A Framework for Ontology Creation and Management for Semantic Web

Imran Sarwar Bajwa, *Member, IASCIT*

***Abstract*—An ontology is a model of reality of the world and the concepts in the ontology must reflect this reality. Ontologies are building blocks of Semantic Web based systems. Creating ontologies is not an easy task and obviously there is no unique correct ontology for any domain. There are many other important issues related to the ontology domain engineering some of which are ontology integration, ontology mapping, ontology translation, ontology reuse and ontology consistency check. Due to unavailability of any standard for ontology building, ontologies on the same subject are different. There are different ontology tools that use different ontology languages. Due to these reasons, interoperability between the ontologies is very low. Current ontology tools concentrate mostly on the functions: create, edit, and inference. Most of the tools do not support the merging of heterogeneous domain ontologies. Moreover, the issues of duplicate information across documents and redundant annotations are major challenges of automatic ontology creation as the automatically populating ontology from diverse and distributed web resources poses significant challenges.**

***Index Terms*—Ontology Creation, Ontology management, Natural Language Processing, Semantic Web.**

I. Introduction

An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them. Ontologies are used to represent domain knowledge in a digital form. This information can be shared and reused. Ontology editors as Protégé-2000 and Apollo are used for ontology creation [4]. These ontology editors are difficult to use and require skilled knowledge. Ontology is the necessity factor of the semantic web and there are languages tools for ontology creation and management. For constructing on ontology, many ontology languages have been developed, OWL, RDF(S), DAML+OIL and so on [5]. Many ontology tools such as *Protege*, *OILEd*, *KAON*, etc have been designed based on these languages. Ontologies are being used in the various fields such as medical service, e-business, e-commerce, knowledge management and information retrieval.

In ontology creation, ontologies have two domains: class ontologies and domain ontologies [3]. The specific and general concepts are identified in the given domain. Depending on the domain, special techniques can be employed to identify concepts and their relationships. The identified concepts or classes are then inserted into the taxonomy at appropriate insertion points. The particular properties for a concept are defined, helping to differentiate among various concepts [7]. The restrictions or facets for these properties are also defined and at last, the ontology designed is revised according to the requirements. A common ontology retrieval system consists of the crawler that brings and parses the HTML documents from web, the classifying module that classifies the parsed the OWL/RDF(S) [2] documents into each domain, the ranking module that decides the order of the classified domain ontologies, and the retrieval module that provides the interface to users to retrieved domain ontology easily. The ontology crawler executes the core function that brings OWL/RDF(S) ontologies from web to local repository. For the OWL/RDF(S) ontologies, the crawler parses Running WordNet HTML documents. There is the HTML Parser in the crawler. After parsing the HTML documents [5], if the parser considers the link that includes documents, the parser saves the link in the queue.

Manuscript received November 25, 2010. This research work was conducted in the department of Computer Science & IT, The Islamia University of Bahawalpur and supported in part by the Higher Education Commission, Pakistan. This paper is the continuation of the series of publications under the research project of Natural Language Processing based Software Designing by the Information Processing Research Group (IPRG).

I. S. B. Author is Assistant Professor in the Department of Computer Science & IT, The Islamia University of Bahawalpur. He is regular member IACSIT. (phone: +92 (062)9255466; e-mail: imran.sarwar@iub.edu.pk)

II. Challenges in Ontology Managment

Due to unavailability of any standard for ontology building, ontologies on the same subject are different. There are different ontology tools that use different ontology languages. Due to these reasons, interoperability between the ontologies is very low. Current ontology tools concentrate mostly on the functions: create, edit, and inference [13]. Methods for merging heterogeneous domain ontologies are not in most tools. Issues of duplicate information across documents and redundant annotations are still major challenges of automatic ontology the knowledge extractor. The extractor has a component population. Automatically populating an ontology from diverse and distributed web resources poses significant challenges.

Ontologies are building blocks of Semantic Web based systems. Creating ontologies is not an easy task and obviously there is no unique correct ontology for any domain. There are many other important issues related to the ontology domain engineering some of which are ontology integration, ontology mapping, ontology translation, ontology reuse and ontology consistency check. An ontology is a model of reality of the world and the concepts in the ontology must reflect this reality [6]. After we define an initial version of the ontology, we can evaluate and debug it by using it in

applications or problem-solving methods or by discussing it with experts in the field, or both. As a result, we will almost certainly need to revise the initial ontology. This process of iterative design will likely continue through the entire lifecycle of the ontology. By providing, a standard for ontology building, ontologies on the same subject will be easy to create and reuse.

