Knowledge Zone: A Public Repository of Peer-Reviewed Biomedical Ontologies

Kaustubh Supekar, Daniel Rubin, Natasha Noy, Mark Musen

Stanford Medical Informatics, Stanford University School of Medicine

Abstract

Reuse of ontologies is important for achieving better interoperability among health systems and relieving knowledge engineers from the burden of developing ontologies from scratch. Most of the work that aims to facilitate ontology reuse has focused on building ontology libraries that are simple repositories of ontologies or has led to keyword-based search tools that search among ontologies. To our knowledge, there are no operational methodologies that allow users to evaluate ontologies and to compare them in order to choose the most appropriate ontology for their task. In this paper, we present, Knowledge Zone – a Web-based portal that allows users to submit their ontologies, to associate metadata with their ontologies, to search for existing ontologies, to find ontology rankings based on user reviews, to post their own reviews, and to rate reviews

Keywords:

Biomedical Ontologies, Ontology Search, Ontology Ranking, Knowledge Management

Introduction

The medical community has seen widespread application and use of ontologies. With the ever-increasing amount of competing knowledge available for computation, workers in the medical domain who wish to reuse this computational knowledge find it difficult to assess and keep track of all available ontologies. Consequently, more often than not, institutions expend their valuable time and resources to develop their own ontologies, thus creating an assortment of ontologies of varying degree of quality that are not interoperable. In the healthcare system, this translates to the creation of solutions that are not interoperable, thereby hampering the information flow between these solutions, which is critical to the operation and long-term sustainability of the system.

There have been some efforts to develop Web-based libraries of ontologies: Table 1 provides a listing of some of the popular systems. These efforts have served varying purposes, from developing repositories specific to a knowledge representation (DAML Ontology Repository, Protégé-OWL Library) to supporting collaborative development of ontologies (Ontolingua). Most of these resources are mere listings of ontology resources and use different logical and structural organization to present the information relevant to the ontology. A survey [1] of these systems has rightly identified their key inadequacies, and also has suggested important requirements for structuring an ontology library system to enhance ontology management, adaptation, and standardization. Furthermore, with the limited search facility available in these resources, finding the desired ontology requires manual perusal of these Web resources, a cumbersome and a time-consuming process.

Table 1. Ontology Libraries

DAML Ontology Library	http://www.daml.org/ontologies
WebOnto	http://kmi.open.ac.uk/project/Webonto
Ontolingua	http://www.ksl.stanford.edu/software/ontolingua
Protégé-OWL Library	http://protégé.stanford.edu/plugins/owl/owl-library
OntoWeb	http://www.ontoWeb.org
OBO	http://obo.sourceforge.net

Researchers at the University of Maryland have developed Swoogle [2] - a Web based tool that provides a keyword based query facility to search and access ontologies. Through its basic and advanced query interfaces, users can search across a large collection of knowledge resources (1,898,651 Semantic Web documents, ~10,000 ontologies)[1] that exists in the cyberspace. This tool is noteworthy in its ability to search a large collection of ontologies, and to present the user with a list of relevant ontologies that match his query.

We argue that simply listing relevant ontologies is not sufficient for users who need to select the most appropriate ontology for their task. Users need a facility to help them evaluate ontologies and to compare them. Users need to ascertain how well an ontology or a part of an ontology covers the domain of interest, what is the maturity of ontology content, and how the content is related to standard ontologies such as CYC, GO, and UMLS. Most of this information, if available, is provided by the creators of the ontology or by the institution that hosts the ontology, and is usually made available to the public through their Website. For evaluating an ontology, it is also

[1] Statistics obtained from http://swoogle.umbc.edu/ as of November 28, 2006

valuable for a user to find out what were the experiences of other users with that ontology. Unfortunately, most of this important metadata is usually not included as part of the ontology content. This problem can be attributed to the limitations of the underlying knowledge representation language to specify such support information and the lack of tool support to associate metadata with an ontology.

