A Survey on Approaches, Methodologies, Standards, and Tools for Ontologies

Mahesh. D. Titiya and Dr. Vipul. A. Shah

Computer Engineering Department, Gujarat Technological University, India,
Instrumentation and Control Department, Dharmsinh Desai University, India,

Received February 2016; Accepted 18 April 2016; Available 25 April 2016

ABSTRACT

The increasing volume and unstructured nature of data available on the World Wide Web (WWW) makes information retrieval a tedious and mechanical task. Lots of this information is not semantic driven and hence not machine processable, but it's only in human readable form. The WWW is designed to builds up a source of reference for web of meaning. The idea of Semantic Web and its importance has grown tremendously in the recent years. Ontology has become the key technique to annotate semantics and provide a common foundation for various complex resources on the semantic web. Ontology, a branch of artificial intelligence, is a formal representation of concepts of a particular domain and relationships amongst those concepts. Ontology acts as a powerful tool or a driving factor for many real world applications. The use of semantic web and ontologies in information system has become more and more popular in the various fields such as web technologies, database integration, natural language processing etc. ontologies can be also be used to better organize information resources and assist users in retrieving relevant information. This paper aims to introduce semantic web, ontologies and issues concerning ontology representation. Then it discusses the various methodologies, approach, standards and tools which are used for the ontology construction.

KEYWORDS: Concepts, Ontology, Ontology engineering, Ontology learning, Semantic web.

INTRODUCTION

With the enormous amount of data which is existing on the internet. The Semantic Web is a set of techniques and technologies used for knowledge representation in a given domain of application. According to the World Wide Web consortium (W3C), it provides a common framework, enabling data to be shared and reused across applications and community boundaries. In fact, Knowledge representation uses the semantic web languages combined AI. This technique is usually Used to improve information retrieval, knowledge sharing/reuse and systems interoperability. The vision of Semantic Web is the idea of having data on the Web defined and linked in such a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications. In order to achieve the above vision, the layers of Semantic Web according to Tim Berners-Lee are described in Figure.1 in a layered architecture.
Extensible Markup Language (XML [5]) helps in achieving a common syntax while RDF [6] and RDFS [7] provide data models and modeling primitives like class hierarchies, relationships respectively. RDFS is a primitive ontology language.[4] Ontology is the layer that achieves the semantics of a domain. Ontology[7] is a specification of a conceptualization. Ontologies are developed to make data machine processable. Ontologies have been developed in various domains to capture the semantics of the domain, for example Gene Ontology [8] (Genetics), Agrovoc [9] (Agriculture) etc. Various tools like Protégé [10], powl [11], OilEd [12], SWOOP [13] etc used to develop ontologies. These are the tools available for free.

II. Overview of the semantic web Technologies:

The key components of semantic web technology for representing knowledge are usually classified into three main categories: The Resource Description Format (RDF), RDF Schema and The Ontology Web Language (OWL) Standard. RDF enables to add meta-data to web resources. RDF Schema is an extensible knowledge representation language, providing basic elements for the description of the ontologies. OWL enables to create ontologies using logic techniques and inferences. In general, the resulting knowledge models are kinds of ontologies. Indeed, ontology is an explicit specification of a conceptualization [14]. On other words, ontology is such a conceptual network that includes a complete set of concepts and notions for describing a specific domain. These concepts are linked to each other by taxonomic relationships and by semantic. The formal content of ontology consists of classes, properties, individuals and relations where classes are concepts from specific-domain knowledge, properties are specific attributes related to classes, individuals are instances of classes and finally relations are ways in which classes and individuals can be related to one another. Thus, the main purpose of ontology is to improve the communication between humans, between humans and computers and between computers with a unique and a standardized vocabulary [14].

2.1. Resource Description Format (RDF) Standard:

RDF is a standard model for data interchange on the Web. It enables to describe semantically web resources in order to improve their referencing and their retrieval by users or by applications. The syntax of RDF consists of a triplet of assertions: subject, predicate and object [10]. The subject represents the resource to describe; the predicate is a type of property related to the resource and the object is the value of the property such as data or other resource (Figure 2).
The general structure of RDF forms a structured graph with nodes and edges which represent the triplet. The subject is the source node, the object is the targeted node and the predicate is the label of the edge. Hence, unlike OWL, RDF is not specific to ontologies, it focuses on web resources. In fact, RDF allows annotating web resources or any addressable subjects with URIs (Uniform Resource Identifier) with meta-data. The meta-data are often very basic and have a low semantic abstraction because RDF is mainly devoted to reference resources on the web. RDF is supported by many tools such as Protégé, OntoEdit etc.

