Linking Data for an Information Support System in Traditional Korean Medicine

Hyunchul Jang[1,*], Yong-Taek Oh[1], Sang-Kyun Kim[1], Anna Kim[1], Sang-Jun Ye[1], Chul Kim[1], Mi-Young Song[1]

[1] Korea Institute of Oriental Medicine,
1642 Yuseong-daero, Yuseong-gu, Daejeon, Republic of Korea
{hcjang, ydydxor, skkim, ankim2012, tomita, chulnice, smyoung}@kiom.re.kr

Abstract. Medical information in a clinical setting is primarily centered on patients' symptoms and the methods for treating these symptoms. Traditional medicine in East Asian regions, including South Korea, may also employ a process of pattern identification to determine treatment methods. Such a treatment process leads to the collection of patient symptoms, doctor's diagnostic decisions, and the selection of the method of treatment. Depending on the patient's condition or the doctor's clinical decision, this process may be repeated in full, or part of the process can be omitted. The use of such medical knowledge to build an ontology for each of these processes can facilitate the application of appropriate knowledge according to each process, which allows for a broad variety of medical information to be leveraged using information systems. Various applications are possible by linking data based on classic literature, medicinal materials, formulas, acupuncture, and diseases.

Keywords: Linked data analysis, Ontology, CDSS, Traditional medicine, Traditional Korean medicine

1 Introduction

Throughout history, there have been efforts to observe patients' symptoms and select appropriate methods of treatment to care for patients. These efforts brought us to our present situation via various processes and methods that are unique to various types of medicine and cultures. Symptoms currently present in a patient form the basis of diagnostic data, patient's physical information, and past medical history, which are designed to accurately describe the overall patient condition. In a similar manner, the treatment method is determined at the end of the doctor's actions, such as the diagnosis of the disease to be treated, decisions regarding treatment medications and combinations, and the determination of dosages.

Various processes are also used in traditional Korean medicine (TKM), depending on the specific care provider, although the universal composition of knowledge in TKM, excluding basic theoretical concepts, consists of medications, acupuncture

* Corresponding author: Hyunchul Jang, fax +82-42-861-9421, email hcjang@kiom.re.kr

points, symptoms, and diseases (or patterns) [1]. Medical literature on TKM has focused on specific diseases, formulas, or medicinal materials, and the ontologies of diseases, formulas, medicinal materials, or acupuncture points can be built based on these reports. Nevertheless, it is difficult to declare that a formula for treating a disease, as found in a book that mainly describes diseases in a semantic web environment, is identical to the disease-treating formula obtained from another book that mainly describes formulas; similarly, it is difficult to determine that the two described diseases are actually the same disease.

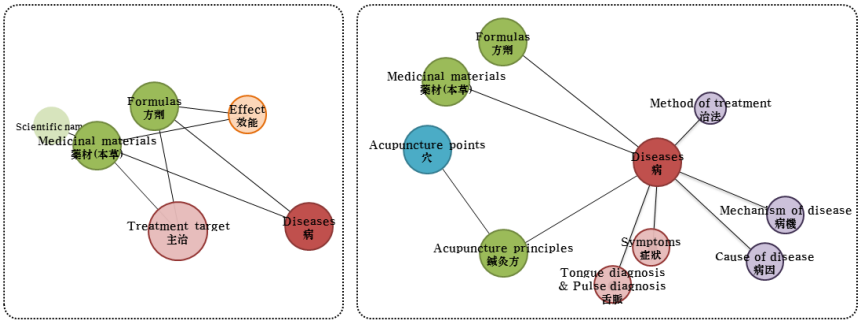
2 Related Work

Content link detection aims to discover similar content across different input and make such links explicit. For example, when reading a news article, content link detection can discover other articles that could serve as background for the current story. A number of machine learning based approaches [2], [3], [4] and [5] showed how machine learning can be used to identify significant terms within unstructured text, and enrich it with links to the appropriate other articles.

But automated approaches cannot be applied in medical domain even though any reduction method like [6] is provided. In medical domain, information systems are demanded content based approach like [7]. Furthermore, contents have to be interpreted semantically.

3 Basic Knowledge

Figure 1 illustrates TKM knowledge, where a node represents class or instance and a link represents a property. Because a pattern in TKM or the traditional Chinese medicine is similar to a disease, it was omitted from the diagrams in Figure 1 to increase visibility. Treatment targets are often represented as a combination of diseases or symptoms.



