Applying Traditional Korean Medical Terms to SUI in the Unified Medical Language System(UMLS) Metathesaurus

Seong-Cheon Hong[1], Heon-Young Jeong[2], Byong-Uk Jeon[3]

[1]Professional Graduate School of Oriental Medicine, Wonkwang University

[2]College of Oriental Medicine, Wonkwang University

[3]Korea Institute of Oriental Medicine

Objective: Various controlled vocabulary such as thesaurus and classification make us to reuse and share effectively by defining different concept and linking terms each other. The UMLS(Unified Medical Language System) is one of the most universal medical terminology systems. It is needed various methods to share and reuse information of traditional Korean medicine. We will research on method that adopt SUI of the UMLS(that is de facto standard in medical terminology system) in traditional Korean medical terminology.

Method: We described major problems and applying process when we tried to add traditional Korean medicine in the part of meridian into the UMLS metathesaurus. Comparing western medical terms and traditional Korean medical terms for applying UMLS metathesaurus, there is not only many consistency, but also differences.

Result: We confirmed what is the differences and consistency between western medical terms and traditional Korean medical terms. And then reviewed methods that apply the CUI, LUI, SUI in traditional Korean medical terms. Traditional Korean medical terms are not discriminated by singular or plural string. In addition, traditional Korean medical terms have vary string by initial law: the law of initial sound of a syllable. Character is described with Korean, traditional Chinese, modern Chinese, etc. According to meaning, language, initial law, SUI has a distinct value respectively.

Conclusion: There are many differences to apply the UMLS between western medical terms and traditional Korean medical terms. For the better implementation to traditional Korean medicine into the UMLS, further research is needed in standardization and classification of traditional Korean medical terms, medical information system, etc. We hope this study helps the implementation UMLS, EHR, knowledge based system in Oriental medicine in the future.

Key words : UMLS, Metathesaurus, Semantic Network, traditional Korean Medical terms

Introduction

Information retrieval system is a major issue for the international use of medical knowledge.

Various controlled vocabulary such as thesaurus and classification make us to reuse and share effectively by defining different concept and linking terms each other.

National Library of Medicine has developed the Unified Medical Language System(UMLS) to solve the problem of information retrieval

and integration resulted from the difference of concepts between different sources.[1]

Despite a few research (including Kim[2] and Han's[3] research) for development in korean medical knowledge, no traditional Korean medical language is considered in medical knowledge system.

In this paper, we researched the process of integration, problems and solutions for applying traditional Korean medical terms in the part of meridian to UMLS metathesaurus by focusing on CUI, SUI, LUI.

Background

The UMLS facilitates the development of computer systems that behave as if they understand the language of biomedicine and health. To that end, NLM produces and distributes the UMLS Knowledge Sources (databases) and associated software tools (programs). Developers use the knowledge sources and tools to build or enhance systems that create, process, retrieve, and integrate biomedical and health data and information. The knowledge sources are multi-purpose and are used in systems that perform diverse functions involving information types such as patient records, scientific literature, guidelines, and public health data. The associated software tools assist developers in customizing or using the

UMLS Knowledge Sources for particular purposes. The lexical tools work more effectively in combination with the UMLS Knowledge Sources, but can also be used independently.[4]

There are three UMLS Knowledge Sources: the Metathesaurus, the Semantic Network, and the SPECIALIST Lexicon.

The Metathesaurus is a very large, multi-purpose, and multi-lingual vocabulary database that contains information about biomedical and health-related concepts, their various names, and the relationships among them. It is built from the electronic versions of many different thesauri, classifications, code sets, and lists of controlled terms used in patient care, health services billing, public health statistics, indexing and cataloging biomedical literature, and/or basic, clinical, and health services research.

The term Metathesaurus draws on Webster's Dictionary third definition for the prefix "meta," i.e., "more comprehensive, transcending." In a sense, the Metathesaurus transcends the specific thesauri, vocabularies, and classifications it encompasses.

The Metathesaurus is organized by concept or meaning. In essence, it links alternative names and views of the same concept and identifies useful relationships between different concepts.

The Metathesaurus is organized by concept. One of its primary purposes is to connect different names for the same concept from many different vocabularies. The Metathesaurus assigns several types of unique, permanent identifiers to the concepts and concept names it contains, in addition to

1) Hye-Sun Kim. A Review of Structure and Application of UMLS. Research of information management. 2001;32(2):26-39
2) Hye-Sun Kim. A Review of Structure and Application of UMLS. Research of information management. 2001;32(2):26-39
3) Seong-Bin Han, Jin-Wook Choi. The comparative study on concept representation between the UMLS and the clinical terms in Korean Medical Records. IJMI 2005;74:67-76
4) UMLS Knowledge Sources. Bethesda(MD). National Library of Medicine. 2007AA

retaining all identifiers that are present in the source vocabularies. The Metathesaurus concept structure includes concept names, their identifiers, and key characteristics of these concept names.

