International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471
Vol. 3 Issue 7, July-2014, pp: (24-28), Impact Factor: 1.147, Available online at: www.erpublications.com

Literature Review of Ontology in Semantic Web

Sandeep
Department of Computer science, M.D. University, Rohtak, Haryana

Abstract: Ontology is an important emerging discipline that has the huge potential to improve information organization, management and understanding. It has a crucial role to play in enabling content-based access, interoperability, communications, and providing qualitatively new levels of services on the next generation of Web transformation in the form of the Semantic Web. The issues pertaining to ontology generation, mapping and maintenance are critical key areas that need to be understood and addressed. This timely survey is presented in two parts. This first part reviews the state-of-the-art techniques and work done on semiautomatic and automatic ontology generation, as well as the problems facing these researches. The second complimentary survey is dedicated to ontology mapping and ontology evolving. Through this survey, we identified that shallow information extraction and natural language processing techniques are deployed to extract concepts or classes from free-text or semi-structured data. However, relation extraction is a very complex and difficult issue to resolve and it has turned out to be the main impedance to ontology learning and applicability. Further researches are encouraged to find appropriate and efficient ways to detect or identify relations through semi-automatic automatic means.

Keywords: Ontology generation, ontology knowledge representation.

Introduction

Ontology is the term referring to the shared understanding of some domains of interest, which is often conceived as a set of classes (concepts), relations, functions, axioms and instances (Gruber, 1993). Now in knowledge representation community, the comm. Only used or highly cited ontology definition is from Gruber (1993): "ontology is a formal ,explicit specification of a shared conceptualization. 'Conceptualization' refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine readable. 'Shared' reflects that ontology should capture consensual knowledge accepted by the communities".

Ontology is a complex multi-disciplinary field that draws upon the knowledge of information organization, natural language processing, information extraction, artificial intelligence, knowledge representation and acquisition. Ontology is gaining popularity and is touted as an emerging technology that has a huge potential to improve information organization, management and understanding. In particular, ontology as the shared and common understanding of the domain that can be communicated between people and application systems, has a significant impact on areas dealing with vast amounts of distributed and heterogeneous computer-based information, such as World Wide Web and Intranet information systems, complex industrial software applications, knowledge management, electronic commerce and e-business. For instance, ontology plays the strategic role for agent communication (Hwang, 1999); ontology mapping is the capable way to break the bottleneck of B2B marketplace (Fensel et al., to appear) and ontology is the enabler to improve the intranet knowledge management systems (Kietz, Maedche and Volz, 2000). Ontology itself is an explicitly defined reference model of application domains with the purpose of improving information consistency and reusability, systems interoperability, and knowledge sharing. It describes the semantics of a domain in both a human-understandable and computer-process able way. The development at the World Wide Web Consortium (W3C) indicates that the first generation of World Wide Web will make its transition in future into the second generation of Semantic Web (Berners-Lee, Hendler and Lassila, 2001; Fensel et al.,2001). The term "Semantic Web" is coined by Tim Berners-Lee, the inventor of the

World Wide Web, to describe his vision of the next generation Web that provides muchmore automated services based on machine-process able semantics of data and heuristics (Berners-Lee and Fischetti, 1999). Ontologies, that provide shared and common domain theories, will be a key asset for this to happen. They can be seen as metadata that explicitly represent semantics of data in a machine-process able way. Ontology-based reasoning services can operationalize such semantics and be used for providing various forms of services (for instance, consistency checking, subsumption reasoning, and query answering, and so on). Ontologies help people and computers to access the information they need, and effectively

