Indian Journal of Science and Technology, Vol 8(24), DOI: 10.17485/ijst/2015/v8i24/82808, September 2015
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Survey of State of the Art of Ontology Construction and Merging using Formal Concept Analysis

M. Priya* **and Ch. Aswani Kumar**

School of Information Technology and Engineering, VIT University, Vellore-632 014, Tamil Nadu, India;
sumipriya@gmail.com, cherukuri@acm.org

Abstract

The goal of Formal Concept Analysis (FCA) and Ontologies are modeling concepts with varying needs. An ontology is an explicit formal conceptualization of some domain of interest. Ontologies are widely used in various fields such as E-commerce, Semantic Web and knowledge management. FCA is a method of deriving a formal ontology or a concept hierarchy from a group of objects with their properties. FCA facilities an environment to make the data simpler by analyzing, structuring and visualizing. This paper presents a survey of Ontology construction and merging using FCA.

Keywords: FCA, Formal Context, Ontology, Ontology Construction, Ontology Merging

1. Introduction

The goal of Formal Concept Analysis (FCA) and Ontologies are modeling concepts with varying needs. The goal of FCA is to provide a platform to the user in analyzing and structuring a domain of interest[1-3]. On the other hand, the purpose of the Ontology is to model a "shared understanding of the domain of interest", where "shared" means an Ontology captures consensual knowledge should be accepted by a panel of experts in the given domain[3,4]. FCA can be used to construct the new Ontology as well as to merge the existing Ontologies[5-18].

In Ontology construction using FCA, initially the formal context of the given domain can be derived. Then from the derived concepts, concept lattices can be generated. Finally, domain Ontology can be generated from the concept lattice. In Ontology merging using FCA, two different Ontologies (O1 & O2) can be taken as input and the two formal contexts K1 & K2 can be identified respectively. Then, using any of the FCA merging algorithms, the common formal contexts and concept lattices can be derived. From the derived concept lattice, the new (merged) Ontology will be constructed.

The motive of the paper is to provide a survey on Ontology construction methods and Ontology merging methods using FCA. The content of the paper is organized as follows: Section 2 introduces the basic concepts of Ontology and FCA, Section 3 presents the Ontology construction methods using FCA, Section 4 describes the Ontology Merging process using FCA and Section 5 concludes the survey.

2. Ontology and Formal Concept Analysis

2.1 Ontology

In [Gruber 1993], Ontology was defined as "an explicit specification of a conceptualization". The difference between Ontology and conceptualization is that Ontology is language-dependent while conceptualization is language-independent[5].

The main components of the Ontologies are given below:

- Individuals: Individuals are the ground level objects.
- Classes: Classes can be set, collections, types of objects or domain concepts.
- Attributes: Attributes can define the properties, features or characteristics of the objects.

*Author for correspondence

- Relations: Relations shows the way how classes and individuals can be related to one another.
- Functions: Function is a special case of relations in which the n^{th} element of the relationship is unique for the n-1 preceding element
- Restrictions:Restriction is formally stated descriptions of what must be true in order for some assertion to be accepted as input.
- Rules: The rules are the statements which are in the form of if-then sentence.
- Axioms: Axioms are the model sentences that are always true.
- Events: Events arethe changing of relations or attributes.

2.2 Formal Concept Analysis

In 1982 [Rudolf Wille], FCA was introduced. FCA is a mathematical theory which deals with concepts and concept hierarchies[19-23]. FCA analyzes the data to show the association between a set of objects and a set of attributes.

- Formal Context

A formal context K: = (*O*, *A*, *R*) where *O* is a set of objects, *A* is a set of attributes (properties), and *R* is an incidence which shows the association between *O* and *R*. o*Ra*is a binary association where (*o*,*a*)R, then the "object *o* has attribute *a*" or "the attribute a applies to the object o". A formal context K is denoted as a cross table where the rows denote *O*, the column denotes *A* and the incidence relation *R* is denoted by a series of crosses[1-3].