To facilitate ontology reuse, we have developed Knowledge Zone, a resource where users can submit their ontologies, associate metadata with their ontologies and search for existing ontologies that satisfy their requirements. Unlike existing libraries, our system allows users to create and view peer reviews, ratings, experience reports and to find out rankings of an ontology. In Knowledge Zone, we have implemented an Open Rating System model [3] that uses reviews and ratings to compute a Web of Trust, which is then used to compute ranking of an ontology and to filter reviews.

Methods

Ontology Submission: Associating Metadata with an Ontology

The Knowledge Zone portal provides a Web interface (Figure 1) that allows users to associate metadata with their ontologies. We have developed a comprehensive ontology of features (metadata) that characterize an ontology. Some of the categories included in the metadata 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; reliance on other ontologies
- Peer reviews; experience reports; usage data; ratings along different axes, such as correctness; coverage; degree of formality.

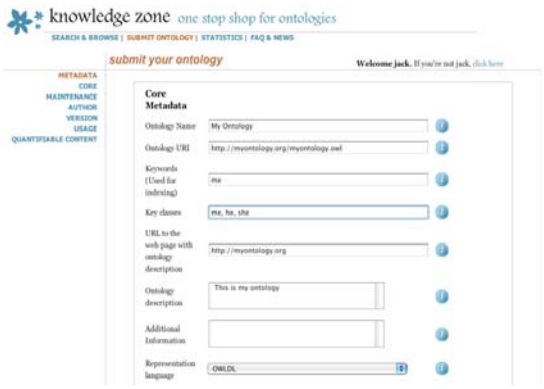

Figure 1. ***Knowledge Zone - Ontology submission interface.*** *Users can enter textual information such as name and URL of the ontology, and specific information about the ontology, such as its representation language.*

These metadata features are principally categorized along two axis (1) source metadata, which include metadata provided by the ontology authors, and (2) third-party metadata, which are provided by the ontology users.

The Knowledge Zone portal employs a dynamically generated UI mechanism that leverages the semantics encoded in the Metadata ontology to provide a contextual interface based on the representation format of the ontology being submitted, level of expertise of the user, type of user (author vs. third-party user), and importance of a particular metadata element. Accordingly, the ontology submission interface is dynamically constructed to generate a Web form that will solicit the user to enter all the metadata information that has been categorized as "source metadata". This process of ontology-driven dynamic UI generation, which is used to generate every page in the Knowledge Zone portal, is depicted in figure 2. Through the dynamically generated ontology submission interface, users can enter, among other things, textual information such as the name of the author and a description of the ontology, as well as controlled information such as intended use, conceptual representation {e.g. Directed Acyclic Graph, Logic based, Frames} and the representation language {e.g. XML, Text, RDF, OWL} can be entered through the drop down boxes, which are populated from the metadata ontology.

The ontology development community in the medical domain comprises of a wide range of researchers, from those who have little or no formal training in logic, to expert knowledge engineers. To accommodate these diverse users, it is imperative to develop a solution that would not overwhelm the naïve user, while at the same time enabling capture of detailed information from the expert user. The dynamic UI generation process employed by Knowledge Zone, and the semantics encoded in our comprehensive metadata ontology allows us provide this flexible behavior in the user interface. For example, based on the axioms in the metadata ontology, an expert user who wishes to annotate description-logic ontology is presented with a different set of metadata information than would be a novice user.

The metadata ontology employed in Knowledge Zone is comprehensive, and has undergone several revisions through a rigorous ontology development process. However, the inclusion and exclusion of metadata features from our ontology could, potentially, evolve; similarly, users might have different opinions about the features attributed to the novice users, controlled terms for a metadata feature, and so on. To accommodate these diverse user opinions and to ensure wide usability of Knowledge Zone, we allow the user to extend the metadata ontology. Currently, these extensions can be made on one or more of the following sections of our ontology: Ontology features (metadata), Controlled Vocabulary, and Tool Tip.

If users extend the metadata ontology, Knowledge Zone can be configured to use their extended ontology (by simply specifying its location). The metadata information that we collect is stored as instances of the metadata ontology.

