

Leveraging Ontology-Based Traditional Chinese Medicine Knowledge System: Using Formal Concept Analysis

Kwoting Fang¹ Chingwei Chang² Yenping Chi²

¹National Yunlin University of Science & Technology (fangkt@yuntech.edu.tw)

²National Chengchi University (channing@mis.nccu.edu.tw)

Abstract

The main purpose of this paper is to show the use of FCA to support the building of ontology-based Traditional Chinese Medicine with herbal medicine serving as an example. This knowledge-sharing platform may make expert's knowledge clearer, may connect with average people, and may provide patients with an opportunity to participate in and become knowledgeable of the whole prescription for his/her illness.

Keywords: FCA, Herbal Medicine, Information Retrieve.

1. Introduction

Given the striking progress of web service technology associated with the continuing rapid growth in knowledge management, it is imperative that we reduce the knowledge gap which exists between the patient and physician by using the medical knowledge sharing platform and by retrieving self symptoms or medical history, and as well as, through patients' themselves spontaneously learning medical knowledge. This may improve the interactive condition through the process of interrogation enquiry.

How to make information retrieval mechanism-friendly is always an important issue for system building and the creation of a User-Interface field. However, the way we use it to retrieve the information is still rife with problems:

- User has to figure out the exact subject and keywords he/she would like to retrieve.
- The scope of retrieving is limited; the same concepts but expressed by difference semantics are hard to find out with keyword searching.
- The accuracy of the retrieval results are difficult to establish.

How should information retrieval (IR) be set to more closely compare with the users' mindset in order to help the retriever more easily find the knowledge

and to execute the function included in the system. The main objective of this study is to adopt formal concept analysis (FCA) to effectively break the restriction of IR, and to exploit the intention of the document by building a support structure and inter-document relationship automatically.

2.1 Using ontology in knowledge retrieval

Ontology applied to knowledge retrieval definitely changes the strategy from a pull style to a push style [1]. Formerly the retriever inserted a certain keyword for looking up the information in the system, which is a kind of pull style of knowledge retrieval; at present, ontology concepts which provide the vocabulary for expressing personal interest profiles for information push service automatically deliver knowledge and information for the categories a user is interested in.

This study uses Formal Concept Analysis, FCA, as the mathematical principle to bring the benefits of FCA to a menu and folder hierarchy structure under an "user-unconscious" state. Because of the preceding, we can construct the menu settings more logically, and place the relative documents and knowledge so that all the information about Chinese herbal medicine is in the correct logical locations in the system.

2.2 Formal concept analysis

Formal concept analysis was originally developed as a field of applied mathematics based on the mathematization of concept and concept hierarchy, but after more than a decade of development, it has been applied in many quite different realms such as psychology, sociology, anthropology, medicine, biology, linguistics, computer sciences, math etc. From the philosophical point of view, a concept is a unit of thought, and is an idea for reinvigorating connections to our general culture by interpreting the theory as concretely as possible.

With the FCA advantage of the main ideas being quickly understood by formal lattices, linguists often characterize datasets by using distinct features, such as semantic components, phonemes or syntactical; or the terms of schema [3], linguistic applications often involve the identification and analysis of features, which were responded to by the participants, which explains the extraction schema from the interview of adopting new information technology more easily, therefore, the reason this study adopts formal concepts of FAC as an automatic technique to elicit the attributes of dependency schemas extracted from the documents and Chinese herbal medicine references.

3. Implementation the TCHM

In this section, we will introduce the traditional Chinese Herbal medicine Knowledge Management System, TCHM, in more detail, by following three principle steps. While put FAC in practice of IR, usually involves the three steps listed below: (1)extracting a set of index terms that describe each collected document, (2)relating the document-terms generated from the step before with each other by concept lattice (3)visualization of the concept lattice constructed by step (2) to a structure diagram. The outcome of each step may profoundly affect the overall effect of the application [1].