(a) (b)

Fig. 1. (a) A summary diagram of TKM Medicinal Material Ontology, (b) A summary diagram of TKM Disease Ontology

A disease contains the following information: the symptoms of a disease that can appear in patients, the causes of a disease, the mechanisms of a disease, and the methods of treatment. The medicinal materials, formulas, and acupuncture points (meridian and collateral), for example, contain information on the effects as treatment methods, as well as information on the treatment target or disease.

In addition, methods of treatment are linked to the corresponding proper effects, whereas treatment targets can be linked to the proper diseases, symptoms, diagnoses, or causes of a disease.

Therefore, to link these fragmentary data, we must determine whether two given diseases can be declared to be the same disease or that two formulas are the same, duplicated, or independent of one another. If they are found to encompass the same concepts at a certain level based on the determined results, then the concepts of the two ontologies must be declared to be the same and linked.

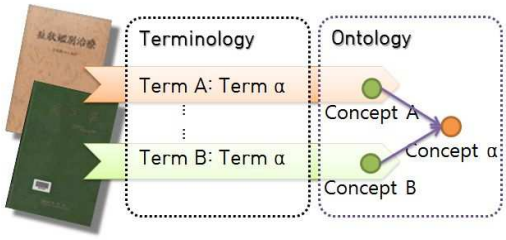


Fig. 2. Linking literature-based knowledge in traditional medicine domain

Such knowledge is described in books on traditional medicine, and depending on the book, it is described with a focus on the diseases, medicinal materials, or formulas. However, the contents of these books, such as textbooks, are organized according to the table of contents and thus cannot actively provide information at an appropriate time, even when information search systems are used. Relevant information is used when the search is conducted at the direction of a Korean medicine doctor, and such information becomes fragmentary information centered on a node.

The abovementioned manner of utilizing knowledge requires a continuous information search process, and many difficulties are encountered when integrating and utilizing a series of concepts obtained in this manner, which makes it unreasonable to use this type of knowledge in clinical practice.

From the perspective of computer engineering, such knowledge can be separated and built into a medicinal material ontology, formula ontology, or disease pattern ontology, as shown in Figure 1, and the sum of knowledge, as shown in Figure 3, can be reproduced by linking the ontologies together.

4 Method

A resource description framework (RDF) [8] was used to build the TKM ontology, and the Jena Ontology application programming interface (API) was used to process the appropriate data. The ontologies of medicinal materials, formulas, or disease patterns described above were linked together by experts in TKM to prevent any

restriction in accessing the linked data or ontologies from the systems perspective. In the real world, each element of the expert knowledge will be published by the experts in each field and could be represented as fused knowledge through a mutual connection.

Figure 3 shows the process of linking concepts, such as medicinal materials, formulas, acupuncture points, and diseases, from each of the ontologies and represents the linking of three ontologies described above into a graph.

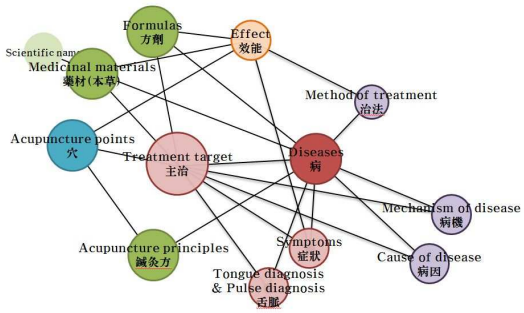


Fig. 3. A graph of the TKM Ontology

The representative formulas and diseases in TKM were linked under the following criteria.

Linking formulas: Because formulas are distinguished by the drug components and the quantities of the drugs used, the formulas are determined to be the same and can be linked if the components or the amounts are inferred to be identical when the drug components and the amounts for the formula used in the disease ontology are compared with the drug components and the amounts built into the formula ontology.

Linking diseases: Depending on the methods of treatment or the specific book, the contents describing the diseases may be comprehensive, a disease may be a subclass of another disease, or the method of treatment may have been described depending on the causes of a disease or displayed symptoms, rather than the names of the diseases. For these reasons, it is difficult to link two concepts as being identical only because they have the same disease names. Of course, it is not impossible to link two concepts at an appropriate level because doctors can determine a more appropriate method of treatment among several that were presented depending on the displayed symptoms, even if the methods were described in the same book. Accordingly, clinical studies, such as the objectification of diagnosis for all diseases, may be necessary, whereas linking concepts at an appropriate level and the use of a simple support system based on this understanding of the concept linking at an information presentation level is thought to be appropriate at this stage.

