Lister Hill National Center for Biomedical Communications

Medical Informatics Training Program

Lowell Vizenor

FINAL REPORT

09/02/2005 – 08/31/2006

A Formal-Ontological Perspective on the UMLS

The following reports research done at the Lister Hill Center for Biomedical Communications as part of the Medical Informatics Training Program between the period of September 2, 2005 and August 31, 2006. This research also represents a contribution to the Medical Ontology Research group headed by Olivier Bodenreider.

Research Overview

The overall theme of my research has been (and continues to be) the study of how a strict adherence to formal-ontological principles can provide a more coherent and consistent means for the organization of biomedical information. Central to this research has been the Unified Medical Language System (UMLS), which has proved a very useful resource for the testing of ideas. In what follows, I list a number of broad research topics that I have worked on during my stay at the Lister Hill National Center for Biomedical Communications. Some of these research topics have borne fruit in the form of conference and journal papers while others have not. Nevertheless, all of the research has provided a rich background for future research.

Work on the UMLS Semantic Network

My research on the UMLS Semantic Network can be divided into two distinct yet related topics. First, there is the study of the semantic types and the ISA (or subsumption) relations that exist between semantic types. Second, there is the study of what are referred to as associative relations between semantic types. I begin with the former.

One persistent task has been to view the Semantic Network through the lenses of a formal ontology. One reason it useful to make the Semantic Network compliant with basic formal ontology is as follows. It is important for any evaluation of the Semantic Network that the question of the proper breath and depth of the Semantic Network be addressed. In other words, how much of the biomedical domain and healthcare domain should be covered and in how much detail? From an ontological point of view, the primary concern has been to ensure that the high-level categories are broad enough in scope to comprehend the wide spectrum of biomedical and healthcare entities. One way to ensure this is to make the Semantic Network consistent with basic formal ontology.

One way we've tried to accomplish this goal was to exhaustively partition the 135 semantic types into exactly one of three formal-ontological categories: independent continuant, dependent continuant and occurrent. It was apparent from the outset that some semantic types were more amenable to categorization than others. Still, even though there were a number of incongruities between the Semantic Network and the aforementioned ontological categories, the differences were not insurmountable. A couple of remedies suggested themselves right away. One solution, for example,

would be to remove the high-level semantic type conceptual entity and to create a number of new semantic types that would correspond to the ontological categories of function, disposition and role.

My research has also focused on the associative relations between semantic types. One area of particular interest has been to establish direct links between Metathesaurus relationships and Semantic Network relationships. We developed a number of methods for aligning the two UMLS knowledge sources. The goal was two-fold. First, we wanted to improve the usefulness of the vocabulary-specific relationships in the context of the UMLS. Second, we wanted these methods to serve as a first step in identifying and classifying biomedical relations beyond the existing 54 Semantic Network relationships. This research has been important since it supports on-going efforts to use the augmented set of relations as the basic building blocks for a broader and more comprehensive ontology of biomedical relations. Ultimately, this work should go some way toward ensuring that the ontology of biomedical relations will remain relatively stable as new terminologies are added to the UMLS and as changes are made to existing terminologies.

Overview of Papers Submitted to Conferences

Second International Symposium on Semantic Mining in Biomedicine

I submitted (with Olivier Bodenreider) a paper entitled “Using dependence relations in MeSH as a framework for the analysis of disease information in Medline.” Olivier Bodenreider presented the paper in April, 2006 at the *Second International Symposium on Semantic Mining in Biomedicine* in Jena, Germany. The paper was

published in the *Proceedings of the Second International Symposium on Semantic Mining in Biomedicine (SMBM-2006) 2006:76-83*.

There are many types of relationships that exist between biomedical entities. For example, a viral meningitis is located in the meninges and is caused by some virus; it can be treated by some antiviral drugs. Such relations are recorded explicitly as symbolic relations in biomedical knowledge bases and, to a lesser degree, in terminological resources such as SNOMED CT and MeSH. Moreover, the association between indexing terms (i.e., term co-occurrence) in the citations from a bibliographic database such as MEDLINE also represents statistical relations among these indexing terms. For example, the MeSH term Viral meningitis co-occurs frequently with MeSH terms for various virus species, including Enterovirus B, Human and Herpervirus 2, Human. One major difference between symbolic and statistical relations is that, whereas the nature of the symbolic relations is explicit (e.g., location of), the nature of the statistical relations is implicit. However, the frequency of co-occurrence can be analyzed to assess the salience of the association. Formal ontology provides another perspective on relations, distinguishing between two major kinds of symbolic relations: dependence relations (inherent to the nature of related entities) and contingent relations.