Such as, like how to organize queries about domain knowledge related articles or to understand the principles introducing system retrieval as constructed by the integrated software tool, Protégé. As we mentioned in the previous section, ontology could be seen as defining the set of concepts and the relationships between the concepts; such relationships facilitate knowledge-acquisition tools designed to be domain-specific, which allows domain experts to easily and naturally embed their knowledge of the area . The resulting knowledge base can then be used as a problem-solving method to answer questions and to solve problems regarding the key point, according to the features of this study, Protégé [2]. This application is the end product which is created when the knowledge base is used in solving an end-user problem employing appropriate problem-solving, expert-system, or decision-support methods. Moreover, there are four strong points which can be achieved by Protégé,

- The modeling of an ontology of classes describing a particular subject
- The creation of a knowledge-acquisition tool for collecting knowledge
- The entering of specific instances of data and creation of a knowledge base, and
- The execution of applications. These characteristics being concluded are the reason we adopt Protégé to construct the

framework of Search Data Management subsystem (SDM).

SDM is the core of the TCHM knowledge retrieval platform for the building of TCHM ontology by Protégé for SDM; this design makes SDM have the capability for handling ambiguous information queries from the different users.

The functions of searching for specific news and for searching Chinese herbal medicine knowledge are providing users with a retrieving mechanism. By inserting the keyword that the user is interested in, the system will return the relative Traditional Chinese Medicine news report, article or documents.

3.1 Extracting a set of index terms

There are many techniques that could improve the effect of a retrieval knowledgebase that we have discussed, but the knowledge itself always plays a major role in the knowledge system. It would be useless, if there were nothing in the knowledge sharing platform, in the event of the latest technology being adopted. Emphasizing the source of knowledge itself is always the best policy for information retrieval system building.

With the Protégé application being developed for the creation of the ontology based system, in this step, we centre on creating the slots for each information item such as documents, clinical diagnosis etc., with these slots also known as roles in description logics and relations in FCA construction and other object oriented notions. It also in some other formalism is called 'attributes'.

This study using TCHM related domain knowledge from literature, documents, CD etc., and creating TCHM ontology by the integrated software Protégé, the process of setup interfaces are as the lower diagram represents. For instance, Traditional Chinese Medicine argues that there are three main kinds of medicines with different curative effects: Heat-clearing and Cold-dispelling Prescriptions, Blood-Fluids tonic Prescriptions and Internal Organs Prescriptions. After identifying these, we set them as kind of a super-class and then locate the detail region which can be treated as a sub-class of the super-class. Chinese Medicine says that either excessive heat or cold injure the human body. After Heat-clearing and Cold-dispelling Prescriptions are created for the dispelling and the elimination of the injury, caring for both the internal and external parts of the body become more and more important; therefore, Inside and Outside Prescriptions should be included in the Heat-clearing and Cold-dispelling Prescription class. Hence, with the same logic Dispelling internal Cold Prescriptions, Relieving Exterior Disorder Prescriptions, Clearing Summer-Heat Prescriptions, Relieving Dryness with moistening drugs, and Sedative and tranquilizing

prescriptions should belong to the Heat-clearing and Cold-dispelling Prescription class. As we go down to the bottom level, the scope is narrowed and made more specific, this is the class where we set slots for Chinese Medicine.

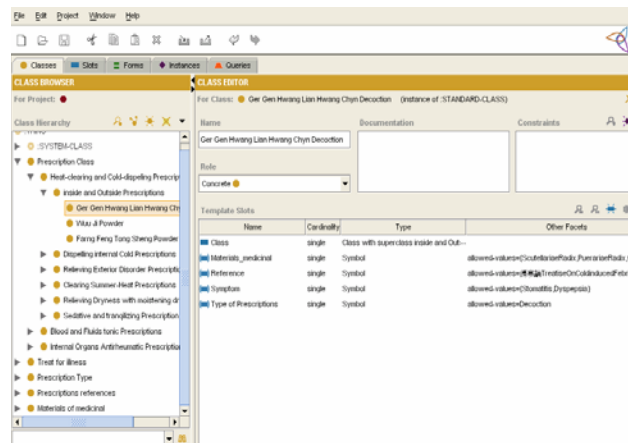


Fig 1: Interface of Protégé constructing classes

Slots are the same as roles in description logics and relations in FCA construction. In some FCA research, attributes are called attributes, which are part of the characteristics of the object; for example, Ger Gen Hwang Lian Hwang Chyn Decoction is described by 5 slots. Different prescriptions may bring out a different nature and may be provided with difference characteristics, such as prescriptions made up of different kinds of medicinal materials. What references and what kinds of symptoms determining the prescriptions could be used for these prescriptions. All of these characteristics were specified by difference allowed-values of slots, as the above Fig.1 mention.