In the stage of determining the methods of treatment, the formulas linked to the entered symptoms are selected or the formulas linked to the decided diagnosis results are selected. As shown in Figure 3, information regarding the given formulas are initially given for the disease. After the treatment method is selected by the doctor, the effects linked to the treatment method are searched, and additional formulas with corresponding effects can be found.

Moreover, if the major indications of the searched formulas were analyzed and it was found such that there is a match between the information on the linked disease and the disease as a result of a diagnosis, then the selection of the corresponding formulas can be presented as an appropriate choice. The formulas linked to the collected symptoms can be selected from among the possible selections of random formulas.

5 Results

We searched formulas from the TKM disease ontology by a symptom and then searched after linking effect to method of treatment. This result shows more information can be obtained by linking data not merely because an additional ontology was used. By linking medicinal materials used for diseases to formulas, more formulas can be retrieved.

The primary property statistics for finding formulas to care symptoms in the disease ontology are shown in Table 1 and Table 2.

Table 1. Number of formulas to care cough.

Number of having effect to "cough"	Number of diseases having cough	Number of formulas treat diseases having cough
22	56	109
(a)		5 overlapped with (a)

Table 2. Number of methods of treatment for diseases having cough.

Number of methods for 56 diseases having cough	Number of effects EXACTLY corresponding these treatment methods	Number of formulas having these effects
58	12	43
		10 overlapped with above 136

From the TKM disease ontology, 109 referable formulas from diseases having cough were found and 58 methods of treatment were found. It is dependent on how to find effects corresponding to these methods, but 12 effects were linked by name. From the TKM formula ontology, 43 formulas having these effects were found and 8 formulas were overlapped. But, 35 referable formulas could be presented after linking.

6 Conclusion

Knowledge regarding TKM is largely based on traditional medical literature, and such knowledge actually exists independent of TKM. This information exists in the minds of Korean medicine doctors through the incorporation and interpretation of such knowledge, and the interpretation and application of TKM theory is determined by

Korean medicine practitioners. However, knowledge based on traditional medical literature, as well as clinical knowledge, must be accumulated and shared by linking these types of knowledge to achieve the standardization and objectification of traditional medicine.

If the knowledge discussed in Section 4 is divided and expressed in a different space in which the data storage or access methods are varied, and the management of the knowledge is conducted independently, it will be virtually impossible to integrate and manage such knowledge; therefore, it will be extremely difficult to merge and use elements of knowledge that are not linked.

However, if each piece of knowledge is shared or linked using a unique uniform resource identifier (URI) through RDF/OWL [9], this knowledge can be readily accessed on the Web, and each set of data can be shared rather than becoming subordinate to a specific system [10]. Furthermore, knowledge that is made public or shared based on a URI would be reviewed and refined by many individuals and could be realized as user-agreed knowledge.

Acknowledgments. This research was conducted with funding from the Korea Institute of Oriental Medicine (K12090).

References

1. Jang, H., Kim, J., Kim, S.-K., Kim, C., Bae, S.-H., Kim, A., Eum, D.-M., Song, M.-Y.: Ontology for Medicinal Materials Based on Traditional Korean Medicine. Bioinformatics 26(18), 2359—2360 (2010)
2. Milne, D., Witten, I. H.: Learning to Link Wikipedia. In Proc. Of the 17th International Conference on Information and Knowledge Management (2008)
3. Kaptein, R., Serdyukov, P., Kamps, J.: Linking Wikipedia to the Web. In Proc. Of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2010)
4. Nomoto, T.: Two-Tier Similarity Model for Story Link Detection. In Proc. of the 19th International Conference on Information and Knowledge Management (2010)
5. West, R., Precup, D., Pineau, J.: Automatically Suggesting Topics for Augmenting Text Documents. In Proc. of the 19th International Conference on Information and Knowledge Management (2010)
6. West, R., Precup, D., Pineau, J.: Completing Wikipedia's Hyperlink Structure through Dimensionality Reduction. In Proc. Of the 18th International Conference on Information and Knowledge Management (2009)
7. Yi, X., Allan, J.: A Content based Approach for Discovering Missing Anchor Text for Web Search. In Proc. of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2010)
8. Klyne, G., Carroll, J.: Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation (2004)
9. McGuinness, D., Harmelen, F.: OWL Web Ontology Language Overview. W3C Recommendation (2004)
10. Tim, Berners-Lee: Design Issues: Linked Data. http://www.w3.org/DesignIssues/LinkedData.html (2006)