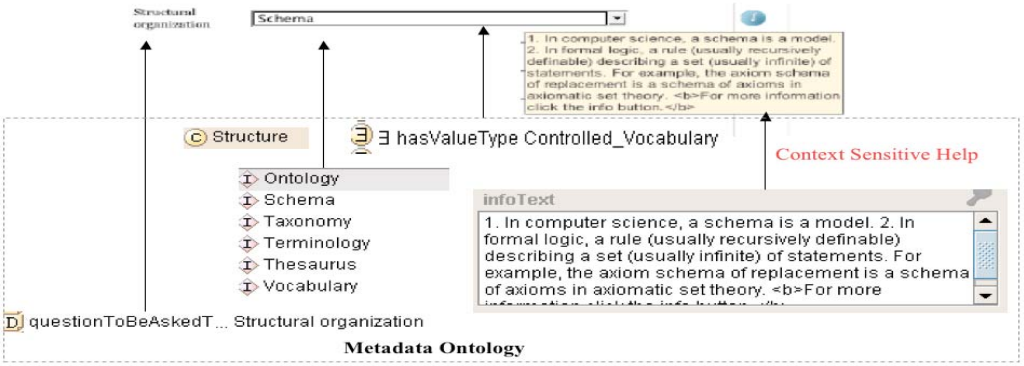

Figure 2. **Dynamic UI Generation in Knowledge Zone**: Process of generating a drop-down form element to capture information about a "Controlled_Vocabulary" metadata element - structural organization of ontology is shown. The label of the form element, tool tip text, and the drop-down box elements, is retrieved from the "questionToBeAskedToAuthor", "infoText", and the "Structure" properties associated with the "Structural_Organization" class in the metadata ontology.

Ontology Search

Most of the current ontology repositories provide a simple keyword search facility. Even though keyword searches are preferred by many users, and work partially for searching Web pages, they are not desirable for searching ontologies, which have an inherent structure. Also, keyword-based searches suffer from poor precision. For example, a search for "anatomy" in Swoogle yields 59 "hits". It is then up to the user to scour through these XML files to look for the ontology that suits his requirements. Also, keyword-based searches are likely to miss out on ontologies that do not have those keywords. For example, the GALEN ontology, one of the popular anatomy ontologies, which, even though crawled by Swoogle, does not appear in the search results when users enter "anatomy" into Swoogle.

Knowledge Zone provides not only a simple keyword-based search interface, but also a structured query interface. For example, through the structured query interface, users can query for anatomy ontologies, which have been encoded in OWL, have been used for data integration, have a GNU license, and have heart, lung, and kidney among its classes. The search engine executes this query by looking up the values of the associated metadata terms across ontologies stored in our repository, and returns all the ontologies that satisfy the search criteria.

Ontology Evaluation

Peer-review of Ontologies

As more ontologies become available, even when users make structured and specific queries, they will obtain multiple "hits". These results will include ontologies that are developed by different institutions, and that are of varying quality. Currently, there are no operational methodologies to compare and evaluate quality of ontologies to select the best ontology for the purpose. We argue that deciding whether some ontology is "best" for a particular purpose is a purely subjective matter.

In Knowledge Zone, users can enter their subjective evaluations and reviews of an ontology. Our idea is akin to that employed by Epinions and Amazon to collect user reviews and ratings on products so that users can subjectively evaluate which "book" or "MP3 player" is best suited for their needs. This peer-review approach is well suited to ontologies, as, like "books" or "mp3 players", ontologies lack universally agreed upon criteria to denote "goodness", and thus users have to rely on other user evaluations and opinions to select a good one.

To evaluate ontologies, it is not only useful to have user evaluations of the ontology itself, but also it is particularly useful if users can comment on particular aspects of an ontology. For example, it is valuable to know what other users think about the "correctness" of a particular ontology. In Knowledge Zone, currently, users can provide numeric ratings and a free text review along the following dimensions: Syntactic Correctness of the ontology, Maturity of the ontology content, Expressibility, Semantic Consistency, Degree of Formality, Availability of documentation, and Usability of the Ontology.

