H4</td>
<td>wind-heat dispersing medicinal</td>
<td>散風熱藥</td>
<td>exterior pattern/syndrome</td>
<td>the effect of dispersing wind and heat</td>
<td>exterior-releasing medicinal pungent in flavor and cool in property</td>
</tr>
<tr>
<td>IST_6.1.53_H5</td>
<td>pungent-cool exterior releasing medicinal</td>
<td>辛涼解表藥</td>
<td>wind-heat exterior pattern/syndrome</td>
<td>treating a wind-heat exterior pattern/syndrome</td>
<td></td>
</tr>
<tr>
<td>IST_6.2.62_F1</td>
<td>exterior-effusing formula</td>
<td>發表剤</td>
<td>exterior patterns/syndromes</td>
<td>diaphoretic, muscle-releasing and eruption-promoting effects</td>
<td>formula that is composed of exterior-releasing medicinals</td>
</tr>
<tr>
<td>IST_6.2.63_F2</td>
<td>emetic formula</td>
<td>漲吐剤</td>
<td>phlegm syncope, food accumulation, and ingestion of poisons</td>
<td>formula that induces vomiting</td>
<td></td>
</tr>
<tr>
<td>IST_6.2.63_F3</td>
<td>emetic formula</td>
<td>吐剤</td>
<td>synonym of 漲吐剤</td>
<td>formula mainly composed of purgatives</td>
<td></td>
</tr>
<tr>
<td>IST_6.2.64_F4</td>
<td>interior-attacking formula</td>
<td>攻裏剤</td>
<td>interior excess patterns/syndromes</td>
<td>bowel-moving, heat-purging, accumulation-attacking and water-discharging effects</td>
<td></td>
</tr>
</tbody>
</table>

2.2 Collection of medicinal classification

In general, a medicinal includes herbal, animal and mineral medical substances, but herbal medicinal substances account for most of the medicinal classes and hence is commonly referred to as herbs.

There are several methodologies to classify the medicinal substances. Of these, “Shennongbencaojing” [17] distinguished them using three grades – top, medium and low – in terms of longevity, life nurturing and disease treating, whereas “Bencaogangmu” [18] presented 17 categories including water, fire, earth, metal and stone based on the medicinal substances natural state – animal, herbal or mineral – and on its origin. Meanwhile, “Yixuerumen” [19] classified the medicinal substances into seven groups depending on its efficacy such as...
wind-dispelling, heat-clearing and dampness-draining. Also Hwang has seven categories including tonifying, astringent, dispersing and purgative, whereas Ga’s classification includes qi – blood, five viscera, phlegm and fire. Kang’s classification centered on viscera-bowels such as heart, liver, spleen, lung and kidney [15].

The subjects of this study, “Herbal Medicine Study” [14] and “Shin’s Clinical Herbal Medicine Study” [15], classified the medicinal substances centering on their efficacy. “Herbal Medicine Study” [14] lists 65 classification names including synonyms, whereas “Shin's Clinical Herbal Medicine Study” [15] contains 89 classification names, and each classification name was assigned a unique set of IDs, [Herbal_H#] and [Clinical_H#], consecutively for collection. They all have a two-depth classification system, and considering this, terms for chief virtue and efficacy, which are essential bases for conceptual distinction, were extracted from a classification name’s definition, and nature-flavor information of a medicinal substance, where available, was collected additionally in order to utilize it in the mapping of the same yet more sophisticated concepts (table 2).

Table 2. Examples of collected data from “Herbal Medicine Study” and “Shin's Clinical Herbal Medicine Study”

<table>
<thead>
<tr>
<th>ID</th>
<th>Chinese Hierarchy</th>
<th>Chief value</th>
<th>Efficacy</th>
<th>Nature of medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herbal_H1</td>
<td>解表藥</td>
<td>表證</td>
<td>发汗解表</td>
<td>辛</td>
</tr>
<tr>
<td>Herbal_H2</td>
<td>發散風寒藥</td>
<td>表寒證</td>
<td>明溫解表/發散風寒</td>
<td>辛温-warm</td>
</tr>
<tr>
<td>Herbal_H3</td>
<td>發散風熱藥</td>
<td>風熱表證</td>
<td>明涼解表/發散風熱</td>
<td>辛凉</td>
</tr>
<tr>
<td>Clinical_H15</td>
<td>解表藥</td>
<td>表證</td>
<td>发汗和表邪發散</td>
<td>辛温</td>
</tr>
<tr>
<td>Clinical_H16</td>
<td>辛溫解表藥</td>
<td>表寒證,風寒表實證</td>
<td>明溫解表</td>
<td>辛温-warm</td>
</tr>
<tr>
<td>Clinical_H17</td>
<td>發散風寒藥</td>
<td>表寒證</td>
<td>明涼解表</td>
<td>辛凉</td>
</tr>
<tr>
<td>Clinical_H18</td>
<td>辛涼解表藥</td>
<td>表寒證</td>
<td>明涼解表</td>
<td>辛凉</td>
</tr>
</tbody>
</table>