2.2. Ontology:

Aristotle was the first to define what Ontology is. From then and till now people like Thomas Gruber, Tim Berners Lee and many more have redefined Ontology. From all those definitions we can state that Ontology, a branch of artificial intelligence, is a formal representation of concepts of a particular domain and relationships amongst those concepts. In more simplified words, Ontology is the knowledge representation of a domain of interest. Ontology is commonly used as a structure capturing knowledge about a certain area via providing relevant concepts and relations between them. Ontology which is created with different approaches like pattern based extraction, Ontology pruning, data mining etc., acts as a powerful tool in many modern knowledge based systems. It supports natural language processing, information filtering, information retrieval and data access. Ontology acts as a driving force for many real world applications like the Semantic Web, Business Process Networks.

III. Need of Ontology:

The Web became an object of our daily life and the amount of information in the web is ever growing. Besides plain texts, especially multimedia information such as graphics, audio or video has become a prevalent part within the web's information transport. But how to find some useful information within this huge information space? How to achieve homogeneity in heterogeneous digital information resources? The increasing volume and unstructured nature of data available on world. World Wide Web (WWW) makes information retrieval a complex and mechanical task. Lots of this information is not semantic driven, and hence not machine understandable, but it’s only in human readable form. Almost since the emergence of computers, one basic goal of computer science research has been to be able to process the meaning of symbols and not only the syntactic structures of the language. The World Wide Web is a system of interconnected hypertext documents that are accessed via the Internet. These web pages follow the HTML, XML languages which are unstructured heterogeneous resources so unable to make an efficient search. To overcome with this problem, ontology is used to conceptualize the domain and represent an area of knowledge. There is a need to analyse or conceptualize a domain to invent an effective knowledge representation system hence ontology vocabulary to share the information and to make machine interpretable concepts. Ontology is important because of the reasons are:

IV. Ontology as a tool for knowledge management:

In the current economy, knowledge has become a key success factor. Knowledge is derived from thinking, and it is a combination of information, experience and insight. Deriving knowledge from information requires human judgment, and is based on context Ontologies are used for organizing knowledge in a structured way in many areas from philosophy to Knowledge.

4.1. Ontology learning:

Ontology learning is a task of information extraction. The aim of Ontology learning is to extract relevant concepts and relations from data sets either automatically or semi automatically to form Ontology. Global knowledge bases were used by many existing models to learn Ontology for web information gathering. Other models used user local information for mining background knowledge. Li and Zhong [10] used pattern recognition and association rule mining techniques to extract knowledge from user local documents and used the same for ontology construction. Tran et al. Converted keyword queries [13] to Description Logics’ conjunctive queries and used ontology for representation of user’s background knowledge. Zhong [3] proposed a domain ontology learning approach that employed various data mining and natural-language understanding techniques. The data mining techniques used in these models led to more user background knowledge discovery. The drawback is that these models contained noise and uncertainties. Ontology was used in many works to improve the performance of knowledge extraction. Using a fuzzy domain ontology extraction algorithm, a mechanism was developed by Lau et al. [16] to construct concept maps based on the posts on online discussion forums. Quest and Ali [8] used ontology to help data mining in biological databases. Jin et al. [15] integrated data mining and information retrieval techniques to further to enhance the knowledge discovery.
4.2. Personalized ontology mining:
Xiaohui Tao [17] proposes a personalized ontology model to attempt the challenge to use semantic relations of “kind-of”, “part-of”, “related-to” and synthesize common sense and expert knowledge in a single computational model. This proposed model is evaluated by assessing the success of its application to a web information gathering system. This model attempts to facilitate the user possessed concept model and to generate the personalized user profile for web information gathering. There is no mathematical model to formalize these three relations namely “kind-of”, “part-of”, “related-to” together. This two-dimensional method [18] can quantitatively analyze these three semantic relations.

V. Ontology construction:
Ontology building which is an automatic or semi-automatic process has the following as its fundamentals [6]: which needs to follow for the better construction of the ontology.

5.1. Components of Ontology:
There are five kinds of ontology components: classes, relations, functions, formal axioms and instances.

1) Classes represent concepts, which can be considered generic entities.
2) Relations represent a type of association between concepts of the domain.
3) Functions are a special case of relations.
4) Formal axioms serve to model sentences that are always true. They are normally coherent description expressions.
5) Instances are used to represent elements or individuals in ontology. An ontology can be characterized as comprising of four tuples: \( O = C, R, I, A \).
   - \( C \) is a set of classes representing concepts in the given domain.
   - \( R \) is a set of relations holding between classes.
   - \( I \) is a set of instances, where each instance can be an instance of one or more classes and can be linked to other instances by relations.
   - \( A \) is set of axioms

5.2. Methodology for ontology construction:
The Methodology to produce the ontology includes various stages which is explained below.

5.2.1. Specification:
Identify the purpose and scope of the ontology. The Purpose is obtained by answering the question —Why is the ontology being built? and the scope is obtained by answering the question —What are its intended uses and end users?

5.2.2. Knowledge acquisition:
Acquire knowledge about the subject either by using elicitation techniques on domain experts or by referring to relevant bibliography. Several techniques can be used to acquire knowledge, such as brainstorming, interviews, questionnaires and text analysis.