communicate with each other. They therefore have a crucialrole to play in enabling content-based access, interoperability, and communication across the Web, providing it with a qualitatively new level of service: the Semantic Web (Fensel, 2001).
Ontology learning is starting to emerge as a sub-area of ontology engineering due to the rapid increase of web documents and the advanced techniques shared by the information retrieval, machine learning, natural language processing and artificial intelligence communities. The majority of existing ontologies have been generated manually. Generating ontologies in this manner has been the normal approach undertaken by mostontology engineers. However, this process is very time-intensive, error-prone, and poses problems in maintaining and updating ontologies. For this reason, researchers are looking for other alternatives to generating ontologies in a more efficient and effective way. This survey aims to provide an insight into this important emerging field of ontology, and highlights the main contributions of ontology generation, mapping and evolving1 whose inter-relationships are shown in Figure 1.The survey is carried out over two parts, namely, the state-of-the-art survey on ontology generation and state-of-the-art survey on ontology mapping and evolving. In this first partof the survey on ontology generation, the areas of semi-automatic or automatic ontology generation will be covered. A subsequent paper will report on the ontology mapping and evolving.

Ontology generation in general

Although, the exist large-scale ontologies, ontology engineers are still needed to construct the ontology and knowledge base for a particular task or domain, and to maintain and update the ontology to keep it relevant and up-to-date. Manuallyconstructed ontologies are time-consuming, labor-intensive and error-prone. Moreover, a significant delay of updating ontologies causing currency problems actually hinders the development and application of the ontologies. The starting point for creating ontology could arise from different situations. An ontology can be created from scratch, from existing ontologies (whether global or localontologies) only, from a corpus of information sources only; or a combination of the latter two approaches (Uschold, 2000). Various degrees of automation could be used to build ontologies, ranging from fully manual, semi-automated, to fully automate. At present, the fully automated method only functions well for very lightweight ontology under very limited circumstances.