2.3 Collection of formula classification

Medical professionals throughout history have attempted to categorize and organize formulas using diverse standards to ensure better understanding of formulas and convenience of use. “Huangdirineijing” [20] placed formulas into seven categories including major-minor, relax-tension depending on the seriousness of illness, body part, pace, and number of medicinal substances used, whereas “Shengjizonglu” [21] offers ten categories including diffusion, unblock, tonifying, and heavy-light based on efficacy. Meanwhile, “Jingyuequanshu” [22] and “Yixueixinwu” [23] classified formulas based on eight laws (principles) of treatment, whereas “Yifangjijie” provided 22 categories including tonifying-nourishing, exterior-effusing, harmonizing-releasing, and regulating qi, similar to today’s categorization [24]. From “Formula Study” [16], the study’s subject, 83 classification names including synonyms were collected, and they were assigned a unique ID [Formula_F#] in a consecutive manner. They have a three-depth classification system; considering this, chief virtue and efficacy of a formula, which are essential bases for conceptual differentiation, as well as characteristics of medicinal ingredients, an additional basis for conceptual differentiation, were extracted from the definitions of classification names (table 3).

Table 3. Examples of collected data from “Formula Study”

<table>
<thead>
<tr>
<th>ID</th>
<th>Chinese Hierarchy</th>
<th>Chief value</th>
<th>Efficacy</th>
<th>Nature of medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula_F1</td>
<td>解表劑</td>
<td>表證</td>
<td>发散解表</td>
<td>exterior-releasing medicinals</td>
</tr>
</tbody>
</table>


2.4 Mapping of classification names of IST and text

In order to map the IST written in English and the classification names written in Chinese in texts, the chief virtue, efficacy and medicinal properties of medicinal substances, which are essential conceptual differentiation factors, were linked to each other. For instance, 風寒感冒 (Herbal_H2) has 外感風寒 as its chief virtue and 辛溫解表 as its efficacy and medicinal property. For these elements, their IST equivalents, which are essential conceptual differentiation factors, were linked to each other. For instance, 解感風寒 (Herbal_H2) has 外感風寒 as its chief virtue and 辛溫解表 as its efficacy and medicinal property. For these elements, their IST equivalents, which are essential conceptual differentiation factors, were linked to each other. For instance, 解感風寒 (Herbal_H2) has 外感風寒 as its chief virtue and 辛溫解表 as its efficacy and medicinal property.

Regarding classification names of medicinal substances, some of the mapping results involving 64 terms from the IST, 65 terms of “Herbal Medicine Study” and 89 classification names from “Shin’s Clinical Herbal Medicine Study” are shown in the top portion of Table 4. Representative concepts were established using 21 categories and 27 divisions in total, and most of such concepts are based on the IST terms. Lexical matching for “Herbal Medicine Study” and “Shin’s Clinical Herbal Medicine Study” was 60.0% and 48.3% respectively, whereas when semantic matching was also included, the values were 87.7% and 74.2% respectively, for “Herbal Medicine Study” and “Shin’s Clinical Herbal Medicine Study.” Also 29 concepts including 温化寒痰藥 (Herbal_H40), 清化熱痰藥 (Herbal_H41), 止渴藥 (Clinical_H81) and 柔肝藥 (Clinical_H37) could not be mapped to IST terms. On the contrary, three IST terms, wind-dampness-dispelling and cold dispersing medicinal: 发散风寒 (IST_6.1.49), wind-dampness-dispelling and heat clearing medicinal: 发散风热 (IST_6.1.67_H19), and liver-emolliating medicinal: 柔肝藥 (IST_6.1.104_H59), were not used in the texts.

For formula classification names, 41 terms were mapped with classification names of “Formula Study,” and the results are shown at the bottom of Table 4. Of the total 83 formula classification names, lexical matching was 28.9%, and was 30.1% when semantic matching was also included. The reason why IST’s coverage is so low is because the IST’s formula classification names use the terms that belong to categories of “Formula Study,” and the sub-terms that belong to sections are rarely listed. 83 classification names consist of 22 categories and 62 sections. Of these 21 category names representing 87.5% of all categories and 3 sections names – altogether 24 terms – were lexically matched.

