Building Domain Ontology based on Formal Concept Analysis

Khin Phyu Phyu Shein, Thi Thi Soe Nyunt
University of Computer Studies, Yangon
Khinphyuphyushain@gmail.com, ttsoenyunt@gmail.com

Abstract

There has been a growing concern on ontology due to its ability to explicitly describe data semantics in a common way, independently of data source characteristics, providing a schema that allows data interchanging among heterogeneous information systems and users. Ontologies are currently being used for explicit representation of the domain knowledge. Ontology provides knowledge about specific domain that is understandable by both developers and computer systems and necessary for knowledge representation and knowledge exchange. Several works have been aimed to improve ontology technological aspects, like representation languages and inference mechanisms, and development method application Using existing taxonomical hierarchies are not enough for knowledge exchange or for informational retrieval. These may cause the problems in knowledge sharing. Therefore we need a better way to describe the concepts and relation for discovering new concepts and relations in ontology. This paper presents an approach of developing ontology for the movie review domain with Formal Concept Analysis (FCA) design.

1. Introduction

One of the main goals of intelligent information processing is domain modeling and knowledge creation for machine process-able knowledge-aware applications. Ontologies are developed and used because they enable among other:

- to share knowledge - by sharing the understanding of the structure of information shared among software agents and people,
- to reuse knowledge – ontology can be reused for other systems operating in a similar domain,
- to make assumptions about explicit – for easier communication.

Domain modeling is the process by which concepts and relations are extracted from the data which can be free-text, semi-structured or data-schema. To enable the desire knowledge, it is necessary to describe the concepts and relations in a better way than just ordering them in taxonomies. Taxonomies are important, since they form backbone of an ontology, but are not enough for knowledge sharing.

We present an approach for designing ontology that is based on formal concept analysis. Formal concept analysis is a theory of data analysis which identifies conceptual structures among data sets [1] [2]. This ontology design method allows for discovering new concepts and relations in an ontology, which leads to an ontology that is suitable for knowledge exchange or for information retrieval.

The rest of this paper is organized as follows. Section 2 describes some related work. Section 3 states the motivation of the proposed approach. Section 4 introduced the proposed approach. In section 5 related query languages is introduced. Finally, the conclusion and future work are presented in section 6.

2. Related Works

Ontology building is the process by which concepts and relations are extracted from the data of free text or data schema. In general, it is unrealistic to use general purpose ontologies in a specific domain. Krovetz and Croft [3] points out that 40% of the words in canonical form in the titles and abstracts of the Communications of the ACM are not included in the LDOCE (Longman Dictionary of Contemporary English). Some researchers have looked for possible ways to build domain ontology by using a number of machine learning approaches applied to ontology building, such as, text structure based learning, syntactic pattern based learning and conceptual clustering based learning [4][5][6] etc. In [7] Philip compare FCA method by showing very good in comparison with the other methods. Another advantages of FCA is that final outcomes can be seen by concept lattice rather than tree like forms produced in other methods. Therefore we choose FCA design to build our domain ontology of movie reviews.

3. Motivation

Using existing taxonomical hierarchies are not enough for knowledge exchange or for informational retrieval. By using the taxonomical ordering, the concepts have no other differentiating attributes. It is not easy to change the frames and their slots once they are defined. Taxonomies are important but are not enough for knowledge sharing. These may cause the problems in knowledge sharing. Therefore we need a better way to describe the concepts and relation. Therefore we proposed a method of FCA based ontology design for creating a new movie review domain.

4. Formal Concept Analysis (FCA)

Formal concept analysis is a principled way of automatically deriving ontology from a collection of objects and their properties builds on applied lattice and order theory. Formal concept analysis refers to both an unsupervised machine learning technique and a method of data analysis.

FCA can be used to analyze the structural properties of conceptual domains that give rise to prototypical categories. The approach takes as input a matrix specifying a set of objects and the properties are called attribute. This duality can often be observed between two types of items that relate to each other in a certain application, such as objects relate to attributes and documents relate to terms. In FCA, the elements of one type are called formal objects; the elements of the other type are called formal attributes. A binary relation exists between the formal objects and the formal attributes. The set of formal objects, formal attributes and the binary relation between the two forms a formal context, which is usually represented as a cross table as shown in table 1. The set of formal concepts from a formal context can be visualized by a concept lattice. A concept lattice consists of the set of concepts of a formal context and the sub concept/super concept relation between the concepts as shown in figure 2.

