A CONSUMER ONTOLOGY ANALYSIS TOOL

Valerie V. Cross and Anindita Pal
Computer Science and Systems Analysis
Miami University
Oxford, OH 45056
{crossv, pala}@muohio.edu

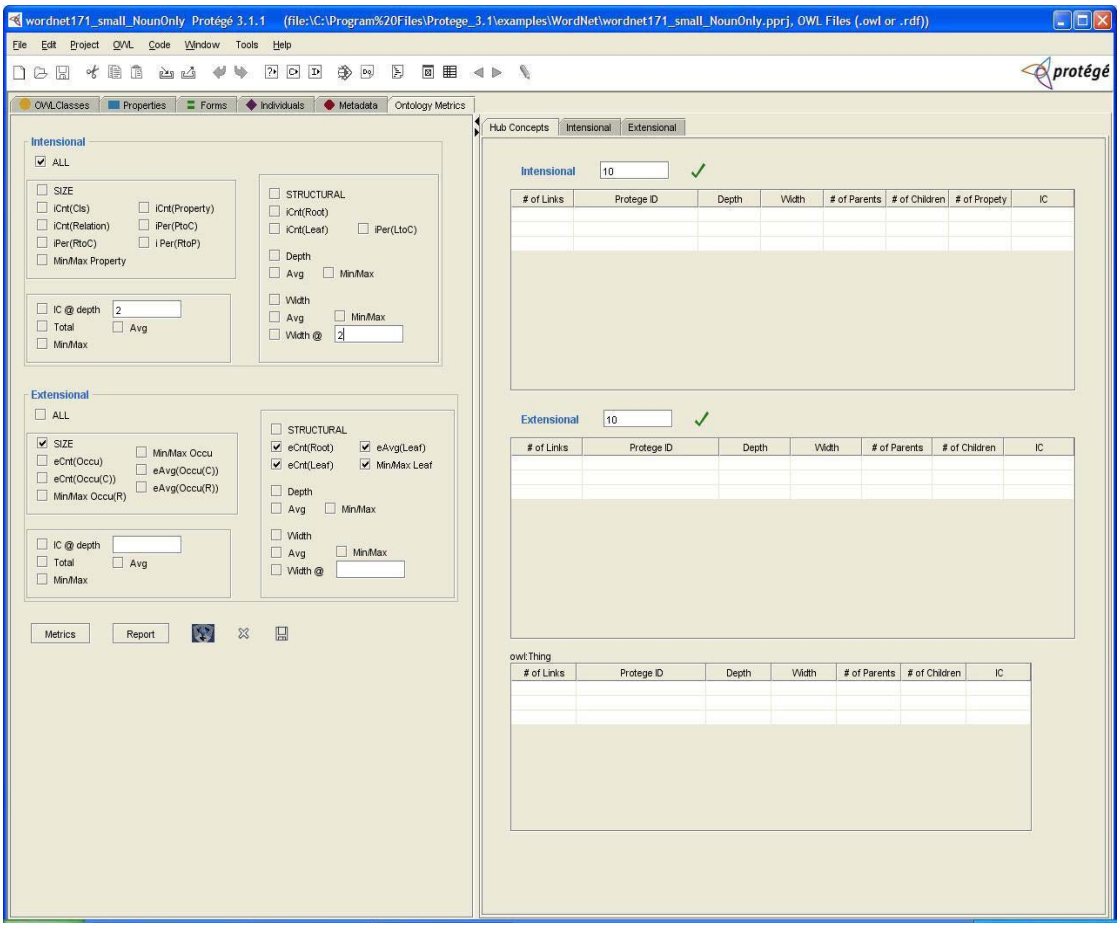
Ontologies capture domain knowledge in a generic way and provide a commonly agreed upon understanding. As such, they are becoming the backbone of the Semantic Web which has as a primary role the sharing of knowledge among Internet users. Numerous ontologies are being developed and published across varying domains such as biomedical, computational linguistics, and business. Development and deployment of extensive ontology-based software solutions represent considerable challenges in terms of the amount of time and effort required to construct the ontology. These challenges can be addressed by the reuse of ontologies. In some domain areas such as e-commerce, recent research indicates that a major concern in the development and maintenance of ontologies for products and services is "the terminological dynamics in markets" [Hepp et al 2005]. Such domains would greatly benefit from research to improve the ontology maintenance process so that the acquisition of dynamic domain vocabulary could be more timely and comprehensive. Another important issue for many Semantic Web applications is the need for interoperability between interacting software agents, each having their own but different ontologies. In this case, the objective is not reuse of an existing ontology but instead one of communication between the two software agents. Establishing a level of semantic interoperability between the two could require both agents to evaluate the quality and usefulness of the other agent's ontology. An ontology analysis tool could assist in the process of ontology evaluation for re-use, facilitate monitoring the development and maintenance of ontologies, and help each agent to determine characteristics of the other agent's ontology for assessing the degree of semantic interoperability.

