A Formal Approach to Evaluating Medical Ontology Systems using Naturalness

Yoo Jung An, Fairleigh Dickinson University, USA
Kuo-Chuan Huang, New Jersey Institute of Technology, USA
Soon Ae Chun, College of Staten Island, USA
James Geller, New Jersey Institute of Technology, USA

ABSTRACT

Ontologies, terminologies and vocabularies are popular repositories for collecting the terms used in a domain. It may be expected that in the future more such ontologies will be created for domain experts. However, there is increasing interest in making the language of experts understandable to casual users. For example, cancer patients often research their cases on the Web. The authors consider the problem of objectively evaluating the quality of ontologies (QoO). This article formalizes the notion of naturalness as a component of QoO and quantitatively measures naturalness for well-known ontologies (UMLS, WordNet, OpenCyc) based on their concepts, IS-A relationships and semantic relationships. To compute numeric values characterizing the naturalness of an ontology, this article defines appropriate metrics. As absolute numbers in such a pursuit are often meaningless, we concentrate on using relative naturalness metrics. That allows us to say that a certain ontology is relatively more natural than another one.

Keywords: IS-A Relationship, Knowledge Sharing, Medical Terminology, Naturalness, OpenCyc, Ontology, Quality of Ontology (QoO), Semantic Relationship, UMLS, WordNet

INTRODUCTION

An ontology represents a data structure that attempts to model human-like knowledge that can be implemented in computational engineering (Gruber, 1993). Concepts (or classes), attributes (or properties) and relationships among concepts that underlie an ontology support the idea of machine-readable data on the Web. The use of ontologies as part of the Web is desirable, since they could support finding answers for users’ queries (Sieg et al., 2007; An et al., 2008). Ontologies may supply generalized terms for a user’s Web search terms. An answer for a query could be derived by using specialization and/or generalization relationships between the concepts of an IS-A hierarchy. For example, a search for Amoxicillin might not give a satis-
factory result, but the user might be satisfied with the search result for the more general term Penicillin. Finding broader or narrower concepts of a given concept is an important technique, which is recommended as a Web search strategy. According to Kalfoglou & Hu (2006), application ontologies are converging with the Web. Thus the knowledge provided by ontologies should be filtered dynamically by understanding the needs of Web users.

There are several well-known ontologies, which many researchers have used and referenced, such as UMLS, WordNet and OpenCyc. Some researchers have presented modified or enriched ontological models by adding new types and trimming some detailed relationships from existing ontologies (Stone et al., 2004). On the other hand, research that investigates these ontologies not only from the viewpoint of experts but also from the perspective of casual users is rare. Assessing difficulties in understanding and using ontologies for emerging user communities on the Semantic Web should be conducted as a stage of implementing the Semantic Web (Finin et al., 2007).

In his original work on ontologies, Gruber (1993) stressed that ontologies are about knowledge sharing. We raise the question whether existing ontologies are constructed so that they may succeed at knowledge sharing. Zeng et al. (2005) showed that communication through terminologies can be significantly facilitated if words labeling concepts are comprehensible to users. Finding concepts which are likely to be recognized by users is a trend in ontology engineering, which is different from the traditional approach of building terminologies understandable mainly by experts of a domain.

We are focusing on an ontology’s role, that is, knowledge sharing supported by an explicit specification of a conceptualization. The key idea of naturalness is based on the need for making terminologies understandable, as described in previous research (An et al., 2006). Some researchers (Staab and Maedche, 2000) have made efforts in making explicit the meaning of some semantic relationships in the form of axioms. However, this declarative knowledge with universal truths about concepts cannot provide answers for all the forms of knowledge inquiries (Mizoguchi, 2004).

It is widely assumed that ontologies represent information in a form that is at least similar to how human knowledge is represented (Smith, 1982). Note that the distinction between primitive and defined concepts (Baneyx et al., 2005) is not employed in this research. It is easy to give precise definitions in mathematically-oriented domains. However, in real world applications this is often not the case.

To many researchers, an ontology concept is a meaningless label, unless it is given a definition. However, any definition itself will contain logical symbols and other labels. Logical symbols do not cause a problem because they are domain independent. However, how can the defining labels themselves be defined? This leads to an infinite regression or circular definitions. Thus, we assume that, at some level, labels have to be understandable by being known to the recipient (program or human). As there are labels that are better known, what we call “more natural,” and labels that are less well known to humans, we prefer to use the more natural labels even for programs. We note that the meaning and the naturalness of concepts are orthogonal factors, as will be explained in more detail below. The author of a document usually has very little choice concerning the meaning he needs to get across. However, he can choose the most natural term for a specific meaning.

In related work about the quality of ontologies, Ram & Park (2004) have focused on semantic interoperability. To enhance the semantics of ontologies, a methodology and analysis have been presented in (Supekar et al., 2004; Brachman, 1992). The ontologies that these researchers presented are domain-specific, so the sizes of the ontologies used are relatively small. In a broader view, some existing ontologies are described in (Noy & Hafner, 1997), however, numeric values and classifications of the ontologies are discussed without being supported by a mathematical model.
Related Content

Hybridizing Artificial Bee Colony Algorithm with Multi-Parent Crossover Operator

Solving Non-Linear Bi-Level Programming Problem Using Taylor Algorithm
www.igi-global.com/chapter/solving-non-linear-bi-level-programming-problem-using-taylor-algorithm/147538?camid=4v1a

A Filter-and-Fan Metaheuristic for the 0-1 Multidimensional Knapsack Problem
www.igi-global.com/article/filter-fan-metaheuristic-multidimensional-knapsack/74738?camid=4v1a
Score Level Fusion of Palmprint, Face and Iris Using Adaptive PSO
www.igi-global.com/article/score-level-fusion-of-palmprint-face-and-iris-using-adaptive-pso/234678?camid=4v1a