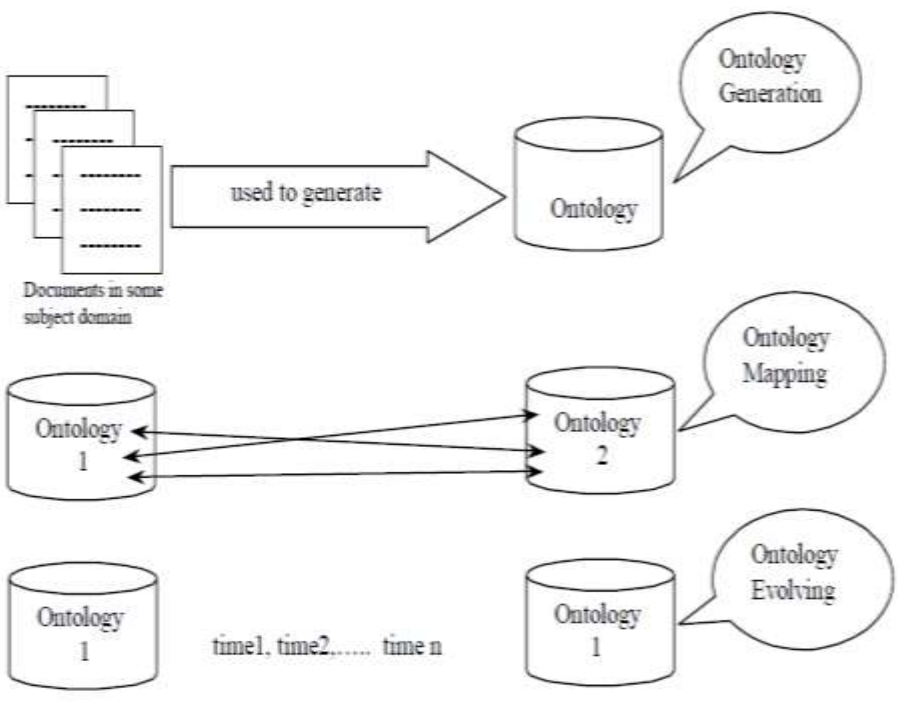

Figure I: General overview of ontology generation, mapping and evolving

Normally, methods to generate ontology could be summarized as bottom-up: from specification to generalization; top-down: from generalization to specification (e.g. KACTUS ontology); and middle-out: from the most important concepts to generalization and specialization (e.g. Enterprise ontology and Methodology ontology) (Fernandez-Lopez, 1999). Most often, lifting algorithms are used to lift and derive different levels of ontologies from a basic ontology (McCarthy, 1993).There are also a number of general ontology design principles that are proposed by different ontology engineers over a period of time. Guarino (1998) was inspired by philosophical research and proposed a methodology for ontology design known as "Formal Ontology" (Cocchiarella, 1991). This design principle contains theory of parts, theory of wholes, theory ofidentity, theory of dependence and theory of universals. He summarized the basic design principles that include the need to (1) be clear about the domain; (2) takeidentity seriously; (3) isolate a basic taxonomic structure; and (4) identify roles explicitly. Uschold & Gruninger (1996) proposed a skeletal methodology for building ontologies via a purely manual process: (1) identify purpose and scope; (2) build the ontology via a 3–step process: ontology capture (identification of the key concepts and relationships and provision of the definitions of such concepts and relationships); ontology coding (committing to the basic terms for ontology (class,entity, relation); choosing a representation language; writing the code);integrating existing ontologies; (3) evaluation (see Gomez-Perez etc., (1995)); (4)documentation; (5) guidelines for each of the previous phases. The final resulting ontology should be clear (definitions should be maximally clear andunambiguous), consistent and coherent (an ontology should be internally and externally consistent), extensible and reusable (an ontology should be designed in such a way as to maximize subsequent reuse and extensibility).