This presentation describes the research, implementation, and evaluation undertaken in order to develop an ontology analysis tool that can help address issues related to ontology re-use, maintenance and interoperability. By examining the current ontology evaluation research literature and reviewing metrics research in other areas such as conceptual modeling, software development, information systems development, and information retrieval, suites of metrics have been identified and incorporated into an ontology analysis tool. First, an overview of related research discusses the various approaches to ontology evaluation. Ontology evaluation is a broad research area. Different frameworks have been proposed for evaluating how "good an ontology is." Some like One-T [Bouillon et al 2002] consider evaluating the content of the ontology for completeness, consistency and correctness in terms of lack of inconsistencies, redundancy and errors. Others suggest developing methodologies to evaluate an ontology during the development process throughout its entire lifetime [Guarino and Welty 2002]. Furthermore, many argue that the only true way of evaluating ontologies is to use them in applications and assess the application performance. Recent research in ontology evaluation such as AKTiveRank [Alani and Brewster 2005], OntoQA [Tartir 2005], and ONTOMETRIC [Lozano-Tello and Gómez-Pérez 2004] is summarized and compared to the consumer ontology analysis tool.

Another earlier and more detailed study compares different ontologies with respect to 28 characteristics grouped into eight categories [Noy and Hafner 1997]. These approaches provide ontology users valuable information, but consumers, who are the developers of applications using ontology, need additional ways to evaluate ontologies to determine how well the ontology fits their application's needs. Noy [2004] suggests that to make reusing ontologies easier as the number of ontology library increases, more research needs to address the evaluation of an ontology from the consumer point of view. The ontology analysis tool's development was based on the consumer point of view [Noy 2004]; i.e., the consumer is shopping for an ontology to re-use or adapt for a particular project or need. The ontology analysis tool is a Protégé tab plug-in, the most popular among the different categories of Protégé plug-ins. The metrics produced by the tool should provide the consumer with insights about the level of sophistication and the amount of detail provided by the ontology. It allows the user to examine the intensional ontology characteristics separately from those of the extensional ontology. An intensional

ontology only includes the ontology schema or definitions. Thus an intensional ontology consists of a set of concepts (or classes), their definitions and their inter-relationships (specified in properties or slots). An extensional ontology includes the instances of an ontology, where instances are occurrences of a concept. Thus an extensional ontology not only references the concept schemas, but also includes the instances of the concepts.

The metrics are divided into size and structural categories to facilitate user interaction. The implementation is generalized in order to handle the structural difference in ontologies. Thus the design is parameterized so that users can easily switch between an intensional and extensional ontology. For an intensional ontology, the taxonomical structure is built using the sub-class relationship. But the extensional taxonomical structure is more complex and differs from ontology to ontology. The users must select the relationship that will be used to build the taxonomical structure of extensional ontology. Much research has focused on extensional ontologies, in some part, because the consideration for reuse of ontologies has often been on terminological ontologies in the biomedical fields. The complete set of metrics produced by the tool and its user interface are presented. The following figure shows the interface of the plug-in. The interface consists of two split panes; a left "Selection" panel for the various parameters to be selected and a "Result" panel to display the result of the metrics calculated.



Users can select all metrics in the group intensional or extensional by checking the "All" buttons in these groups. Users can also select either size or structural metrics within a group by checking the corresponding buttons. The next set of parameters to input after selection of metrics is the root concept to use in metric calculations and the relationship used to build the extensional taxonomical structure. User can select these parameters by clicking the "Metrics" or "Report" buttons. For the "Metric" button if no class is selected, then only the ontology level metrics are displayed since there is no space in the UI to display the metric results for all classes in the ontology. In case of the "Report" button, if users do not select any class, then the metrics are measured on all classes in the ontology and reported.

