

Ontology Research Based on FCA

Jing Yan, Yantao Wang

College of Economics and Management Northeast Dianli University,
Jilin city, 810008, China

Abstract—With the extending of the scope of information sharing and data exchange, ontology, as the formal specification of share concepts, is more and more useful. FCA had remedied deficiency of existing method of ontology building. In this paper, the existing methods of domain ontology building are presented, then under given data model and data information, demonstrates how to construct the ontology.

Keywords-FCA; ontology concept; lattice.

I. INTRODUCTION

Ontology derived from philosophy, used to describe the essence of things. Since 90's of 20th century, ontology was used in knowledge representation, sharing and reuse. The definition of ontology is given by Gruber Yu in 1993[1], that is, "ontology is an explicit formal specification of conceptual model ". Borst Yu made a point in 1997[2] that "ontology is an explicit formal specification description of the shared conceptual model ". Ontology is used to capture knowledge in related fields, to provide a common understanding of knowledge, to determine the common memory of vocabulary, to supply a clear definition of the interrelationships between these terms from different levels of the formal model. Ontology can be described as a concept hierarchy in which there are relationships between them.

FCA (Formal Concept Analysis) is a data analysis method. The main work in this area is research on domain of knowledge representation, taking the concept lattice as the core data structures, in order to find, sort and display concepts. All concepts together with the generalization and specialization relationships between them form a concept lattice with a strict mathematical nature.

II. METHODOLOGY OF DEVELOPING ONTOLOGY

Ontology is used to do automated reasoning and knowledge representation and reusing. Ontology needs to be built before use. In conjunction with the relevant literature [3], the steps of ontology building can be summarized as following:

- i) Ontology obtainment. The main vocabulary in the text, as well as all the concepts and the relationships between concepts are got, and the concept is clear defined.
- ii) Ontology decoding. Using the hierarchical relationships of concept lattice describes the domain concepts and relationships between concepts, and concept lattice classification property is added to the appropriate

concept. The classification relations between domain concepts are expressed clearly.

iii) Ontology integration. The original ontology can be reused to speed ontology development. Due to the different development methodologies, the common concept is difficult to be found.

Ontology building should been given with application environment and clear development purpose, and developers can use a variety of methods to get the domain ontology and formal their definition. Concept learning method is adopted in loop extraction [4], in which classification relationship has been established. Ontology can be identified through building the mapping relationship between concepts in KACTUS [5] [6]. Although the researchers used different architectural approaches to build ontology, many common problems still cannot be solved, if ontology is built in manual way, when facing complex area, the work will be hard. Experts are needed when we want to build domain ontology, which makes it more subjectivity to build ontology. In another word, different experts built the different structures of domain ontology even in the same area. Each method and principles of ontology building is used in different fields, but the general method of ontology building is lack.

Due to the many problems with manual methods, researchers introduce the knowledge acquisition technology to build domain ontology, and make the related technology as the ontology learning technology. The ontology needed can be obtained automatically or semi-automatically from existing data sources in machine learning and statistics method.

There are two main methods of ontology building semi-automatically:

i) Ontology learning with dictionary. With the help of the dictionary definitions of each word synonym, prototypes and word roots, hierarchical relationships of concepts can be built [7]. The advantage of this method is based on the dictionary, which makes it has a certain commonality. While the disadvantage is that the method didn't associate with a particular area of the body, so experts in the field is needed to participate to complete the ontology building. In addition, because of the size of the dictionary, the method can't be used in a diverse range of fields, and also the method is unable to adapt to climate change.

ii) Ontology learning with knowledge base. After knowledge base has been built, when the uses query and

enter the keyword, a large number of unrelated data can be deleted using the existing rules in knowledge base. After that we can get the related documentation collections.

Comparing with ontology building by manual, ontology learning technology provides a way to suffer tremendous workload. But in the practical course of knowledge acquisition, this kind of learning is difficult to find the knowledge hidden in the brain or document. Formal representation of implied relationship between concepts can't be realized.