Given two sets *E, I*, such that $E \subseteq O$ and $I \subseteq A$, consider the dual sets *E'* and *I'* i.e., the sets defined by the attributes applying to all the objects belonging to *E* and the objects having all the attributes belonging to I, respectively[1-3],

$$E' : \{a \in A \,|\, (o,a) \in R \text{ } for\ all\ o \in E\}$$

$$I' := \{o \in 0 \,|\, (a,o) \in I \text{ } for\ all\ a \in I\}$$

Example: Consider the Animals context, where
O = {Wolf, Peacock, Hen, Honey Bee, Cat}
A = {Bird, Mammal, Preying, Flying}

In Table 1, rows denotes the set of objects and columns denotes the set of attributes. 'X' denotes the intersection of an object and the attribute which indicates that the object possesses that attribute.

- Concept Lattice

A Concept Lattice is defined as the collection of all formal concepts of a given formal context. Given a context *(O, A, R*), consider the set of all concepts in this context, indicated as £ *(O, A, R)*. Then

$$(£\ (O, A, R), \leq)$$

is a complete lattice called Concept Lattice (Galois Graph), i.e., for each subset of concepts, the greatest lower bound and the least upper bound exist[2]. The animals context concept lattice is shown in Figure 1.

3. Ontology Construction using Formal Concept Analysis

FCA is a mathematical theory of data analysis. It discovers the conceptual structures between data sets. Following this way making the Ontology construction very effective. It also provides an environment to the user to identifies the necessity of new concepts and relations in an Ontology[23].

Table 1. Formal Context of the Animals

Animals	Mammal	Bird	Preying	Flying
Wolf	X		X	
Peacock		X		X
Hen		X		
Honey Bee		X	X	X
Cat	X			

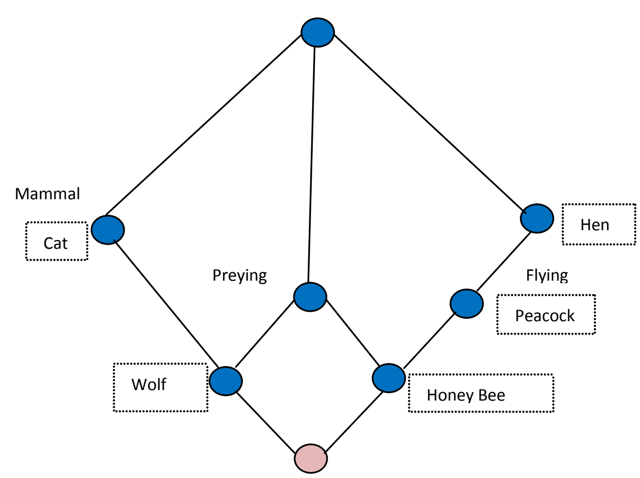

Figure 1. Concept Lattice of Animals context.

In this section some of the formal concept analysis based Ontology construction methods are presented.

Guoqian Jiang et al.[7] proposed a method for constructing clinical domain Ontology based on FCA with a Natural Language Processing (NLP) module. The system user interface was developed in Protégé-2000. The main knowledge source for the clinical Ontology is the set of 368 textual patient discharge reports from cardiovascular and a Japaneese standard dictionary (MEDIS version 2.0). NLP module having three stages namely, Diagnostic Term Dictionary, Morphological Analysis System and Morphological Analyzer Connectivity Driver Model. The Diagnostic Term Dictionary (MEDIS version 2.0) extracts the medical terms from the text document summaries, then Morphological Analysis System (ChaSen version 2.1) used for adding user-defined dictionary and finally Morphological Analyzer Connectivity Driver Model was used to connect the ChaSen system with the Protege´2000. The FCA module takes the terms (input) from the medical dictionary to identify the set of objects in the clinical domain and their attributes, then the set of formal context was identified, then finally clinical Ontology have been generated from the set of formal context (concept lattice). This method has following merits: the Protégé plug-in can automatically extract the formal concept from the given domain concepts and relationships and achieves the semi-automatic Ontology construction by combining with the involvement of domain experts; required concepts can be incurred and the redundant taxonomic structures and concepts can be removed; Ontology construction depends on feedback loop which improves the overall construction process. However there are some demerits: this method is not considered multi-valued attributes hence it is not useful for dealing multi-valued contexts; The Protégé plug-in is used as a medium in the task of discoursing concepts to the formal context, which increases the difficulty. Few associations in the initial prototype of Ontology cannot be mappped into respective formal contexts.