Open Rating Systems and Computing Ontology Rank

A user review of ontologies is a time consuming process and to have its maximum utility it is desirable to keep the review process "open". However, opening the review process to everyone causes the problem of trust. Furthermore, users have varying degree of expertise: biologists are more likely to comment better on the "domain coverage" of the ontology compared to a logician, who is likely to review better on the "semantic correctness" of the ontology. Consequently, we are bound to get some reviews that are of poor quality. To have a "usable" system, it is imperative to be able to filter them.

In Knowledge Zone, we have used the Open Rating System model [3] to accomplish this and to be able to rank ontologies. Users of our system can not only rate the ontology content, but also can rate the reviews as well as the reviewers. We use

these reviews and "trust" statements made by the users to compute a "Web of Trust" (WOT), which is then used to compute rank of ontologies. The Web of Trust model [3] has been successfully used and implemented in systems such as Epinions and Amazon to rank products and to filter reviews.

The Web of Trust model as adapted and implemented in Knowledge Zone has six major components and can be summarized as follows:

- Set of Ontologies O: {O1, O2…On) that are being rated.
- Set of Users U: {U1, U2…Un} that either participate in reviewing the Ontology or providing ratings on other users.
- Set of possible ratings on an Ontology D: {0,1,2,3,4,5}
- Set of possible ratings by a user on another user T: {useful (positive), not-useful (negative)}
- Function that stores the ontology ratings provided by the user R: O x U $\rightarrow$ D
- Function that stores the ratings of users provided by other users: W: U x U $\rightarrow$ T

As mentioned earlier, a structured query by a user might return more than one ontology. We define "ranking" as a problem of being able to rank these ontologies in the descending order of their importance for a particular user. This is equivalent to finding top N ontologies, given all the six major components of our Web of Trust model. Similarly, reviews can be filtered by finding the top N reviews that are relevant to the user.

Given the WOT model; there are many approaches [4] to compute the top N objects. In our case, due to the relatively small size of the user set as compared to that available for systems such as Amazon and Epinions, we envision that functions R and W would be relatively sparse, consequently, most of the users will not have a WOT. To tackle this issue, in Knowledge Zone, we have implemented a TrustRank [5], an approach that is well suited to scenarios where most of the people do not have a WOT. However, TrustRank is limited by its applicability to compute rankings based on reviews and ratings on ontology as a whole. To accommodate for reviews that are collected on specific aspects of an ontology (for example: maturity, correctness) we augmented the TrustRank with a topic-specific trust model. This augmented approach is used to compute ontology rankings and to rank reviews associated with an ontology. Space limitations prohibit us from providing details of this approach; interested reader is encouraged to read the supplementary paper [6].

Results and Implementation

Knowledge Zone is a java-based Web portal that is hosted on an Apache Tomcat server, and is publicly available. The portal uses AJAX, DHTML, and XSLT for dynamic UI generation, and MySQL database to store the user information. User reviews and ontology annotations are stored as instances of the Metadata Ontology in OWL format; Protégé-OWL API is used to access, query, and retrieve and store these instances.

From its inception, Knowledge Zone has consistently attracted a sizeable number of hits (Total hits: 12,456) from the user community, with a large portion of it being unique hits. At the time of this writing, Knowledge Zone is host to a total of twenty eight ontologies, with some notable submissions like BioPAX, Foundational Model of Anatomy, OBO Relationship Ontology and GALEN, and a total of eleven user reviews. Most of the current ontologies in our repository cover different domains. Finding the ontology rank in that case is trivial because a keyword-search or a structured query will at the most return one or two ontologies as the search result. For e.g. a structured query made in Knowledge Zone that represents the user query "Find all the anatomy ontologies" returns two ontologies – the Foundational model of Anatomy and GALEN. The number of ontologies and the number of reviews in Knowledge Zone are a limiting factor in doing a real-world evaluation of our Web of trust model to rank ontologies. However, the Web of trust model, implemented in Knowledge Zone has been successfully tested on non-ontological data, which in terms of our model has similar properties compared to the ontological data.