III. ONTOLOGY BUILDING BASED ON FORMAL CONCEPT

A. Constructing steps

We use concept lattice for ontology building, aiming not only to find all existing concepts, but also to find the concepts hidden in the human brain or in a document, and to determine the relationships between concepts. The problem, that building ontology by hand would be affected easily by subjective of experts in the field, can be solved. We use Haav method, Cimiano method [8] [9] and so on to build ontology. Although there are various differences in initializing the form background, corresponding relationships, transmitting the concept lattices into ontology and the size of the ontology, the main steps can be summarized as follows:

- i) Plain text document has been pre-processed by using natural language processing (NLP), and we can get the word combination.
- ii) Getting the keyword in the text by using statistical methods. TF-IDF method is usually used in this step. First, we should figure out the frequency of concepts emergence in the text. Second, we will judge the frequency. If it is greater than the threshold value, the concept is a domain ontology concept. Last, we can identify all binary relations table formed by concepts and text.
- iii) Transmitting the multi-value of binary relations table into a single-value form, in order to form a single valued binary relation table to construct a concept lattice.
- iv) Converting concept lattice into corresponding ontology. Attributes can be instead of formal concept, and it is important that its mark is appears only once only. Since property and keyword are all vocabularies, it is feasible that the attribute of formal concept takes place of the concept of ontology. After that we can complete the transformation of formal concepts into ontologies.
- v) Automatic update to the resulting concept lattice.

B. Constructing instance

We can analyse the documents (number 1-8) in semantic gallery by using ICTCLAS system [10]. After analysis, we can get the words and their attributes. We can extract the nouns as project accordance with Chinese grammatical rules. We can filter the highest degree of correlation words (number a-h) by using TF-IDF [11], and find out the high frequency words in each document.

This binary relation table is constituted by the documents and words, in which we can defined formal background of the ontology as L, collection of documents

as object set G, high frequency vocabulary as of attribute set M, and occurrences as rules I, Then the $L := (G, M, I)$, $G = \{1,2,3,4,5,6,7,8\}$, $M = (a, b, c, d, e, f, g, h)$. Because corresponding rule represents times the word appears in the document, the result we got is a multi-valued background. We can transmit the multi value into single value by using yes or no to represent whether the word appear in the document.

TABLE I. FORMAL BACKGROUND BASED ON TEXT RESOURCES

Document	a	b	C	d	e	f	G	H
1	1	1	1		1	1		
2			1	1		1	1	1
3						1	1	1
4							1	
5					1	1		1
6	1	1	1	1				
7		1	1	1				
8				1				

Initial and updating concept lattice can be constructed by the Godin Algorithm [12]. The basic idea of the algorithm is that, when the form background $k := (g, m, I)$, the concept lattice L and new object g^* are given, we can generate a new formal background $k^* = (g \cup \{g^*\}, m, I)$ and the new concept lattice L^* corresponding. Godin algorithm, primarily addresses three issues:

- i) To generate a new node
- ii) Avoid duplication of node generation
- iii) Update

Godin Algorithm:

Input: L, new object x^* , property $f(\{x^*\})$ corresponding with x^*
Output: Lattice L updated
BEGIN
/*To adjust the top node*/
if $\text{sup}(L) = (\emptyset, \emptyset)$ then $(\{x^*\}, f(\{x^*\}))$ instead of $\text{sup}(L)$
else if not $\text{sup}(L) = (\emptyset)$ then $(f(\{x^*\})) \subseteq x'(\text{sup}(L))$
if $x(\text{sup}(L)) = \emptyset$ then
 $x'(\text{sup}(L)) := x'(\text{sup}(L)) \cup f(\{x^*\})$
else adding new nodes H
update
/* Storage in order of the cardinality of the object's attribute*/
 $c[i] \leftarrow \{L : \|x'(H)\| = i\}$
 $c'[i] \leftarrow \emptyset$
for $i : 0 \text{ max}(\text{cardinality})$
for (number of storage in each layer $c[i]$)
if $x'(H) \subseteq f(\{x^*\})$ then
add x^* to $x'(H)$
add H to $c'[i]$

if $x'(H) = f(\{x^*\})$ then algorithm end
else $\text{int} \leftarrow x'(H) \cap f(\{x^*\})$
if (there is no $H \in c'[\text{int}]$ st $x'(H) = \text{int}$) then
 generate a new node H
 generate a new side
 adjust the side