Hele-Mai Haav[8] presented a new method which combines rule-based language with FCA to construct a semi-automatic domain Ontology. In this approach, initially formal contexts for the given domain can be extracted from the input domain data based on the natural language processing techniques. Using FCA and reduction procedures, the initial Ontology can be constructed as a concept lattice from the formal context. Then the initial Ontology can be represented as a set of rules in first order logic and visualized to the Ontology designer. Then the designer can further extend the Ontology by adding concepts and relationships (related to, part-of, etc.) by using a rule language based on Horn clauses. This method has the following merits: this method completely deals with the Ontology reasoning and the Non-taxonomic relations of a domain Ontology; an Ontology can be represented as a first-order logic which validates the Ontology. However, there are some demerits: the process of converting the initial Ontology to the first-order predicate logic requires rule language mapping and FCA which is tedious and difficult to accomplish; the conceptual extension is the number of domain texts, which consequences in the lexical gap of the extension representation of the domain.

Marek Obitko et al.[9], proposed a new method for designing an Ontology using FCA. This method has the following properties:

- The concepts are discovered by properties.
- The properties specify the hierarchy of concepts.
- When the properties of different concepts are the same, then the concepts are the same.

Marek Obitko et al. algorithm performs the following steps to generate the Ontology:

- Initially the Ontology design starts with an empty set (no concepts and properties).
- The designer can append new concepts and properties one by one then identifies the formal contexts of the concepts and generates the concept lattice.
- The generated concept lattice is visualized using FCA with their properties.
- Based on the concept lattice visualization, an Ontology designer can either perform direct editing or Ontology design tool editing :
- In "Direct" editing the designer can add or drop the concept, add or drop the property and specify a property to concept or drop a property from the concept.
- In Ontology design tool editing, while visualizing the concept lattice, if two concepts fall into same place either these two concepts should be merged together or differentiation should be shown between the concepts. The FCA can produce concepts that are formed by properties and are super-concepts of defined concepts, but are not explicitly mentioned in the concept table.

These four steps are repeated until Ontology designer is satisfied with the generated Ontology.

This method has following merits: creates a distributed Ontology environment; this method accomplishes the visualization of concept lattices. However, there are some demerits: the extraction of formal context is completely manual process. Hence this method is unsuitable for larger domain Ontology construction.

Xin Peng et al.[10] have proposed an incremental FCA method for constructing Ontology for semantics–based component retrieval. In the incremental FCA method, the component providers are involved in constructing the collaborative Ontology. The component providers provide the various I/O semantics which includes the business objects and their features while submitting a component to the repository. Then incremental FCA is constructed on the new object. Then based on the FCA, either a new action concept can be created or no action concepts can be created.

In new action concepts creation the component enforces a new business function. An Action directly respective to the component and many suitable ancestor concepts may be generated. Some of the newly created ancestor concepts may not be significant in the domain. Hence the created concepts are given to the component providers for evaluation. Based on the evaluation, the insignificant concepts can be dropped. In No Action concepts creation no changes in the business function. The component has the same I/O semantics with an existing Action concept. Now the submitted component can have the newly created action or existing action based on the I/O semantics. Finally, the system will communicate with the component providers to capture the property specification for the component. This method has following merits: Ontology Construction is based on the contribution of many component providers and the conceptual structure is automatically constructed, which facilitate the efforts of Ontology construction remarkably and ensures the quality of the constructed Ontology.

