Experience in Aligning Anatomical Ontologies

Songmao Zhang, Chinese Academy of Sciences, China
Olivier Bodenreider, U.S. National Library of Medicine, USA

ABSTRACT

The objective of this article is to recapitulate our experience in aligning large anatomical ontologies (Foundational Model of Anatomy, GALEN, Adult Mouse Anatomical Dictionary and NCI Thesaurus) having different representation formalisms. Our approach to aligning concepts (directly) is automatic, rule-based, and operates at the schema level, generating mostly point-to-point mappings. It uses a combination of lexical, structural and semantic techniques. It also takes advantage of domain-specific knowledge (lexical knowledge from external resources, such as the Unified Medical Language System, as well as knowledge augmentation and inference techniques). In addition to point-to-point mapping of concepts, we present the alignment of relationships and the mapping of concepts group-to-group. We have also successfully tested an indirect alignment through a reference ontology. We present an evaluation of our techniques, both against a gold standard and against a generic schema matching system. The advantages and limitations of our approach are analyzed and discussed throughout the article.

Keywords: anatomy; artificial intelligence; knowledge models; medical knowledge; ontologies; semantic matching

INTRODUCTION

An ontology is a formal representation of a domain modeling the things in that domain and the relationships between those things. Generally speaking, ontologies are composed of concepts (or classes) organized in taxonomies and other hierarchical structures, including partonomies. Moreover, concepts are often connected by various kinds of associative relationships (e.g., spatial, temporal, functional, etc.). In addition to relations to other concepts, concepts can be represented as having properties, often used to differentiate them from other concepts. An ontology, more formally, is “a set of logical axioms designed to account for the intended meaning of a vocabulary” (e.g., Guarino, 1998a). Different ontologies are created to support different tasks, including data integration (e.g., Goble, Stevens, Ng, Bechhofer, Paton, & Baker, 2001), reasoning (e.g., Horrocks & Sattler, 2001) and the semantic annotation of resources in the Semantic Web.
A given domain is often represented by multiple ontologies, providing overlapping, yet different coverage and possibly differing in their representation of the domain knowledge. There is a need for creating mappings among such ontologies in order to facilitate the integration of data annotated with these ontologies and reasoning across ontologies. The goal of ontology alignment is to identify correspondences among entities (i.e., concepts and relationships) across ontologies with overlapping content. Manual alignment of large ontologies is slow, difficult, labor intensive, and error prone. Moreover, it is not suitable for applications in which ontologies need to be aligned on the fly. Semi-automatic and fully automatic approaches to aligning ontologies have been developed instead.

Anatomy is central to the biomedical domain and many anatomical representations have been created over the past fifteen years. While some of them are mere lists of names for anatomical entities (e.g., Terminologica Anatomica), others are full-fledged ontologies, organizing anatomical entities in a rich network of relations. Different knowledge representation formalisms have been used to represent anatomical ontologies, including frame-based structures (e.g., the Foundational Model of Anatomy) and description logics (e.g., GALEN common reference model, SNOMED CT®). While most anatomical ontologies available represent human anatomy, the study of model organisms by biologists (Bard, 2005) has prompted the development of anatomical ontologies for other species (e.g., the Adult Mouse Anatomical Dictionary). Like domain ontologies in general, most anatomical ontologies are developed for a given purpose, for example to support cancer research (NCI Thesaurus) or clinical applications (SNOMED CT). In contrast, some ontologies, called reference ontologies, have been developed independently of specific objectives. For example, the Foundational Model of Anatomy, a reference ontology of structural anatomy (Rosse & Mejino, 2003), could be used as a reference for describing physiology and pathology.

Over the past few years, we have developed domain knowledge-based techniques for aligning large anatomical ontologies, with the objective of exploring approaches to aligning representations of anatomy differing in formalism, structure, and domain coverage. We started by aligning concepts point-to-point in two large ontologies of human anatomy, using lexical and structural techniques (Zhang & Bodenreider, 2003). We later tested these techniques on other pairs of anatomical ontologies, both within and across species (Bodenreider, Hayamizu, Ringwald, de Coronado, & Zhang, 2005; Bodenreider & Zhang, 2006). We also investigated the complex alignment of groups of concepts (Zhang & Bodenreider, 2006a) and that of relationships (Zhang & Bodenreider, 2004a). Finally, we investigated the possibility of deriving the indirect alignment of two ontologies through their direct alignment to a reference ontology (Zhang & Bodenreider, 2005). The objective of this article is to recapitulate our experience in aligning anatomical ontologies and to reflect on some of the issues and challenges encountered along the way. In particular, we want to show the importance of domain-specific knowledge in our alignment strategies.

The article is organized as follows. We first briefly review related work on ontology alignment. Then, we present our experience in aligning anatomical concepts directly, both point-to-point and group-to-group. We follow by the presentation of the alignment of relationships. Finally, we present the indirect alignment techniques we developed. The evaluation of our techniques is presented next, both against a gold standard established manually and against a generic schema matching system. The advantages and limitations of our approaches are analyzed and discussed throughout the article.

BACKGROUND

The general framework of this study is that of ontology aligning, merging, matching, and integration. More than merging or integrating
Related Content

Sarcasm Detection in Twitter Data: A Supervised Approach
[www.igi-global.com/chapter/sarcasm-detection-in-twitter-data/215068?camid=4v1a](www.igi-global.com/chapter/sarcasm-detection-in-twitter-data/215068?camid=4v1a)

Enhancing Folksonomy-Based Content Retrieval with Semantic Web Technology
[www.igi-global.com/chapter/enhancing-folksonomy-based-content-retrieval/64022?camid=4v1a](www.igi-global.com/chapter/enhancing-folksonomy-based-content-retrieval/64022?camid=4v1a)

A Review of Fuzzy Models for the Semantic Web
[www.igi-global.com/chapter/review-fuzzy-models-semantic-web/30384?camid=4v1a](www.igi-global.com/chapter/review-fuzzy-models-semantic-web/30384?camid=4v1a)
Integrating Interactive TV Services and the Web through Semantics
www.igi-global.com/article/integrating-interactive-services-web-through/41960?camid=4v1a