According to the above algorithm concept lattice in Figure 1

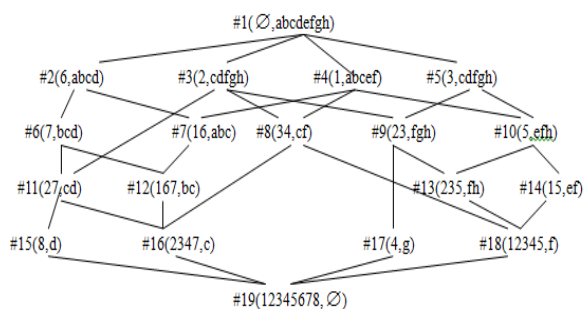


Figure 1. Concept lattice

IV. CONCLUSIONS

This paper is based on the concept lattice theory, and the ontology is transmitted into formal background to build. After introduction of ontology building methods, the author generalized the steps for ontology building, and summed up the problems in ontology building. In order to solve the problem, the author tries to build the ontology by using concept lattice.

In the paper, basic steps of transmitting ontology into formal background have been summed up. A given documents from the database has been extracted to build the ontology and convert it into single-valued formal contexts. Concept lattice existing can be added and updated automatically by using the Godin algorithms, and through this way ontology updating automatically is completed. It can be seen, to introduce the formal concept into ontology

building, can largely solve the problem of ontology automatically updating. Other problems of ontology building will be solved in the next step.

REFERENCES

- [1] Stackelberg H. *The Theory of the Market Economy*. Oxford: Oxford University Press, 1952.
- [1] Gruber T R.A *Translation of Engineering Ontologies for Knowledge Sharing and Reuse*. University of wente, Enschede, 1997.
- [2] W.N.Borst, *Construction of engineering ontologies for knowledge sharing a resue*. Enschede :University of Twente ,1997
- [3] Bi Qiang and Teng Guangqing . *Foreign study on the application of theory of formal concept analysis and concept development and at the forefront of hotspot analysis*. Library and information technology,pp.17-23,2010
- [4] Kieta J U, Maedche A, Volz R. *Extracting a Domain-Specific Ontology from a Corporate Intranet*. In: *Proc.of the Second Learning Language in Logic(LL)Workshop*. The International Conference in Grammar Inference and Conference on Natural Language Learning Lisbon, Portugal, September, pp. 13-14, 2000
- [5] Bernaras A, Laresgoiti I, Coreral J. *Building and Reusing Ontologies for Electrical Network Applications*. In: *Proc.of the European Conf.on Artificial Intelligence*, Budapest, Hungary ,pp. 298-302, 1996
- [6] Martil R, Turner T, Terpstra P. *Knowledge Reuse in Technical Domains: The KACTUS Project*. In: *Proc.of the Impact of Ontologies on Reuse, Interoperability and Distributed Processing*, Unicom Seminar, London, 1995
- [7] Tan K-W, Han H, Elmasri R. *Web Data Cleansing and Preparation for Ontology Extraction using WordNet*. In: *1th Intl.Conf.on Web Information Systems Engineering (WISE'00)*, 2000
- [8] Cimiano P, Staab S, Tane J. *Automatic Acquisition of Taxonomies from Text: FCA meets NLP*. In: *Proc.of the Intl.Workshop on Adaptive Text Extraction and Mining*, pp.10-17, 2003
- [9] Haav H M.A *semi-automatic method to ontology design by using FCA*. Proceedings of the 2nd International CLA Workshop, TU of Ostrava
- [10] <http://ictclas.org/>
- [11] Salton, G. and Buckley, C. *1988 Term-weighting approaches in automatic text retrieval*. Information Processing & Management, pp.513-523, 2005
- [12] Godin. *Incremental concept formation algorithms based on Galois(concept) lattices appeared in computational Intelligence*. Library and information technology, pp.246-267, 1995