Suqin Tang et al.[11] have proposed a new method called Tourism Ontology Construction Method (TOCM) using FCA. TOCM comprises tourism information pre-processing module, FCA module and Ontology construction module. In Tourism Information Pre-processing module, the first step is to collect the tourism information from the relevant websites using tools like reptile software and then the collected information needs to be processed for obtaining relevant texts and removing irrelevant information and finally acquire the valid unstructured tourism information. The second step is processing the preserved terms. Finally the preserved terms can be saved into the database table. In FCA Module, the set of formal contexts for the tourist information can be derived from the preserved terms and concept lattice can be generated. In Ontology Generation module, tourism information Ontology can be constructed from the concept lattice. Then the values of slots in each class in the Ontology can be filled either manually or automatically. This method has the following merits: the tourism Ontology construction method combines the context knowledge information with the linguistic knowledge information using FCA, which provides a mass of information in both domains to the knowledge engineers.

Liu Ning et al.[12] have proposed a new method to create a maritime Ontology using FCA. The Maritime Ontology construction process performs four steps. The first step is computing initial Ontology from the thesaurus (Chinese) with the help of maritime domain experts. The second step is performing maritime Ontology creation based on FCA. In this step initially text processing can be done using NLP technology to identify the objects and attributes in the domain and then based on the objects and attributes formal context can be generated and then from the formal context concept lattice can be constructed and finally the conceptual hierarchy can be generated. The third step is mapping new concepts between initial Ontology and newly generated Ontology. Finally the Ontology can be described formally using Protege. This method has following merits: the maritime Ontology construction is purely depends only on objects and their attributes, which permits to attain a new objects and attributes; degree of automation of Ontology building is improved.

Chien Duy et al.[13] have presented an improved formal concept analysis algorithm to construct the domain ontology. The improved algorithm, generates an ontology using Threshold value (T) along with the Information Gain (IG) and Entropy (E). The input for this algorithms are categories and objects and attributes from the domain of computer science and engineering. The categories was taken from ACM (Association for Computing Machinery). From Wikipedia and other corpora they have taken the objects and attributes.

Then the Entropy (E) and Information Gain (IG) of attributes are calculated as follows:

$$IG(a) = E(B-a) - E(a)$$

$$E(a) = -\sum_{j=0}^{C-1} (\text{Pj} \log 2 \text{ P j})$$

$$IG(a \mid Ci) = E(X \mid Ci) - E(a)$$

Where

- B is the set of attributes.
- C is the set of categories.
- E(a) is the entropy of attributeof 'a' in B.
- E(B-a) is the entropy of attributeof 'a' in B after deleted from B..
- Pj is the propability of distribution of attribute 'a' in B.
- IG(a|Ci) is the Information Gain of 'a' in category Ci.
- E(X|Ci) is the entropy of all attributes in category Ci after'a' is removed from Ci.

After calculating Entropy (E) and Information Gain(IG) they have taken Threshold T for constructing the domain ontology. Threshold T is some floating point value based on experience.

The Information Gain ($IG(a_i)$) of each attributes if compared with Threshold T. If the $IG(a_i)$ is greater than T then the attribute will be added to the domain ontology in corresponding category C_i. Else if $IG(a_i)$ is less than 0 and greater than T then the attribute will not be added into domain ontology. It will be added in the new category. This process will continue for the entire attributes. At the end of the algorithm the domain Ontology is connstructed using the relavent attributes from the category C_i.

4. Ontology Merging using Formal Concept Analysis

Ontology merging method is a process of generating a unique Ontology by merging the original Ontologies. The Ontology mapping is involved in the merging process to establish the links between the Ontologies. In this section, we have presented how the Ontologies can be merged using FCA.

Gerd Stumme et al.[14] have proposed the bottom-up merging method to merge the given two Ontologies from the same domain based on the FCA. This merging method is based on the application-specific instances of the input Ontologies *O1* and *O2*. Bottom-up merging method performs the following three steps to merge the Ontologies. The first step is extracting the instance from the input Ontologies *O1* and *O2* based on natural language processing and computes the two formal contexts *K1* and *K2* respectively. Then FCA-MERGE algorithm takes the two formal contexts *K1* and *K2* as input and generates the common context K between *K1* and *K2* and constructs the pruned concept lattice for the context *K*. Finally, the merged Ontology can be constructed from the pruned concept lattice.