With the slot symptom identifying results caused by prescriptions, mapping for certain disease should be improved or eased; the slot class is to specify the superclass of certain prescriptions locality; the slot type of prescriptions specify the appearance of medicines like powder, decoction, pill and paste.. etc.. All kinds of prescriptions were set up for slots in Protégé for the construction of TCM ontology. After this procedure, prescriptions were described in all aspects by different slots; however, the relationship building and constructing concept lattices must be based on these slots which belonged to each prescription in the next step.

Slots are as roles in description logics but relations in FCA construction, in some FCA research call attributes as slots, which is part of the characteristic of the object; for example, Ger Gen Hwang Lian Hwang Chyn Decoction is described by 5 slots. Different prescriptions may bring out a different nature and are provided with difference characteristics, such as prescriptions made up of what kinds of

medicinal materials, what references bring out the prescriptions, and what symptoms could use this prescription. All of these characteristics were specified by difference allowed-values of slots, as the above Fig.1 mentions.

The slot symptom implies that results caused by prescriptions, and mapping to certain disease should be improved or eased; the slot class is to specify the super-class of certain prescriptions found; the slot type of prescriptions specify the appearance of medicine, like powder, decoction, pill and paste, etc. All kinds of prescriptions were given setup slots in Protégé for constructing TCHM ontology. After this procedure, prescriptions were described in every aspect by different slots; moreover, the relationship building and constructing concept lattices must be based on these slots which belong to each prescriptions in the next step.

3.2 Concept lattice constructing

Ontology describes the concepts in certain domains and also describes the relationships that hold between those concepts. In this section we will map the concepts into the correct class, classes which are interpreted as sets that contain instances with the construction of the linking between the concepts by the related degree of attributes which subordinate each different concept and find out the common part of two classes and then determine how close the relationships are. Furthermore, several similar objects are merged with new concepts to represent them.

As the bellow diagram shows, the TCM ontology which we built is huge and complicated. Therefore, setting the concepts for representing the sub-class is quite necessary. In this process, five concepts gradually emerge to become the main essence of TCHM ontology. They are as follows: Prescription Class, Prescription references, Materials of medicine, Prescription Type, and Treatment for illness.

3.3 Queries Outcome Visualization

The relationship between concepts is constructed at step2 with users allowed to view certain herbal medicine domain knowledge in which they are really interested, with it represented by a nest line diagram, as the following Fig.3 displays. Although this diagram may show only partial knowledge of TCHM, areas selected by the user will provides sufficient knowledge of certain traditional Chinese Herbal medicine.

After constructing the ontology of certain domain knowledge which is set up by the experts and builder, ontology is described by an accredited standard text format, Extensible Markup Language, XML [4], which is generated by Protégé. While we finish the construction of TCHM, ontology is simultaneously and

automatically translating the ontology of TCM domain knowledge to the web page, which represents the returns of user queries more smoothly than through a browser.