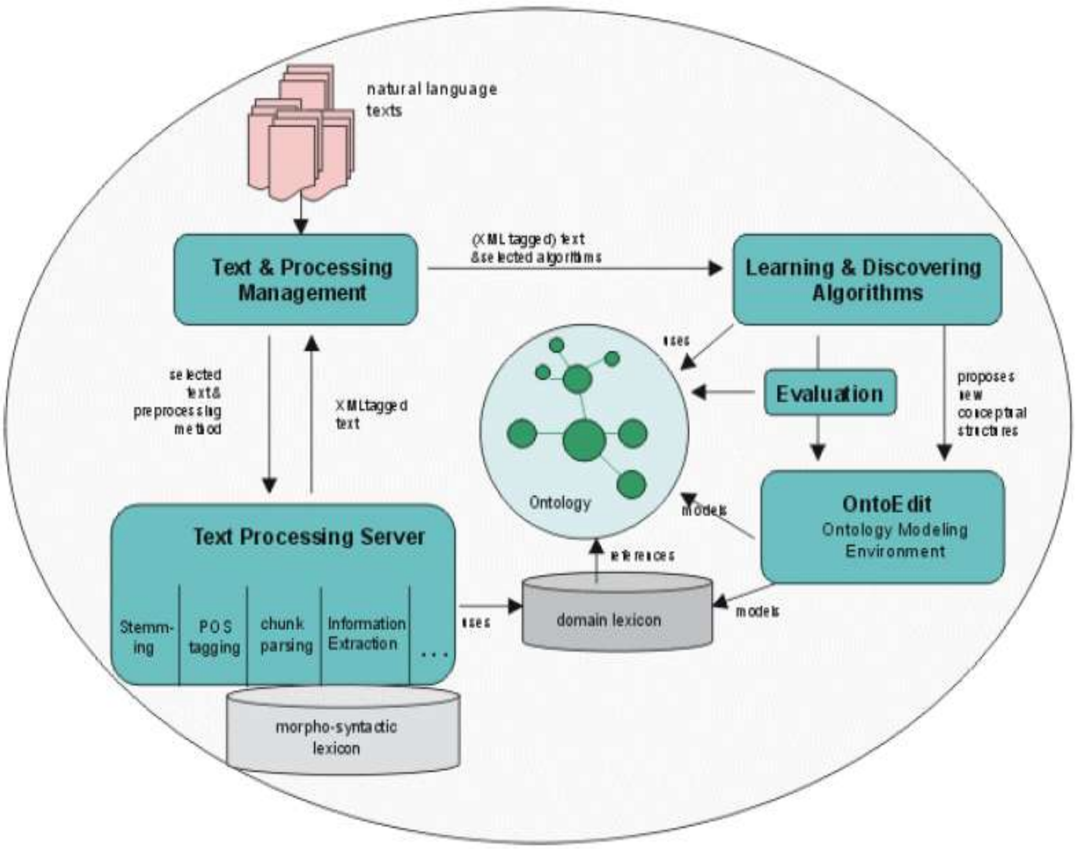

Figure II: General Overview of ontology learning system

Ontological Design Patterns (ODPs) (Reich, 1999) were used to abstract and identify ontological design structures, terms, larger expressions and semantic contexts. These techniques can separate the construction and definition of complex expressions from its representation to change them independently. This method was successfully applied in the integration of molecular biological information (Reich, 1999).
Fundamental steps which are key in the development of an ontology using a tool are [6]:

- Obtain domain knowledge: A deep insight into and a thorough knowledge of the respective domain is prerequisite to construction of any domain ontology.
- Identify the key concepts: Concepts that represent the domain are identified and hence implemented by means of classes.
- Build the taxonomy: the class hierarchy is created by creating the classes and their respective subclasses, and instances of classes.
- Identify relationships between classes: Properties are used to represent relationship between classes.

- Consistency checking: the constructed domain ontology must be checked for consistency using reasoning.
- Implementation of ontology: involves deployment of ontology to enable machine-to-machine communication.

Hwang (1999) proposed a number of desirable criteria for the final generated ontology tobe (1) open and dynamic (both algorithmically and structurally for easy construction and modification), (2) scalable and interoperable, (3) easily maintained (ontology should have a simple, clean structure as well as being modular), and (4) context independent. The remaining sections highlight the major contributions and projects that have been reported with respect to ontology generation. In each project, an introduction and background is first provided. This is followed by a description of the methods employed and concluded with a summary of problems that have surfaced during the process of ontology generation.

- Deeper semantics for class/subclass and cross-class relationships
- Ability to express such concepts and relationships in a description language; and
- Reusability and "share-ability" of the ontological constructs in heterogeneous system
- Strong inference and reasoning functions

The learning mechanism is based on the algorithm for discovering generalized association rules proposed by Srikant and Agrawal (1995). The learning module contains four steps: (1) selecting the set of documents; (2) defining the association rules; (3) determining confidence for these rules; and (4) outputting association rules exceeding the user-defined confidence. Higher levels of conception of descriptive vocabulary

Summary and Conclusions

As the first part of the survey of ontology generation, we have examined researches that are related to semi-automatic or automatic ontology generation. Table 1 summarizes the general pattern and characteristics of the various methods adopted by different research groups or researchers along the dimensions of the source data, methods for concept extraction and relation extraction, ontology reuse, ontology representation, associative tools and systems and other special features.
In general, we can observe the following salient points and features in ontology generation-

- Source data are more or less semi-structured and some seed-words are provided by the domain experts for not only searching for the source data but also as the backbone for ontology generation. Learning ontology from free-text or heterogeneous datasources are still within the research lab and far beyond the real applications.
- For concept extraction, there already exist some quite-matured techniques (such as POS, word sense disambiguation, tokenizer, pattern matching, etc.) that have been employed in the field of information extraction, machine learning, text mining andnatural language processing. The results of these individual techniques are promising as basic entities and should prove most useful in the formation of concepts in ontology building.
- Relation extraction is a very complex and difficult problem to resolve. It has turned out to be the main impedance to ontology learning and applicability. Further researches are encouraged to find appropriate and efficient ways to detect or identify the relations either semi-automatically or automatically.
- Ontologies are highly reused and reusable. Based on a basic ontology, other forms of ontologies may be lifted off to cater to specific application domains. This is important due to the cost of generation, abstraction and reusability.
- Ontologies can be represented as graph (conceptual graph), logic (description logic),web standards (XML), or a simple hierarchy (conceptual hierarchy). Currently there is the standard ontology representation language called DAML+OIL which combines the merits of the description logic, formal logic and web standards.
- A number of tools have been created to facilitate ontology generation in a semiautomatic or manual way. For instance, University of Karlsruhe (Germany) developed and commercialized the semi-automatic ontology editor called protégé. Stanford University exploited and provided an ontology-editing environment called Protégé with massive users.
- It is evident that much needs to be done in the area of ontology research before any viable Large scale system can emerge to demonstrate ontology's promise of superior information organization, management and understanding. Far beyond ontology generation, evolving and mapping existing ontologies will form another challenging area of work in the ontology field.