Bernard Ganter and Stumme[15] presented a new method called OntEx (Ontology Exploration) which performs both Ontology creation and merging based on knowledge acquisition techniques known as attribute exploration. OntEx ensures that the knowledge engineer deals all relevant possibilities both for the creation and merging of Ontologies. In Ontology creation, the first step is initialization of the exploration contexts based on the users provided concepts in the domain. The exploration process accomplished by providing the exploration dialogue with the user comprising of questions. At the end of the exploration process, the concept lattice of all conjunctions of the input concept is constructed. Finally, the user can make the needed changes in the hierarchy using any of the Ontology editors. The Ontology merging process involves the same steps of the Ontology construction process, the only difference is instead of creating Ontologies from scratch they have taken two Ontologies as input and generates the new merged Ontology using attribute exploration techniques. This method has following merits: The OntEx method provides high accuracy and is suitable for small parts of the Ontologies. However there are some demerits: OntEx methods should be integrated with a heuristic method and needs the interaction of knowledge engineers. When high interaction is required from the knowledge engineers, the cost will be very high.

Li Guan-Yu et al.[16] presented the FCA-Ont Merge method for merging the two Ontologies based on FCA. FCA-Ont Merge method performs four steps to merge the Ontologies. The first step is unifying the format of input Ontologies and describing the Ontologies in OWL format using Protege. In the second step, the input Ontologies can be parsed by the Jena i.e. ontology analysis tool to extract the formal context based on the concepts and attributes. In the third step, attribute mapping table can be generated by matching the attributes in the formal contexts.

The attribute matching process is done by comparing the two attributes. The attributes can be antonym, synonym, hypernymy or hyponymy. If the attributes are antonym then the process will be stopped. If the attributes are synonym the matching process will be stopped and the attributes will be added in the mapping table. If the attributes are either hypernymy or hyponymy the matching process will be stopped and the attributes will be added in the mapping table. The last step in this method is merging the two formal contexts, then constructing the concept lattice for the merged context and finally generates a new Ontology.

Olivier Cure[17] presented a new algorithm for merging the expressive spatial ontologies using FCA. This algorithm is very much helpful for spatial data, in specific the nature of land parcels characterized by a geographical ontology. The input for the algorithm is two spatial ontologies and the output is merged spatial ontology. The steps for merging the two spatial ontologies are as follows:

- Facilitating the construction of concepts not originally from the source ontologies.
- Once concepts has been enabled then the definition to these concepts should be specified with respect to the elements of the source ontologies.
- Finally dealing the construction of merged ontologies based on the uncertainties encountered at the object and alignment levels.

Rung-Ching Chen et al.[18] presented an innovative Ontology merging method which is purely depends on WordNet and Fuzzy FCA (FFCA-Merge). In FFCA-Merge method, the fuzzy ontology can be created based on two extent Ontologies with the similar domain. The newly generated fuzzy Ontology is a unique rational Ontology with high standards. This ontology is unrestricted than a general Ontology.

5. Conclusion

Ontology and Formal Concept Analysis play an important role in the knowledge representation. This paper describes the interplay of Ontology and Formal Concept Analysis in detail manner. In this survey we have provided the summary of different Ontology construction methods and merging methods using Formal Concept Analysis. This survey can provide a better understanding of Ontology and Formal Concept Analysis.