5.2.3. Conceptualization:
Structure a conceptual model that describes ontology to be built, so that it meets the specification found in the previous step. The conceptual model of ontology consists of concepts in the domain and relationships among those concepts. Concepts can be organized with a class hierarchy, including super classes and subclasses concept. Relationships between concepts can be grouped in two main groups: hierarchical relationships and associative relationships. Hierarchical relationship identifies the hierarchy between super classes, subclasses. Associative relationship connects concepts which are not in the same hierarchy.
5.2.4. Formalization:
Transform the conceptual description into a formal model, that is, the description of the domain found in the previous step is written in a more formal way, although not yet its final form. Concepts are usually defined through axioms that restrict the possible interpretations for the meaning of those concepts.

5.2.5. Implementation:
Implement the formalized ontology in a knowledge representation language. For that, one chooses a representation language and writes the formal model in the representation language (OWL) using the representation ontology.

5.2.6. Evaluation:
Process of technically judging the quality of the ontology.

5.2.7. Documentation:
Report what was done, how it was done and why it was done. Documentation associated with the terms represented in the ontology is particularly important, not only to improve its clarity, but also to facilitate maintenance, use and reuse.

5.2.8. Maintenance:
Updation and correction of the implemented ontology.

5.3. Approaches to construct an Ontology:
5.3.1. Manually-driven by domain-experts:
This approach relies totally on experts of the area. The experts will set the rules and concepts and sort of relationships based upon experts knowledge and experience of the Knowledge domain.

5.3.2. Automatic approach:
This approach will construct ontology by using a computer program, whereas the program will be produced according to rules and conditions posed by developer with the help of experts and the computer.

5.3.3. Semi-automatic approach:
This approach also uses computer program but the ontology builder will have product’s accuracy and rules verified and confirmed by expert who created the rules.

5.4. Tools for building Ontologies:
When we are about to build ontology, several basic questions arise related to the tools to be used: Which kinds of tools are the most convenient for building my ontology? Which kinds of tools give better support to the ontology development process? Etc. There are many kinds of tools which are available for building anthologies. We have grouped them in the following clusters:

5.4.1. Ontology development tools:
This group includes tools, environments and suites that can be used for building a new ontology from scratch or reusing existing ontologies. Apart from the common edition and browsing functionality, these tools usually include ontology documentation, ontology exportation and importation from different formats, graphical views of the ontologies built, ontology libraries, attached inference engines, etc.

5.4.2. Ontology merge and integration tools:
These tools have appeared to solve the problem of merging or integrating different ontologies on the same domain. This need appears when two companies or organizations are merged together, or when it is necessary to obtain a better quality ontology from other existing ontologies in the same domain.

5.4.3. Ontology-based annotation tools:
These tools have been designed to allow users inserting and maintaining (semi)automatically ontology-based markups in Web pages. Most of these tools have appeared recently, along with the emergence of the Semantic Web. Most of them are already integrated in an ontology development environment. Ontology storage and querying tools.- These tools have been created to allow using and
querying ontologies easily. Due to the wide acceptance and use of the Web as a platform for communicating knowledge, new languages for querying ontologies have appeared in this context.

5.4.4. Ontology learning tool:
They are used to (semi)automatically derive ontologies from natural language texts. However, these tools will not be revised in this deliverable, since there will be a specific deliverable from OntoWeb.

VI. Challenges:
6.1. Vastness:
The World Wide Web contains enormous Numbers of web pages. The existing technology has not yet been able to eliminate all semantically duplicated terms. Any automated reasoning system will have to deal with truly huge inputs.

6.2. Vagueness:
Its mainly imprecise concepts like "young" or "tall". This arises from the vagueness of user queries and the concepts represented by content providers; Fuzzy logic is the most common technique for dealing with vagueness.

6.3. Uncertainty:
These are precise concepts with uncertain values. For example, a patient might present a set of symptoms which correspond to a number of different distinct diagnoses each with a different probability. Probabilistic reasoning techniques are generally employed to address uncertainty.

6.4. Inconsistency:
These are logical contradictions which will inevitably arise during the development of ontologies. Deductive reasoning fails catastrophically when faced with inconsistency, because "anything follows from a contradiction".

6.5. Deceit:
This is when the producer of the information is intentionally misleading the consumer of the information. Cryptography techniques are currently utilized to alleviate this threat.

Conclusion:
Even if many technologies exist in knowledge representation field for improving the “findability” of information, few of them are suitable to deal with the problem of large quantities of information stored in different places. Ontology is the most suitable standard for representing knowledge in problem solving situations. We have discussed the steps for construction of the ontology and to extend ontologies using unstructured text sources. In future the modeling. However it is open question that whether formal specifications are the only way to reach the goal, or whether the manual approach of hand coding semantics will be outperformed by inconsistentm, statistical black box methods again.

REFERENCES