Table 4. Examples of mapping results

<table>
<thead>
<tr>
<th>No.</th>
<th>IST concept in supercategory</th>
<th>Mapped concept in supercategory</th>
<th>No.</th>
<th>IST concept in supercategory</th>
<th>Mapped concept in supercategory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>解表藥 IST_6.1.49_H1</td>
<td>解表藥_Clinical_H15</td>
<td>1-1</td>
<td>發散風寒藥 IST_6.1.50_H2</td>
<td>辛溫解表藥 IST_6.1.51_H3</td>
</tr>
<tr>
<td></td>
<td>解表藥_H1</td>
<td></td>
<td></td>
<td>解表藥_H1</td>
<td></td>
</tr>
</tbody>
</table>
4. DISCUSSION & CONCLUSIONS

The purpose of this paper is to examine the IST’s coverage and limitations of the terms used in the domestic education of oriental medicine by liking classification names found in texts of herbal medicine study and formula study used in Korean schools of oriental medicine with the IST and to suggest ways of improvement based on such examination.

To this end, this study examined the IST’s 64 medicinal classification names and 41 formula classification names, 65 medicinal classification names from “Herbal Medicine Study” [14], 89 medicinal classification names from “Shin’s Clinical Herbal Medicine Study” [15] and finally, 83 formula classification names from “Formula Study” [16]. From their definitions, data on each material’s chief virtue, efficacy and medicinal properties were extracted, and based on this data and using the IST and dictionaries, Chinese-English mapping was performed. The mapping results were then analyzed by dividing them into lexical matching with identical strings and into semantic matching in which the terms have different strings but the same conceptual elements.

In terms of classification names for medicinal substances, lexical matching for “Herbal Medicine Study” [14] and “Shin’s Clinical Herbal Medicine Study” [15] was 60.0% and 48.3% respectively, whereas when semantic matching was also included, it was 87.7% and 74.2% respectively, for “Herbal Medicine Study” and “Shin’s Clinical Herbal Medicine Study.” The mapping rate for “Shin’s Clinical Herbal Medicine Study” was somewhat lower than that of “Herbal Medicine Study” because its classification names include various synonyms and section names, and it follows a dual-classification system using diverse criteria based on efficacy, as in the case of 淨瘀劑 (Clinical_H81) and 止瘀藥 (Clinical_H80).

Some issues that arose from the mapping process of the medicinal substances’ classification names are as follows:

First, when semantic matching was attempted by using a medicinal substance’s efficacy (including its nature-flavor) or its chief virtue as conceptual elements, some terms were defined as different concepts in the IST and thus were assigned separate codes but were bound as the same concepts. For instance, wind-cold-dispersing medicinal: 發散風寒藥 (IST 6.1.50_H2) is defined as “a medicinal that has the effect of dispersing wind and cold in the treatment of an exterior pattern/syndrome,” while pungent-warm exterior-releasing medicinal: 辛溫解表藥 (IST 6.1.51_H3) is defined as “an exterior-releasing medicinal pungent in flavor and warm in property, used for treating a wind-cold exterior pattern/syndrome.” Because the two had the same chief virtue (wind-cold exterior pattern/syndrome) and efficacy (dispersing wind and cold), they were classified as the same concept. The same principle was applied to strangury-relieving diuretic medicinal: 利尿通淋藥 (IST 6.1.71_H24) / strangury-relieving medicinal: 通淋藥 (IST 6.1.72_H25), blood-activating and stasis-resolving medicinal: 活血化瘀藥 (IST 6.1.83_H48) / blood-activating and stasis-resolving medicinal: 活血化瘀藥 (IST 6.1.84_H39) / blood-activating medicinal: 化瘀藥 (IST 6.1.86_H40), and exterior-securing anhidrotic medicinal: 固表止汗藥 (IST 6.1.107) / sweat-constraining exterior-securing medicinal: 敷汗固表藥 (IST 6.1.108). While the two texts include definitions of classification names as well as data on the medicinal substances concerned, the IST should depend on the definitions based on conceptual elements of the classification names, and it would be much clearer to define two terms as the same concept if they have an identical chief virtue or efficacy (medicinal properties), which constitute a definition.

