A Peer-review Approach for Ontology Evaluation

Kaustubh Supekar, M.S.
Section on Medical Informatics, Stanford University School of Medicine, Stanford, CA

Ontology Evaluation

The semantic web is intended for knowledge sharing among agents as well as humans [Berners-Lee *et al*., 2001]. To achieve this goal, ontologies, which express knowledge in a certain representation as well as in machine interpretable form were introduced, and have grown considerably in number. To have a sustained growth of the semantic web and to have better interoperability between intelligent systems and applications it is highly desirable and is very critical to reuse existing ontologies. Furthermore, ontology engineering is an exceedingly intricate and challenging task requiring specialized design skills as well as comprehensive domain knowledge. Reuse of ontologies present on the web will ease this burden. Users who wish to reuse these ontologies, however, are often confronted with knowledge resources that cover overlapping domains of interests, and that vary in quality of their content. To our knowledge, currently, there are no operational quantitative or qualitative methodologies to assess the quality of ontology content. Consequently, more often than not knowledge engineers will develop their ontologies from scratch rather than reuse existing ones.

We propose a peer-review based approach for ontology evaluation. Our approach will allow the user to provide qualitative ratings on the ontology content. We believe that this approach will lead to the discovery and selection of the most relevant ontology for the purpose and will facilitate ontology reuse. In this paper, we provide results of our preliminary work on the development of an ontology model of metadata elements that captures quality features of an ontology, and a prototype system that will allow the user to associate these elements to an ontology.

Related Work

Two recent works [Patel *et al*., 2003; Ding *et al*., 2004] on ontology search have proposed a pagerank [Page *et al*., 1998] based algorithm to rank ontologies on the semantic web. These approaches work on the assumption that ontologies are akin to web pages that are heavily interconnected using hyperlinks. In the case of ontologies on the semantic web, the hyperlinks are relationships between classes and properties of classes (owl:sameAs, rdfs:isDefinedBy, rdfs:subClassOf)[1].However, this assumption is inappropriate as long as the rate of ontology reuse is very low, and, therefore, cross-links between ontologies are sparse. Therefore, there is a chicken-and-egg problem: the existing methods to find appropriate ontologies do not work due to lack of links between ontologies, and users do not link their ontologies due to lack of methods to find the existing ontologies.

A competing approach [Supekar *et al*., 2004] that is not based on this assumption provides a formal criterion and computational methodology to gauge the semantic correctness and validity of ontology. Other researchers have proposed quantitative criteria about ontological structure, which can be distinguished from qualitative comparison criteria [Magkanaraki *et al*., 2002].

---

[1] W3C OWL Web Ontology Language Guide http://www.w3.org/TR/owl-guide/

However, these methods usually do not provide information on whether the ontology is suited for a particular task or a purpose. We argue that even though it is valuable to know the semantic validity and consistency through formal means, it is equally important to know subjective information about the ontology, such as what is the maturity of ontology content? What is the intended use and purpose of the ontology? How do other users rate the ontology under consideration? For the purposes of ontology evaluation we have proposed to implement a framework that will allow the user to associate this and other subjective information with the ontology. The proposed system will also implement the functionality to verify the syntactic correctness and semantic consistency of ontologies, and we will also use and develop automated methods to compute and extract ontology metadata such as number of concepts, instances, and relationships, and links to external knowledge resources.

Metadata Ontology

To enable the creation and retrieval of ontology metadata, we have designed an ontology of metadata elements and have implemented a prototype tool that supports creation of metadata for ontology resources based on the Metadata ontology. For each resource, the Metadata ontology provides two types of information: (1) source metadata, which include the metadata provided by ontology authors and generated by ontology-development tools; and (2) third-party metadata, which are provided by ontology users and that include peer reviews of ontologies, usage and experience information, and ratings. Some of the metadata categories included in the ontology are:

- Domain of the ontology (using controlled terminology, when possible); informal description of the content; intended use of the ontology;
- Version number; contact and author information; supporting institutions; availability and licenses; citations and references;
- Verification tools used and development methodology;
- Naming policy; policy for extensions; reliance on other ontologies
- Peer reviews; experience reports; usage data; ratings along different axes, such as coverage; degree of formality.

The Metadata ontology is flexible and extensible and we plan to make enhancements as we discover new requirements when using metadata in the proposed system.

Prototype Tool for Creating Ontology Metadata

We have developed a prototype tool that supports the creation of ontology metadata. The tool uses a Web-based form interface that is driven by the Metadata ontology (Fig.1). Users can enter textual information such as the name of the author, a description of the ontology, as well as controlled information such as intended use, and representation format. The metadata values are stored as properties for instances of the Metadata ontology and can be indexed and retrieved as any other ontology instances.