Each concept or meaning in the Metathesaurus has a unique and permanent concept identifier(CUI). Each unique concept name or string in each language in the Metathesaurus has a unique and permanent string identifier(SUI). Any variation in character set, upper-lower case, or punctuation is a separate string, with a separate SUI. The same string in different languages will have a different string identifier for each language. The basic building blocks or "atoms" from which the Metathesaurus is constructed are the concept names or strings from each of the source vocabularies. Every occurrence of a string in each source vocabulary is assigned a unique atom identifier(AUI). Each string is linked to all of its lexical variants or minor variations by means of a common term identifier(LUI). Like a string identifier, the LUI may be linked to more than one concept. This occurs when strings that are lexical variants of each other have different meanings. In contrast, each string identifier and each atom identifier can only be linked to a single LUI.

In the Metathesaurus, every CUI(concept) is linked to at least one AUI(atom), SUI(string), and LUI(term), but can also be linked to many of each of these. Every AUI is linked to a single SUI, a single LUI, and a single CUI. Each SUI can be linked to many AUIs, to a single LUI, and to more than one CUI – although the typical case is one CUI. Each LUI can be linked to many AUIs, many SUIs, and more than one CUI – although the typical case is one CUI.

In the abbreviated example in Table 1, Atrial Fibrillation appears as an atom in more than one source vocabulary and has a distinct AUI for each occurrence. Since each of these atoms has an identical string or concept name, they are linked to a single SUI. Atrial Fibrillations, the plural of Atrial Fibrillation, has a different string identifier. Since the singular and plural are lexical variants of each other, both are linked to the same LUI. There is a different LUI and different SUIs and AUIs for Auricular Fibrillation and its plural Auricular

Table 1. example of CUI, LUI, SUI, AUI in metathesaurus

Concept (CUI)	Terms (LUIs)	Strings (SUIs)	Atoms (AUIs) * RRF Only
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation (preferred) Atrial Fibrillations	S0016668 Atrial Fibrillation (preferred)	A0027665 Atrial Fibrillation(from MSH) A0027667 Atrial Fibrillation(from PSY)
		S0016669 Atrial Fibrillations	A0027668 Atrial Fibrillations(from MSH)
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	S0016899 Auricular Fibrillation (preferred)	A0027930 Auricular Fibrillation(from PSY)
		S0016900 (plural variant) Auricular Fibrillations	A0027932 Auricular Fibrillations (from MSH)

Fibrillations. Since Atrial Fibrillation and Auricular Fibrillation have been judged to have the same meaning, they are linked to the same CUI.[5]

Applying to SUI of UMLS metathesaurus

Comparing Western Medicine and traditional Korean medicine for applying UMLS metathesaurus to traditional Korean medical terms, there is not only many consistency, but also differences.

We can see differences in character, language and consistency in linking structure, whether being synonym and opposite or not (Table 2).

Table 2. Compare Western Medicine and traditional Korean medicine in SUI of Metathesaurus

		western medicine	traditional Korean medicine
Differences	Character	singular/plural, genitive, capital/small, spelling variants	initial law, spelling variants
	Language	english	hangul, chinese, modern chinese(Jian-hua)
Consistency	Linking Structure	concept	concept
	Synonym	yes	yes
	Opposite	yes	yes

Traditional Korean medicine is represented by Hangul, Chinese, modern Chinese(Jian-hua), that is different from western medicine in marking for medical terms. We must consider as follow in setting SUI in western medicine.

Strings of Singular or Plural

Atrial Fibrillations, the plural of Atrial Fibrillation, has a different string identifier. Since the singular and plural are lexical variants of each other, both are linked to the same LUI.

Strings with Multiple Meanings

In some cases, the same name (with or without differences in upper-lower case) may

apply to different concepts, usually (but not always) in different Metathesaurus source vocabularies. In the abbreviated example that follows, the string "Cold" is a name for the temperature in one vocabulary. In another vocabulary, "Cold" is an alternate name for the "Common cold". In a third vocabulary, "COLD" is an acronym for "chronic obstructive lung disease". As a result, "Cold" or "COLD" appears as a name of more than one concept in the Metathesaurus.

Traditional Korean medical terms to SUI of UMLS metathesaurus

Traditional Korean medical terms are not discriminated by singular or plural string. Also traditional Korean medical terms has vary string by initial law: the law of initial sound of a syllable.

There is many terms that the meaning is

---

5) Seong-Bin Han, Jin-Wook Choi. The comparative study on concept representation between the UMLS and the clinical terms in Korean Medical Records. IJMI 2005;74:67-76

same each other, but the strings is different in traditional Korean medicine.