5. Proposed Ontology Design

The method presented is based on Formal Concept Analysis (FCA), used for analyzing data and forming semantic structures that are formal abstraction of concepts of human thoughts and identify conceptual structures among data sets. It also allows the analysis of complex structures and the discovery of dependencies within the data. In FCA, the elements of one type are called "formal objects", the elements of the other type are called "formal attributes". The adjective "formal" is used to emphasize that these are

formal notions. "Formal objects" need not be "objects" in any kind of common sense meaning of "object". But the use of "objects" and "attributes" is indicative because in many applications it may be useful to choose object-like items as formal concepts and to choose their features or characteristics as formal attributes.

The main characteristics of FCA
- concepts are described by properties
- the properties determine the hierarchy of the concepts
- when the properties of different concepts are the same, then the concepts are the same

Contexts in FCA are triples (O,A,R)

Where　　　　　O=finite set of object

　A=finite set of attributes

R=binary relation on O and A

The procedure of designing ontology supported by a tool that use FCA is described in the Figure 2.

---

1. Start with empty set of concepts and properties

2. Add concepts and properties

3. Modify the ontology by the following ways

   a. Direct editing

      i. Add or remove concept

      ii. Add or remove property

      iii. Assign a property to concept or remove a property from the concept

   b. Editing as suggested by the ontology design tool

      i. When two concepts fall into one place, merge or add a distinction

      ii. Can generate the concepts which are formed by properties and super concepts of defined concepts, that are not explicitly mentioned in the concept table

4. Repeat until complete the ontology

Figure 1. Outline of the algorithm for designing ontology using FCA

Table 1. Example ontology design with FCA

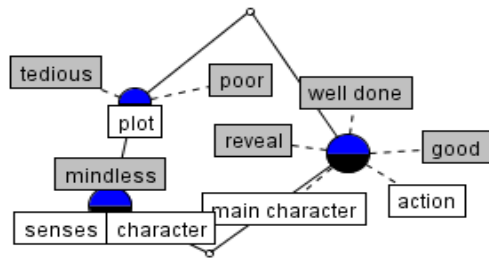|	Good	Poor	Well done	Mindless	Reveal	Tedious
Main character	X		X		X	
plot		X				X
senses		X		X		X
action	X		X			
character		X		X		X

Figure 2. Concept lattice form of some features of movie review domain

We developed the domain ontology in OWL (Web Ontology Language) based on FCA design. Implementation of domain ontology is used protégé 2000. Features that are parts of the domain are extracted by the SPARQL query language.

6. Conclusion

In this paper we proposed the approach of building the domain ontology based on FCA design. By using this approach, we can have better ontologies that are more suitable for knowledge sharing than pure taxonomies. Further development of this approach is still ongoing. More detail descriptions about this approach and comprehensive results will be reported in a follow up article.

7. References

[1] B. Ganter, R. Wille, "Formal Concept Analysis", Mathematical Foundation, Berlin, Springerr Verlag, 1999

[2] M. Benes, V. Snasel, "Deducing Design Class Hierarchy from Object Properties", ISM'2002

[3] R. Krovetz, W.B. Croft, "Lexical Ambiguity and Information Retrieval", 1991

[4] T. Quan, S. Hui, A. Fong, T. Cao, "Automatic Generation of Ontology for Scholarly Semantic Web", In The Semantic Web – ISWC 2004, LNCS, pp726-pp740. Springer, Hiroshima, 2004

[5] G. Zhang, A. Troy, and K. Bourgoin, " Bootstrapping Ontology Learning for Information Retrieval Using Formal Concept Analysis and Information Anchors", 14th International Conference on Conceptual Structures, Alborg, 2006

[6] P. Zhao, M. Zhang, S. Tang, "An Ontological Approach for Scienti, LNCS, pp699-710, Springer, Beijing, 2005

[7] P. Ciniamo, A. Hotho, S. Staab, "Comparing conceptual and divisive and agglomerative clustering for learning taxonomies from text", the European Conference On Artificial Intelligence (ECAI), Valencia, 2004