The tool employs a dynamically generated UI that leverages the rich semantics encoded in the Metadata ontology to provide contextual interface based on the ontology representation format, level of expertise of the user, type of user, and importance of a particular metadata element. For example, based on the axioms in the metadata ontology, an expert user who wishes to annotate a description logics based ontology is presented with different set of metadata information as compared to a naive third-party user annotating a frame based ontology. Controlled information such as conceptual representation {Directed Acyclic Graph, Logic based, and Frames} and the representation language {XML, Text, RDF, and OWL} can be entered through the drop down boxes which are populated from the metadata ontology. Users can extend the metadata ontology

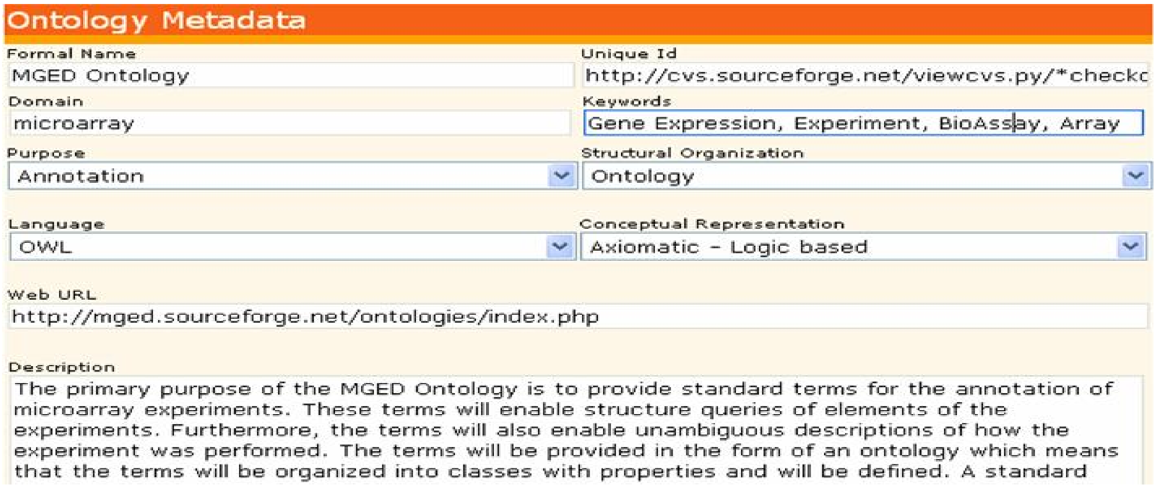

Figure 1: Form to annotate an ontology resource. The interface shows the metadata information captured for the MGED Ontology.

to customize this controlled vocabulary. The tool also allows the user to specify the domain of ontology in terms of an external identifier. For example, users can enter UMLS:CUI C0887809 (*Microarray Analysis of Gene Expression*) to specify the domain of the MGED ontology.

Discussion and Future Work

We intend to extend the prototype system to incorporate features that will allow the user to search, store and retrieve metadata annotations. Indexing information such as keywords that describe the topic of ontology content that is provided, as part of metadata ontology will serve as a mechanism to develop a keyword based index of ontologies. Allowing all the users on the web to post reviews on an ontology resource raises important issues of security and trust. We intend to develop a web of trust that is based on the paradigm of open rating systems that will allow users to rate reviewers as well as the content of the review.

References

[Berners-Lee *et al.*, 2001] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.

[Ding *et al.*, 2004] Li Ding, Tim Finin, and Anupam Joshi. Swoogle: A search and metadata engine for the semantic web. *In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, pages 58–61, Washington DC, November 2004.

[Magkanaraki *et al.*, 2002] Aimilia Magkanaraki, Sofia Alexaki, and Dimitris Plexousakis. Benchmarking RDF schemas for the semantic web. *In Proceedings of the First International Semantic Web Conference*, volume 2348,pages 132–146, June 2002.

[Page *et al.*, 1998] Lawrence Page, Sergey Brin, and Rajeev Motwani. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1998.

[Patel *et al.*, 2003] Chintan Patel, Kaustubh Supekar, and Yugyung Lee. Ontokhoj: A semantic web portal for ontology searching, ranking and classification. *In Proceedings of the Fifth ACM Workshop on Web Information and Data Management*, pages 58–61, New Orleans, Louisiana, November 2003.

[Supekar *et al.*, 2004] Kaustubh Supekar, Chintan Patel, and Yugyung Lee. Characterizing quality of knowledge on semantic web. *In Proceedings of the Seventeenth International FLAIRS Conference*, pages 220–228, Miami, Florida, May 2004.