In the abbreviated example in Figure 1, '상양' appears as an string in more than one string vocabulary and has a distinct SUI for each occurrence. Since each of these strings have an identical lexical or concept name, they are linked to a single LUI. 商陽, Chinese string of 상양, has a different string identifier. 商陽, Chinese modern(Jian-hua)

string of 상양, has a different string identifier. Since the Korean and Chinese are lexical variants of each other, both are linked to the same LUI. There is a different LUI and different SUIs for 상양 and its Chinese 商陽 and modern Chinese(Jian-hua) 商阳. Since 상양 and 商陽, 商阳 have been judged to have the same meaning, they are linked to the same CUI.
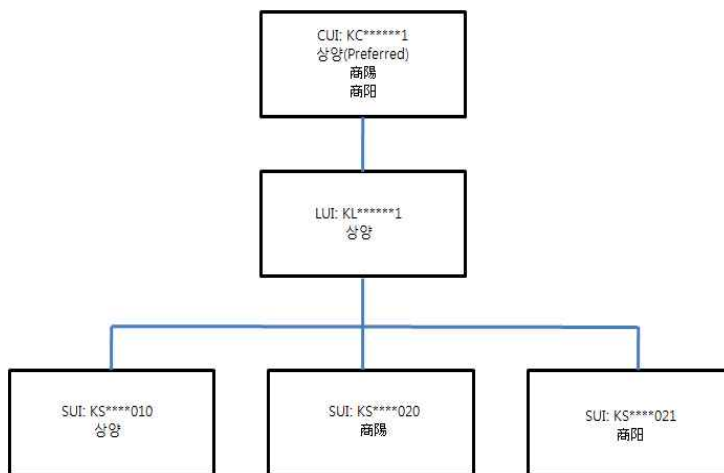

Figure 1. the same name to different names

In some cases, the same name may apply to different concepts, usually (but not always) in different part source vocabularies. In the abbreviated example in Figure 2, that follows, the string "회음" is a name for the acupuncture point name(CV-1, first of controlling vessel) in one vocabulary. In another vocabulary. In another vocabulary, "회음" is a name for the "Perineum". The same string in different meaning (e.g., 회음) will have a different string identifier for each meaning. If the same string, e.g., 회음, has more than one meaning, the string identifier will be linked to more than one concept identifier (CUI).

Coding and Classification for SUI

Coding is the process of assigning an individual object or case to a class, or to a set of classes in the case of a multiaxial classification. In most classification, classes are designated by codes. Coding is, in fact, interpretation of the aspects of an object. Codes may be formed by numbers, alphabetic characters, or, both. The following list describes different types of codes.[6]
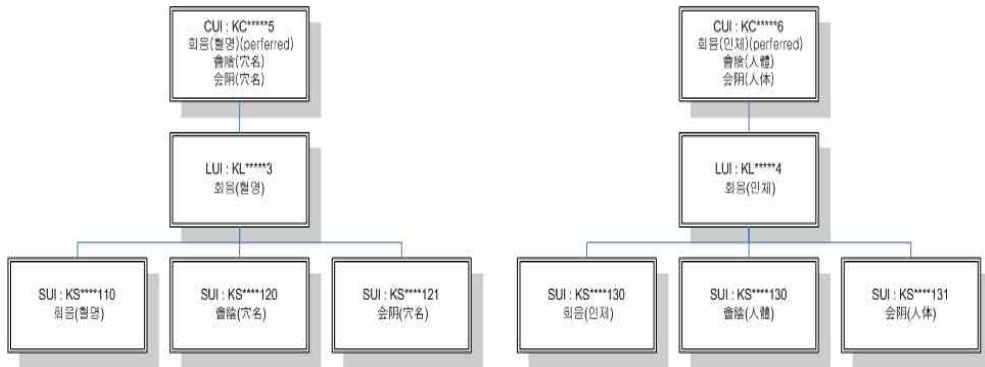
---

6) J.H.van Bemmel. Handbook of Medical Informatics.

Figure 2. the same name to different concepts

Number codes

Number codes may be issued sequentially. This means that each new class will be given the next unused number. The advantage is that new classes can easily be added. Series of numbers can be reserved for sets of classes. Issuing this type of number is only of use with a fixed set of classes, that is, when no expansion of classes is expected.

Mnemonic codes

A mnemonic codes is formed from one or more characters of its related class rubric. this helps users to memorize codes. However, for classifications with many classes this may lead either to long codes or codes with no resemblance to the class rubrics. Therefore, mnemonic codes are generally used for limited lists of class.

Hierarchical codes

Hierarchical codes are formed by extending an existing code with one or more additional characters for each additional level of detail. A hierarchical code thus bears information on the level of detail of the related class and on the hierarchical relation with its parent class.