In ontology creation process, some fundamental rules should be emphasized in ontology design. These rules may seem rather dogmatic [8]. They can help, however, to make design decisions in many cases.

1) There is no one correct way to model a domain— there are always viable alternatives. The best solution almost always depends on the application that you have in mind and the extensions that you anticipate.
2) Ontology development is necessarily an iterative process.
3) Concepts in the ontology should be close to objects (physical or logical) and relationships in your domain of interest. These are most likely to be nouns (objects) or verbs (relationships) in sentences that describe your domain

III. NLP based Ontology Management

A general ontology merging process involves six steps: feature engineering, selection of next search steps, similarity computation, similarity aggregation, interpretation and iteration of this whole process. The methodology designed for merging the heterogeneous ontologies based on WordNet. The proposed framework has four distinct phases. In the first phase, WordNet [3] is used as a similarity measure in different ontologies. Then in a second phase, selection of the mass similar concept is performed. Phase three computes similarity and finally the reconstruction of the new ontological hierarchy is performed.

There is no one "correct" way or methodology for developing ontologies [9]. Here we discuss general issues to consider and offer one possible process for developing an ontology. We describe an iterative approach to ontology development: we start with a rough first pass at the ontology. Then we revise and refine the evolving ontology and fill in the details. Along the way, we discuss the modeling decisions that a designer needs to make, as well as the pros, cons, and implications of different solutions [10, 11, 12].

A. Determine the domain and scope of the ontology

First step is the development of ontology by defining its domain and scope. That is, answer several basic questions: What is the domain that the ontology will cover? For what, we are going to use the ontology? For what types of questions the information in the ontology should provide answers? Who will use and maintain the ontology? The answers to these questions may change during the ontology-design process, but at any given time they help limit the scope of the model.

B. Consider reusing existing ontologies

It is almost always worth considering what someone else has done and checking if we can refine and extend existing sources for our particular domain and task [14]. Reusing existing ontologies may be a requirement if our system needs to interact with other applications that have already committed to particular ontologies or controlled vocabularies.

C. Enumerate important terms in the ontology

It is useful to write down a list of all terms we would like either to make statements about or to explain to a user. What are the terms we would like to talk about? What properties do those terms have? What would we like to say about those terms?

D. Define the classes and the class hierarchy

This step starts by defining classes. From the list which created in Step 3, the terms are selected that describe objects having independent existence rather than terms that describes these objects. These terms will be classes in the ontology and will become anchors in the class hierarchy. Classes are also organized into a hierarchical taxonomy [15, 16].

E. Define the properties of classes—slots

The classes alone will not provide enough information to answer the competency questions from Step 1. Once we have defined some of the classes, we must describe the internal structure of concepts.

F. Define the facets of the slots

Slots can have different facets describing the value type, allowed values, the number of the values (cardinality), and other features of the values the slot can take.

IV. Related Work

From the W3C's (World Wide Web Consortium) inception, there was a perceived need to bring order to the loosely connected networks of digital documents that made up the Web. Although this order was to be realized by consortium's development of standards, it would also reflect the order that libraries have and the Web does not – a consistent structure by which people can access materials. More recently, we can see evidence that this view of the Semantic Web is still widely held in the Hypertext and World-Wide Web communities [8]; Scenario 1 in [9], an information access scenario in which the retrieval is aided by semantic metadata, is a good example. Semantic Web is about the modern web that contains the meanings of information and services available on web. A second perspective for the Semantic Web is one of a globally distributed knowledge base. This perspective on the Semantic Web was put forth early in the Web's development by Berners-Lee, who began his efforts with the aim of eventually creating networked knowledge ontologies [3]. Berners-Lee has gone on to describe the Semantic Web as being able to learn from the experience of Cyc, creating an infrastructure for knowledge acquisition, representation, and utilization across diverse use contexts [4]. In scenarios reminiscent of Apple's Knowledge Navigator vision from the mid 1980's, this global knowledge base will be used by personal agents to collect and reason about information, assisting people with tasks common to everyday life.

Ontobase Ontology Repository [1] is an implementation of a design that allows users and agents to retrieve ontologies and metadata through open Web standards and ontology services. The *Ontobase* provides a knowledge management mechanism by maintaining structural and semantic information about each data source, recording the relationship between attributes of the data sources with terms from a business domain, and computing contextual information gleaned from these linkages and other resource related information. Another method [2] based on *WordNet* [3] has been presented to merge the heterogeneous domain ontologies.