Also, there were cases where mapping was impossible because sections had different bases for classification. Wind-dampness-dispelling medicinal: 烏風濕寒藥 (IST 6.1.65) is classified into wind-dampness-dispelling and cold-dispelling medicinal: 烏風濕寒藥 (IST 6.1.66_H18) and wind-dampness-dispelling and heat-dispelling medicinal: 烏風濕清熱藥 (IST 6.1.67_H19) based on its medicinal properties in IST; both the texts used in Korea categorized it into 止咳藥 (Clinical_H64) / 烏風濕止咳藥
Moreover, while some IST terms include synonyms, the formula classification names in the two texts had no information on synonyms. Building synonyms is an important part of terminology-related studies as it allows a wider use of terms and they serve as useful data in securing interoperability among terms through conceptual linkage with other sources. Therefore, it is necessary to establish synonym data for the classification names that serve as titles of formulas contained in “Formula Study,” the representative formula study educational material in Korea, in order to allow the use of more diverse and rich terminology.

When semantic matching was attempted for the formulas as well, some concepts were defined as different in the IST yet bound as identical, as in the case of purgative formula: 

<table>
<thead>
<tr>
<th>IST_6.2.72_F14</th>
<th>interior-attacking formula:</th>
</tr>
</thead>
<tbody>
<tr>
<td>彭下剤</td>
<td>攻表剤</td>
</tr>
<tr>
<td>(Purgative formula: Within the IST)</td>
<td>(Purgative formula: Outside the IST)</td>
</tr>
</tbody>
</table>

Because IST defines the purgative formula as a sub-concept of “interior-attacking formula,” the semantic matching engine recognized them as different terms, whereas their meanings are almost identical. Furthermore, some IST terms include synonyms, the formula classification names in the two texts had no information on synonyms. Building synonyms is an important part of terminology-related studies as it allows a wider use of terms and they serve as useful data in securing interoperability among terms through conceptual linkage with other sources. Therefore, it is necessary to establish synonym data for the classification names that serve as titles of formulas contained in “Formula Study,” the representative formula study educational material in Korea, in order to allow the use of more diverse and rich terminology.

We hope that this pilot study will serve as a starting point to spread the use of the IST terms related to herbal medicine study and formula study, not only in education but also in clinical settings. Also, efforts should be made in the future to find effective measures to address the issues examined in this study.

5. ACKNOWLEDGEMENT

This research was supported by “Ontology-based Traditional Korean Medicine Knowledge Framework” and “the Study on Reasoning Algorithms of Medicinal Materials based on a Mathematical Prognosis Model” projects. They were funded by the Korea Institute of Oriental Medicine (KIOM) grant K11090 and C11060.

6. REFERENCE


Kim Jin-Hyun
She received the B.S. in mathematical education and oriental medicine from Ehwa woman's university and Daegu haany university, Korea in 2001, 2009 respectively. From 2009, she has been with the Information Research Center, Korea Institute of Oriental Medicine. Her main research interests include oriental medicine contents and terminological standards.

Kim Chul
He received the B.S., M.S in industrial engineering from KAIST, Korea in 1998, 2000 respectively and also received Ph.D. in oriental medicine informatics from Wonkwang university, Korea in 2009. From 2006, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include RFID/USN and oriental medicine informatics.

Yea Sang-Jun
He received the B.S., M.S in computer science from KAIST, Korea in 2002, 2004 respectively. From 2008, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include computer vision and oriental medicine contents.

Jang Hyun-Chul
He received the B.S., M.S in computer engineering from Chungnam national university, Korea in 1996, 1999 respectively and also received Ph.D. in computer engineering from Chungnam national university, Korea in 2007. From 2008, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include bioinformatics and semantic web.

Kim Sang-Kyun
He received the B.S., M.S in computer engineering from Chungnam national university, Korea in 1999, 2001 respectively and also received Ph.D. in computer engineering from Chungnam national university, Korea in 2008. From 2007, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include semantic web and database.
Kim Young-Eun
She received the B.S., M.S in Korean traditional medicine from Wonkwang university, Korea in 2007, 2010 respectively. From 2011, she has been with the Information Research Center, Korea Institute of Oriental Medicine. Her main research interests include oriental medicine contents.

Kim Chang-Seok
He received the B.A. in Chinese language and literature from Chungnam national university, Korea in 2001 and also received M.A. in archival science from Chungnam national university, Korea in 2011. From 2009, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include oriental medicine information and contents.

Song Mi-Young
She received the B.S in chemistry from Sookmyung woman’s university, Korea in 1984 and also received M.S, Ph.D. in chemistry from KAIST, Korea in 1981, 1991 respectively. From 2004, she has been with the Information Research Center, Korea Institute of Oriental Medicine. Her main research interests include oriental medicine contents and bioinformatics standards.