Juxtaposition codes

Juxtaposition codes are composite codes consisting of segments. Each segment provides a characteristic of the associated class. In ICPC, for instance, a diagnostic code is formed by using a code consisting of one letter of the alphabet (a mnemonic code for the tract), followed by a two-digit number. For instance, all codes with the character "D" are related to the tractus digestivus and all codes starting with an "N"describe disorders of the nervous system. In the example of ICPC, two independent characteristics are coded simultaneously, and each characteristic has its own position in the code.

Combination codes

Another example is a classification of medical procedures using ordering principles : action, equipment, aim, and anatomical site. the combination of 100 anatomical sites with 20 different actions, 10 different instruments, and 5 different purposes results in a classification system with a potential of a 100,000 classes and codes. A way to cope with this explosion is the use of a combination code.

Springer. 1997

Value addition codes

In value addition codes only powers of 2 are used as a representation of a data item or class. just as in a combination code, several characteristics can be coded. In this case, however, only one number instead of a segment for each characteristic is used as a code. this is easily illustrated if we code the presence or absence of risk factors.

SUI Coding for traditional Korean medicine

Table 3. SUI Coding for traditional Korean medicine

segment section	transcription	unit	feature
association segment	"KS"	two digit	string
first numeral segment	"000001"	six digit	odd(n mod 2 = 1) : Korean even(n mod 2 = 0) : Chinese
second numeral segment	"0"	last one digit	the others

Rule for second numerical segment can be provisional, folllow is one of the example for last one digit. (Table 4)

Table 4. Example for second numerical segment in SUI

last one digit	allot
0	hangul, chinese
1	modern chinese(Jian-hua)
2	spelling variants
3	english
4-9	others

There are some examples of SUI in traditional Korean medicine. (Table 5)

Table 5 Example for SUI coding

terms	association segment	first numeral segment	second numeral segment	SUI
상양 商陽 商阳 (shāng yáng)	KS KS KS	000001 000002 000002	0 0 1	KS0000010 KS0000020 KS0000021
요료	KS	000003	0	KS0000030
노료	KS	000003	2	KS0000032

SUI code for traditional Korean medicine is a kind of combination code. That is made up of association session, e.g.,KS(Korean terms string) and numeric session for diversity of representation. Association segment has two characters composed of string with K(initial of Korean) and S(initial of SUI). The next numeric session has seven digits, that is six digits for 999,999 terms and last one digit for the others from now on(Table 3).

Conclusion

This paper is described major problem and applying process when we tried to add traditional Korean medicine in the part of acupuncture into the UMLS metathesaurus. There are many differences to compare western medical terms with traditional Korean medical terms. Traditional Korean medical terms are not discriminated by singular or plural string. In addition, traditional Korean medical terms have vary string by initial law: the law of initial sound

of a syllable. Character is described with Korean, traditional Chinese, modern Chinese, etc. According to meaning, language, initial law, SUI has a distinct value respectively.

For the better implementation to traditional Korean medicine into the UMLS, further research is needed in standardization and classification of traditional Korean medical terms, medical information system, etc. We hope this study helps the implementation UMLS, EHR, knowledge based system in Oriental medicine in the future.

Acknowledgements

References

1. Hye-Sun Kim. A Review of Structure and Application of UMLS. Research of information management. 2001;32(2):26-39

2. Seong-Bin Han, Jin-Wook Choi. The comparative study on concept representation between the UMLS and the clinical terms in Korean Medical Records. IJMI 2005;74:67-76

3. UMLS Knowledge Sources. Bethesda(MD). National Library of Medicine. 2007AA

4. J.H.van Bemmel. Handbook of Medical Informatics. Springer. 1997

5. http://www.opengalen.org

6. Willam Hersh, M.D.et al. Assessing Thesaurus-Babed Query Expansion Using the UMLS Metathesaurus. JAMIA. 2000;7:344-348

7. Nelson, Stuart J., Schopen, Michael, Schulman, Jacque-Lynne, Arluk, Natalie. An Interlingual Database of MeSH Translations. 8th International Conference on Medical Librianship. 2000 Jul 4. London. UK

8. McCray A, Srinivasan S, B rowne A. Lexical methods for managing variation in biomedical terminologies. Proc Annu Symp Comput Appl Med Care. 1994:225-9

9. C.M. Tersa. La Terminology. Amand Colin. Paris. 1998

10. J.J.Cimino. Auditing the Unified Midical Language System with Semantic Methods. J.Am.Med. Inform. Assoc. 1998;5:41-51

11. J.J.Cimino. Linking patient information systems to bibliographic resources. Methods Inf. Med. 1996;35: 122-126