A hub summary lists the hub concepts, those with the largest number of links in and out of them. The "Hub Concepts" tab displays the intensional and extensional hub summaries of the ontology. The tab consists of three tables. The intensional table lists the hub concepts, that is, classes with the maximum number of subclasses and super-classes, for entire intensional ontology. The extensional table lists the extensional hub concepts, that is, instances with the maximum number of links in and out of it for the complete extensional ontology. These instances could be from different classes. The Selected Class table lists the extensional hub concepts from the user-specified class. The summary also consists of the following measure for each hub; depth, width, number of properties and information content measures [Seco et al 2004]. By default the plug-in displays the top 10 hub concepts. Users can specify the number or percent of hubs to display. The hub summary may also be visualized by displaying with the paths connecting the hubs. An individual hub may be selected from the hub summary report and a visualization of three levels up and down from the selected hub concept is displayed.

The consumer ontology analysis tool is applied to several different ontologies, primarily terminological ontologies: WordNet, UMLS, UNSPSC, and eCl@ss. WordNet [Miller 1995] is an online English lexical reference system whose extensional ontology organizes nouns, verbs, adjectives and adverbs into synonym sets into semantic network with sets of synonymous terms, or synsets associated with lexical concepts. The synsets differentiate word senses from each other. UMLS (Unified Medical Language System) (http://www.nlm.nih.gov/research/umls/documentation.html) is an ontology that combines many distinct terminologies and was created by the National Library of Medicine (NLM) in Bethesda MD. It helps retrieve information from different biomedicine and health sources. UMLS consists of a large vocabulary database, Metathesaurus. It contains 1.8 million biomedical and health-related concepts, various string concept names, and their relationships. In Metathesaurus, there are more than 100 source vocabularies including different terminologies, classifications, and some thesauri. UNSPSC (United Nations Standard Products and Services Code (UNSPSC) is a hierarchical convention used to classify all products and services. eCl@ss is a Product and Service classification standards developed by leading German companies and offers a standard for information exchange between suppliers and their customers.

The results from applying the ontology analysis tool to the four ontologies are presented and the usefulness of tool and the insights gained for each ontology are discussed. The significance of and recommendations for future directions of this research are also presented.

REFERENCES

H. Alani and C. Brewster, "Ontology Ranking based on the Analysis of Concept Structures", *International Conference On Knowledge Capture*, Alberta, Canada, 2005.

Y. Bouillon, O. Corcho, M. Fernández-López, and A. Gómez-Pérez, "A survey on ontology tools", *OntoWeb: Ontology-based Information Exchange for Knowledge Management and Electronic Commerce*, May 2002.

N. Guarino and C. Welty, "Evaluating ontological decisions with OntoClean", *Communications of the ACM*, Volume 45, Number 2, February 2002

M. Hepp, J. Leukel, and V. Schmitz. "Content Metrics for Products and Services Categorization Standards," *2005 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'05)*, pp. 740-745, 2005.

A. Lozano-Tello, and A. Gómez-Pérez, "ONTOMETRIC: A Method to Choose the Appropriate Ontology," *Journal of Database Management*, 15(2), 1-18, April-June 2004.

G. Miller, "WordNet: a lexical database for English," *Communications of the ACM 38 (11)*, 39 – 41, 1995

N. Noy, "Evaluation by Ontology Consumers," in "Why Evaluate ontology technologies? Because it works!" *Intelligent Systems, IEEE*, Volume 19, Issue 4, Jul-Aug 2004

N. F. Noy, and C. Hafner, "The State of the Art in Ontology Design", *AI Magazine*, Fall 1997, p. 53-74, 1997.

N. Seco, T. Veale, and J. Hayes, "An Intrinsic Information Content Metric for Semantic Similarity in WordNet", *ECAI 2004*, 1089-1090, 2004.

S. Tartir, I.B. Arpinar, M. Moore, A.P. Sheth, A.P. and Aleman-Meza, B. OntoQA: Metric-Based Ontology Quality Analysis *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, Houston, TX, USA, 2005.