WordNet uses a dictionary to detail the relationships between the concepts like the synonym, antonym, hypernym and hyponym. The main idea was to merge the taxonomies, because they are central components of ontologies. After evaluation, it was determined that the new methodology is very efficient in merging between heterogeneous ontologies.

V. Conclusion

The future requirements in ontology engineering field include these areas: ontology integration, mapping and reuse; developing the core standard ontologies in different fields; integration of ontologies in processes and applying the time notion. Multilinguality is recognized as one of the most important challenges of the Semantic Web. A yet more important issue is dealing with developed ontologies used in cultural requirements of different countries and areas. Different people think about a single concept differently, and this has some roots in cultural and historical backgrounds. When the content is understood by a machine, new pieces of information will be produced. According to the first vision of Semantic Web, it should be able to provide trust and proof at the highest layers. Some essential challenges that the Semantic Web will confront us with are: providing mechanisms for different levels of trust ontologies and secondly providing trust on assertions, which are inferred from existing concepts. Web services enable us to access relevant applications nevertheless the discovery, invocation and composition of web services still need to be supported by human interaction. Some challenging requirements that should be fulfilled by research projects are: establishing ontologies for service descriptions and classification; semantic Web service trust and proof; Knowledge Representation for Semantic Web Services; semantics for service delegation and knowledge aggregation.

References

[1] Ding Pan, Yan Pan, "Using Ontology Repository to Support Data Mining", The Sixth World Congress on Intelligent Control and Automation, 2006. WCICA 2006, Volume: 2, pp. 5947- 5951

[2] Hyunjang Kong, Myunggwon Hwang, Pankoo Kim, "A New Methodology for Merging the Heterogeneous Domain Ontologies based on the WordNet", International Conference on Next Generation Web Services Practices, 2005. NWeSP 2005. page(s): 6 pp.-

[3] Sahoo, K. Vidyasagar, V.E., (2003) "Kannada WordNet - a lexical database", Conference on Convergent Technologies for Asia-Pacific Region, Volume 4, Page(s):1352 - 1356 Vol.4

[4] Jan Uhlir, Magda Machkova, Pavel Hajeck, Tereza Cermanova, "Creation of Architectural Ontology: User's Experience" Proceedings. 14th International Workshop on Database and Expert Systems Applications, 2003. 1-5 Sept. 2003 Page(s):65 - 69

[5] Myunggwon Hwang, Hyunjang Kong, Pankoo Kim, "The Design of the Ontology Retrieval System on the Web", Advanced Communication Technology, 2006. ICACT 2006. The 8th International Conference, Volume: 3, pp.1815- 1818

[6] McGuinness, D.L. and Wright, J. (1998). Conceptual Modeling for Configuration: A Description Logic-based Approach. Artificial Intelligence for Engineering Design, Analysis, and Manufacturing - special issue on Configuration.

[7] Brickley, D. and Guha, R.V. (1999) "Resource Description Framework (RDF) Schema Specification. Proposed Recommendation", World Wide Web Consortium: http://www.w3.org/TR/PR-rdf-schema.

[8] Chimaera, Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B. and Benjamins, V.R. (2000). Ontology Environment. www.ksl.stanford.edu/software/chimaera

[9] WonderTools? A comparative study of ontological engineering tools. International Journal of Human-Computer Studies 52(6): 1111-1133.

[10] Farquhar, A. (1997). Ontolingua tutorial. http://ksl-web.stanford.edu/people/axf/tutorial.pdf

[11] Gómez-Pérez, A. (1998). Knowledge sharing and reuse. Handbook of Applied Expert Systems. Liebowitz, editor, CRC Press.

[12] Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. Knowledge Acquisition 5: 199-220.

[13] Gruninger, M. and Fox, M.S. (1995). Methodology for the Design and Evaluation of Ontologies. In: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.

[14] Hendler, J. and McGuinness, D.L. (2000). The DARPA Agent Markup Language. IEEE Intelligent Systems 16(6): 67-73.

[15] Humphreys, B.L. and Lindberg, D.A.B. (1993). The UMLS project: making the conceptual connection between users and the information they need. Bulletin of the Medical Library Association 81(2): 170.

[16] McGuinness, D.L., Fikes, R., Rice, J. and Wilder, S. (2000). An Environment for Merging and Testing Large Ontologies. Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference (KR2000).

Imran Sarwar Bajwa is Assistant Professor in Computer Science Dept., The Islamia University of Bahawalpur.. He has been teaching and doing research in various universities in Pakistan and University of Coimbra, Portugal & University of Birmingham, UK