In future, we envision Knowledge Zone to be host to a sizeable number of ontologies and their reviews. To facilitate this process, we investigated factors that may hamper the growth of Knowledge Zone as a system. From informal interviews with the curators of biomedical ontologies, we postulate that the following may be the rate-limiting factors,

Associating metadata with ontologies is a time consuming process. To tackle this issue we are working on methods to automatically compute the quantifiable metadata. Another approach is to hire curators to enter information about ontologies on behalf of the ontology authors, and get it verified from them. We have used this approach to seed Knowledge Zone with ontologies from the Protégé-OWL library. For these ontologies we obtained metadata information through manual perusal of relevant publications and Web sites. Even though we were able to capture most of the metadata information, such an approach is not scalable, and we would miss out on critical mass of ontologies, which we are not aware-of or are not hosted on one of the existing ontology repositories. We found that there are approximately 500 ontologies in the popular ontology libraries (see Table 1); on the other hand, the number of ontologies crawled by Swoogle is 10K.

Social issues. There are a number of social issues involved, predominantly in the biomedical domain, which hampers information sharing among researchers. These issues are widely studied and solutions involve making the user aware of the benefits of sharing information, ontologies in our case.

Peer-pressure. Currently, the biomedical ontology development community is small and comprises of closely knitted group of researchers. Our approach results in exposing their work (ontologies) for critiquing from colleagues, an approach to which they are receptive to in a private setting such as through email conversations rather than in a public forum. Having anonymous ontology submissions and reviews could

possibly alleviate this concern. However, having large number of anonymous users would need substantial modifications to the current Web of Trust model.

In order to overcome these issues, work is in progress to automatically compute the metadata information associated with an ontology, and in future we plan to disseminate information in the ontology community about incentives of sharing ontologies.

Discussion and Conclusion

We have developed Knowledge Zone – a Web-based portal that allows users to submit their ontologies, associate metadata with their ontologies, to search for existing ontologies, to find out their rankings based on user reviews, to post their own reviews, and to rate reviews. Our hypothesis is that having a substantial number of ontologies and a large number of reviews provided by the user community would support the user in selecting the suitable ontology for his purpose, which, consequently, would facilitate ontology reuse.

The infrastructure developed as part of Knowledge Zone: metadata ontology, ontology search and indexing mechanisms, and peer-review approach for ontology evaluation, is currently being ported for similar purposes in the Bioportal application [7], which is maintained and developed by the National Center for Biomedical Ontology. With the migration of Knowledge Zone functionality and implementation to Bioportal, planned activity by our Center to disseminate information in the ontology community about "incentives" of sharing ontologies, and current efforts to seed the Bioportal with all the ontologies from the Open Biomedical Ontology (OBO) repository, we believe that there would an increase in the number of ontology submissions, and consequently number of reviews collected in, thereby increasing the utility of our system to facilitate ontology reuse and consequently enabling the building of sustainable, interoperable health systems.

Acknowledgments

This work was supported by the National Center for Biomedical Ontology, under roadmap-initiative grant U54 HG004028 from the National Institutes of Health.

References

[1] Ding Y, Fensel D. Ontology library systems: The key to successful ontology reuse. SWWS. (2001)

[2] Ding L, Finin T, Joshi A. Swoogle: A search and metadata engine for the semantic Web. CIKM (2004).

[3] Guha R. Open rating systems. citeseer.ist.psu.edu/694373.html. (2003)

[4] Noy N, Guha R, Musen M.User ratings of ontologies: Who will rate the raters? AAAI Spring Symp (2005)

[5] Haveliwala T. Topic-sensitive pagerank. WWW (2002)

[6] Lewen H, Supekar K, Noy N, Musen M. Topic Specific and Open Rating Systems : An Approach for Ontology Evaluation. EON (2006).

[7] Bioportal.http://www.bioontology.org/resources-bioportal.html

Address for correspondence

ksupekar@stanford.edu