6. References

1. Ganter B,Wille R. Formal Concept Analysis: Mathematical Foundations. International Conference on Conceptual Structures; 1999. p. 377–88.
2. Wille R. Restructuring lattice theory: An approach based on hierarchies of concepts. In: l. Rival (Ed.). NATO Advanced Study Institutes Series. 1982; 83:445–70.
3. Formica A. Ontology-based concept similarity in Formal Concept Analysis. Information Sciences. 2006 Sep; 176(18):2624–41.
4. Uschold M, Gruninger M. Ontologies: principles, methods and applications. The Knowledge Engineering Review. 1996 Jun; 11(2):93–136.
5. Li S, Lu Q, Li W. Experiments of Ontology construction with formal concept analysis. International Joint Conference on Natural Language Processing; 2005. p.67–75.
6. Wolff KE. A First Course in Formal Concept Analysis, Advances in Statistical Software 4. The 7th Conference on the Scientific Use of Statistical Software; 1993. p. 429–38.
7. Jiang G, Ogasawara K, Endoh A, Sakurai T. Context-based Ontology building support in clinical domains using formal concept analysis. International Journal of Medical Informatics. 2003 Aug; 71(1):71–81.
8. Haav, A semi-automatic method to Ontology design by using FCA. In: V. Snasel, R. Belohlavek (Eds.). Concept Lattices and their Applications. International workshop on Concept Lattices and their Applications; 2004. p. 13–24.
9. Obitko M, Snasel V, Smid J. Ontology Design with Formal Concept Analysis. Concept Lattices and theirApplications; 2004. p. 111–9.
10. Peng X, Zhao W. An incremental and FCA-based ontology construction method for semantics-based component retrieval. Seventh International Conference on Quality Software; 2007 Oct 11-12 Portland OR. p. 309–15.
11. Tang S, Cai Z. Using the Format Concept Analysis to construct the tourism information ontology. Seventh International Conference on Fuzzy Systems and Knowledge Discovery; 2010 Aug Yantai Shandong; 6. p. 2941–4.
12. Ning L, Guanyu L, Li S. Using Formal Concept Analysis for maritime ontology building. International Forum on Information Technology and Applications. 2010 Jul 16-18 Kunming; 2:159–62.
13. Ta CDC, Thi TP. Improving the Formal Concept Analysis algorithm to construct domain ontology. Fouth International Conference on Knowledge and Systems Enginering; 2012 Aug 17-19 Danang. p.74–8.
14. Stumme G, Maedche A. FCA-MERGE: Bottom-Up Merging of Ontologies. International Joint Conference on Artificial Intelligence. 2001; 1:p 225–30.

15. Ganter B, Stumme G. Creation and merging of Ontology top-levels. International conference on Conceptual structures; 2003 Jul; 2746. p. 131–45.
16. Guan-Yu L, Shu-Peng L, Yan Z. Formal Concept Analysis based Ontology Merging Method. International Conference on Computer Science and Information Technology. 2010 Jul 9-11 Chengdu; 8: p. 279–82.
17. Cure O. Merging expressive spatial ontologies using Formal Concept analysis with uncertainty considerations. Methods for Handling Imperfect Spatial Information. Spinger-Verlag. 2010; 256:188–209.
18. Chen RC, Bau CT, Yeh VJ. Merging domain Ontologies based on the WordNet system and Fuzzy Formal Concept Analysis techniques. Applied Soft Computing. 2011 Mar; 11(2):1908–23.
19. Kumar CA, Srinivas S. Concept lattice reduction using fuzzy *k*-means clustering. Expert Systems with Applications. 2010 Mar; 37(3):2696–704.
20. Kumar CA. Fuzzy clustering based formal concept analysis for association rule mining. Applied Artificial Intelligence. 2012 Feb; 26(3):274–301.
21. Kumar CA. Knowledge discovery in data using formal concept analysis and random projections. International Conference on Information Processing. 2011 Dec; 21(4):745–56.
22. Kumar CA. Mining association rules using non-negative matrix factorization and formal concept analysis. International Conference on Information Processing. 2011 Aug; 157(1):31–9.
23. Cimiano P, Hotho A, Stumme G, Tane J. Conceptual knowledge processing with Formal Concept Analysis and Ontologies. International Conference on Formal concept Analysis. 2004; 2961:189–207.