Tools for Mapping and Merging Ontologies

Natalya F. Noy

Stanford University, Stanford, CA, 94305, USA noy@smi.stanford.edu

Summary. Researchers in the ontology-design field have developed the content for ontologies in many domain areas. This distributed nature of ontology development has led to a large number of ontologies covering overlapping domains. In order for these ontologies to be reused, they first need to be merged or aligned to one another. We developed a set of tools to support semi-automatic ontology merging: iPrompt is an interactive ontology merging tool that guides the user through the merging process, presenting him with suggestions for next steps and identifying inconsistencies and potential problems. AnchorPrompt uses a graph structure of ontologies to find correlation between concepts and to provide additional information for iPrompt. We present the tools and results of our evaluation of the tools. We discuss other tools and approaches for mapping between ontologies, both in the field of ontology design and database-schema integration.

18.1 Merging And Mapping Between Ontologies

Researchers have pursued development of ontologies—explicit formal specifications of domains of discourse—on the premise that ontologies facilitate knowledge sharing and reuse [23, 14]. Today, ontology development is moving from academic knowledge representation projects to the World-Wide Web. The e-commerce companies use ontologies to share information and to guide customers through their Web sites. The ontologies on the World-Wide Web range from large taxonomies categorizing Web sites (such as on Yahoo!) to categorizations of products for sale and their features (such as on Amazon.com). Many disciplines now develop standardized ontologies that domain experts can use to share and annotate information in their fields. Medicine, for example, has produced large, standardized, structured vocabularies such as SNOMED [29] and the semantic network of the Unified Medical Language System [16]. With this widespread distributed use of ontologies, different parties inevitably develop ontologies with overlapping content. For example, both Yahoo! and the DMOZ Open Directory [24] categorize information available on the Web. The two resulting directories are similar, but also have many differences.

But what does a user do when he finds several ontologies that he would like to use but that do not conform to one another? The user must establish correspon-
dences among the source ontologies, and determine the set of overlapping concepts, concepts that are similar in meaning but have different names or structure, concepts that are unique to each of the sources. Finding these correspondences manually is a onerous and error-prone process. A domain expert who wants to determine a correlation between two ontologies must find all the concepts in the two source ontologies that are similar to one another, determine what the similarities are, and either change the source ontologies to remove the overlaps or record a mapping between the sources for future reference. Therefore, ontology developers need automated or semi-automated help in reconciling disparate ontologies.

This reconciliation must be done regardless of whether the ultimate goal is to create a single coherent ontology that includes the information from all the sources (merging) or if the sources must be made consistent and coherent with one another but kept separately (mapping).

In this chapter, we describe two tools that we developed to aid users in ontology merging and mapping. The first tool, iPROMPT, is an interactive ontology-merging tool that guides users through the merging process and analyzes the structure of the ontologies to suggest points for further merging. The second tool, ANCHORPROMPT uses pairs of related terms from two ontologies, which it gets either from the user or from iPROMPT, and analyzes the structure of the graphs representing source ontologies to produce new pairs of related terms.

We have implemented the iPROMPT and ANCHORPROMPT tools as extensions to the Protégé-2000 ontology-editing environment.1 Protégé-2000 is an open-source tool that allows users to create and edit ontologies and to generate knowledge-acquisition tools based on the ontologies. Protégé-2000 has a large and active user community and many of these users have asked for ability to manage multiple ontologies. The open architecture of Protégé-2000 allows developers to extend it easily with plugins for specific tasks. We implemented PROMPT as a set of such plugins.

### 18.2 The Knowledge Model

Before presenting the tools themselves, we define the frame-based knowledge model underlying PROMPT and Protégé-2000. The knowledge model, the knowledge primitives that we consider, the semantics of relations between them, inevitably affect any strategy for ontology merging and mapping. In describing the knowledge model, we will use an ontology of academic publications as an example. Our example ontology will include such notions as different types of publications, their authors and titles, dates and places of publication, and so on.

We define a **class** as a set of entities. **Instances** of a class are elements of this set (the class of an instance is called its **type**). Classes constitute a taxonomic hierarchy with multiple inheritance. For example, a class **Publication** can represent a set of all publications. Its subclass, a class **Book** represents books, all of which are also publications.

---

1 [http://protege.stanford.edu](http://protege.stanford.edu)