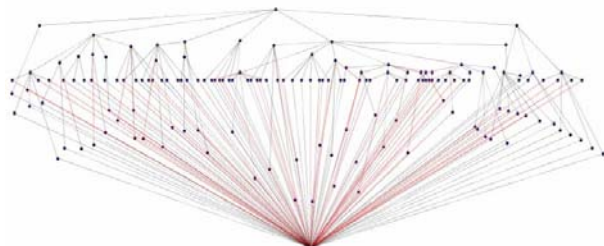


Fig 2: Chinese Herbal Medicine Ontology

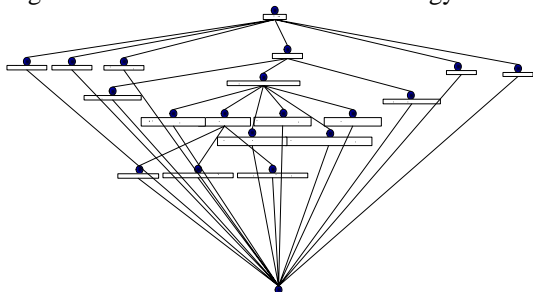


Fig 3. Partial ontology in nested line diagram

Protégé compiles the domain knowledge rule and relationship by XML syntax, which contains the rule of the treatment method, and includes the connecting symptoms, treatments and curative effects with each other. This constructs difference aspects of experts clinical knowledge of how to be a knowledge-triangle, which gives consideration for two or more aspects of health care as the Fig.4 presents in the XML format which is automatically compiled by Protégé integrated software.

The outcome of the query displayed through the website return to users themselves may have advantages such as user-friendly feedback by multiple media, and the generation of dynamic results etc., However, TCHM returns not only relative knowledge of requests, but also shows the whole picture of an FCA nested line diagram, and lets the retriever be aware of the knowledge he/she queried from that location of the whole TCM knowledge ontology.

4. Conclusion

The goal of our developing FCA-constructed retrieve mechanism was to improve the traditional ways of managing information items (such as document, medical science literature) and domain knowledge used in the Chinese Medical clinic, while minimizing the additional effort required from the information retrieve and content management. The above-mentioned also required us to put an emphasis on intelligent query interface—building the hierarchy with flexible information retrieve mechanisms through the ontology concept, which is quite different from

before. It formulates user queries accordingly with ambiguities being avoided in the interpretation of both query and reply. Such normalization and retrieval promise precision in search beyond what is possible with the current keyword-based methods. Not having to be aware of the semantics of terms makes things much easier.

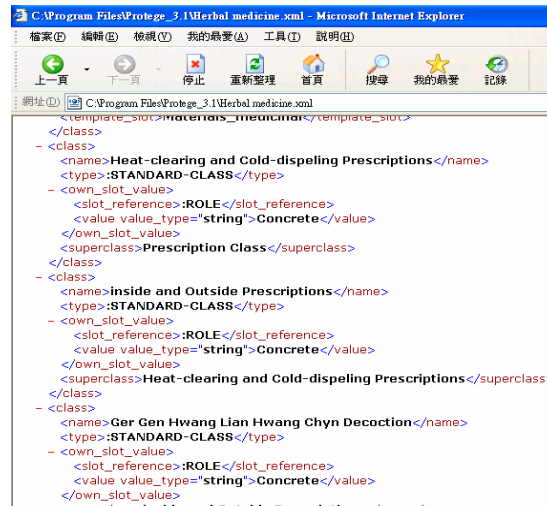


Fig 4. Ontology compile in XML

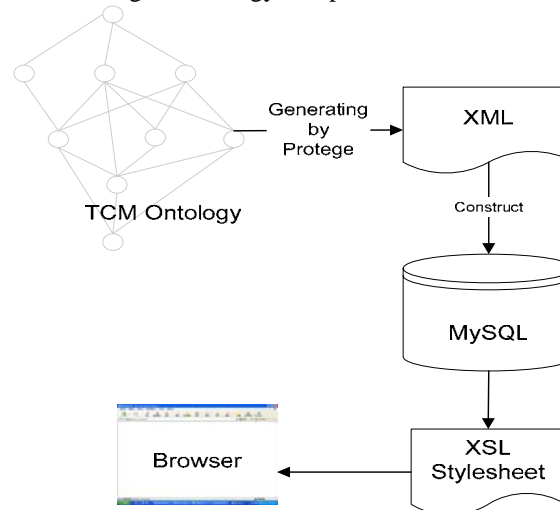


Fig 5. Presentation Construction

● Reference

- [1] Claudio Carpineto and Giovanni Romano, 2005, "Using Concepts Lattices for Text Retrieval and Mining", Formal Concept Analysis Foundations and Applications, Springer Publishing, Inc.
- [2] David Hunter, Andrew Watt etc., 2004, Beginning XML 3rd Edition, Wiley Publishing, Inc.
- [3] Uta Priss, 2005, "Linguistic Applications of Formal Concept Analysis", Formal Concept Analysis Foundations and Applications, Springer Publishing, Inc.