References

[1]. Abe, N. & Li, H. (1996). Learning word association norms using tree cut pair models. In.Proc. Of 13th Int. Conf. On Machine Learning, 1996.
[2]. Berners-Lee, T. & Fischetti, M. (1999). Weaving the Web. Harper San Francisco, USA.1999.
[3]. Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. Scientific American, May 2001.
[4]. Bisson, G., Nedellec, C. & Canamero, D. (2000). Designing clustering methods for ontology building: The Mo'K workbench. In. Proceedings of the First Workshop on Ontology Learning (OL-2000) in conjunction with the 14th European Conference onArtificial Intelligence (ECAI 2000). August, Berlin, Germany.
[5]. Borgo, S., Guarino, N., Masolo, C., and Vetere, G. (1997). Using a large linguistic ontology for internet-based retrieval of object-oriented components. In. Proceedings of1997 Conference on Software Engineering and Knowledge Engineering. Mardrid,Knowledge Systems Institute, Snokie, IL, USA.
[6]. Brin, S. (1998). Extracting patterns and relations from the World Wide Web. In. WebDBWorkshop at EDBT '98, 1998.
[7]. Chalendar, G. & Grau, B. (2000). SVETLAN: A system to classify nouns in context. In.Proceedings of the First Workshop on Ontology Learning (OL-2000) in conjunction withthe 14th European Conference on Artificial Intelligence (ECAI 2000). August, Berlin,Germany.
[8]. Cocchiarella, N.B. (1991). Formal ontology. In H. Burkhardt & B. Smith (eds.),Handbook of metaphysics and ontology (pp. 640-647). Munich: Philosophia Verlag.Constant, P. (1996). Reducing the complexity of encoding rule-based grammars.December 1996.
[9]. Ding, Y. (submitted). Ontology: The enabler for the Semantic Web. Journal ofInformation Science, submitted.
[10]. Faure, D. & Poibeau, T. (2000). First experiments of using semantic knowledge learnedby ASIUM for information extraction task using INTEX. In. Proceedings of the First Workshop on Ontology Learning (OL-2000) in conjunction with the 14th European Conference on Artificial Intelligence (ECAI 2000). August, Berlin, Germany.
[11]. Fensel, D. (2001). Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag, Berlin.
[12]. Fensel, D., Hendler, J., Lieberman, H. and Wahlster W. (Eds., to appear): Semantic Web Technology, MIT Press, Boston.
[13]. Fensel D., Omelayenko B., Ding Y., Schulten E., Botquin G., Brown M., and Flett A.(2001). Information Integration in B2B Electronic Commerce, to appear.
[14]. Fernandez-Lopez, M. (1999). Overview of methodologies for building ontologies. In., Proceedings of IJCAI-99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends, in conjunction with the Sixteenth International Joint Conference on Artificial Intelligence, August, Stockholm, Sweden.
[15]. Gaizausksa R. and Wilks Y. (1998). Information Extraction: Beyond document retrieval, Journalof Documentation, 54(1), 70-105.
[16]. Gomez-Perez, A.; Juristo, N. & Pazos, J. (1995). Evaluation and assessment of knowledge sharing technology. In. Mars, N.J. (Ed.). Towards very large knowledge bases– Knowledge building and knowledge sharing (pp. 289-296). IOS Press: Amsterdam.
[17]. Grishman R. (1997). Information extraction: Techniques and challenges. In M. T. Pazienza, International Summer School SCIE-97, Springer-Verlag.
[18]. Gruber, T.R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5, 199-220.