The objective of this paper was to analyze dependence relations in MeSH and to compare them to statistical relations obtained from co-occurrence data. We restricted our analysis to the relations between disease categories and other categories of biomedical interest. Our hypothesis was that systematic associations will be found between diseases and the types of entities on which they are dependent, namely between diseases classified by location and their corresponding anatomical sites and

between diseases classified by etiology and their corresponding causes or agents. In practice, for a given disease, the largest proportion of relations to another category should be to a category on which this disease is dependent, and this systematically for each disease. In contrast, we expected to find a smaller proportion of relations between diseases and other categories of biomedical interest, corresponding to contingent relations. Besides clarifying the link between dependence relations and co-occurrence, this paper sought to identify associative relations in MeSH (i.e., relations across trees), which, we expected, would support information retrieval and semantic mining applications.

AMIA Annual Symposium 2006

I submitted (with co-authors: Olivier Bodenreider, Lee Peters, Alexa T. McCray) a paper title "Enhancing biomedical ontologies through alignment of semantic relationships: Exploratory approaches" to the *AMIA Annual Symposium 2006*. The paper was accepted and will be published in the *Proceedings of AMIA Annual Symposium 2006*.

This paper investigated several methods for aligning Metathesaurus relationships with their counterparts in the UMLS Semantic Network. Unlike the categorization link defined between Metathesaurus concepts and Semantic Network types, no such correspondence exists between the relationships at these two levels of the UMLS. Methods: The first approach attempted to elicit the semantics of Metathesaurus relationships through an examination of their relata at different levels: concept, high-level ancestors and semantic types. The second approach examines the frequency of association between a given Semantic Network relationship and the actual relationships observed in the Metathesaurus between the concepts categorized by

these semantic types. Results: A total of 139 relationships are present in the Metathesaurus. Using the methods described in this paper, 80 (58%) could be aligned with Semantic Network relationships. The remaining relationships are vocabulary internal, used, for example, for vocabulary management or to indicate strictly lexical relationships. The work reported here is a first step in the attempt to build a more comprehensive ontology of biomedical relationships.

These methods are meant to be seen as complementary. In the ideal case, all the methods would point to a single semantic interpretation of a given relationship. In practice, some methods work better than others for some cases. The hope was that, in combination, these methods would provide, first, a comprehensive strategy for guiding the alignment of Metathesaurus relationships and Semantic Network relationships, and, second, would serve as a good starting point for the development of a comprehensive ontology of biomedical relationships.

It should be noted that this paper was primarily a "methods" paper. We are currently working on a subsequent paper that will deal with the results of these methods.

FOIS 2006: International Conference on Formal Ontology in Information Systems

I submitted a paper with Eric Little titled "Principles for the Development of Upper Ontologies in Higher-level Information Fusion Applications" to *FOIS 2006*. The paper will be published in the *Proceedings of FOIS 2006.*

In this paper we analyzed the types of complex relations typically involved in situation and threat assessment (STA) ontology and discussed how these distinctions can serve as a theoretical framework for the enhancement of existing ontology development tools, especially as these relate to STA ontology. The use of ontologies

within many information science communities is growing at an ever-increasing rate. This is especially true of higher-level multisource information fusion applications, where there is a genuine need for an improved understanding of the complex relational items (e.g. intentions, capabilities, opportunities and vulnerabilities) typically associated with situation and threat assessment. Still, most current ontology development tools lack the resources to support the sorts of ontological distinctions necessary to provide a sufficiently scalable and reusable ontology product for STA purposes. Although multisource information fusion originated in defense research it is by no means limited to it.

Recently, it has been extended to a variety of non-military applications such as robotics, transportation, remote sensing, optical character recognition, medical decision-making, and crisis management. So, the general approach discussed here can be extended beyond ontology development for higher level information fusion related to situation and threat assessment as described in. One such area is healthcare, where decision-making is a key component to providing services associated with medical diagnosis, treatment, emergency service monitoring, and so on.

Papers in Progress

The following two papers are intended to be submitted to journals and the work is still in progress. The overall goal of both papers is to develop an ontology of biomedical and healthcare relations that exploits the preexisting knowledge sources in the UMLS.

JAMIA Paper

Currently I am working with Olivier Bodenreider and Alexa T. McCray on a paper entitled “An Ontology of Relationships for Healthcare and the Life Sciences”. This paper is an outgrowth of the AMIA paper discussed above. The primary goal of this paper is to apply the methods for aligning Metathesaurus relationships with their counterparts in the UMLS Semantic Network from our AMIA paper. The major portion of the manual alignment was completed in early June during a two day workshop at the Harvard School of Medicine.

What remains is to assess the results of the automated methods and then to align the various methods with one another. This paper will bring together a number of research topics that I’ve worked on during my stay at the NLM. It will discuss the benefits of providing a standardized way of expressing knowledge – i.e. reusability. Advance the field of ontology alignment by focusing on aligning relationships and not concepts. Discuss how this work could contribute to the semantic web, especially as this relates to the life sciences. It should potentially add depth to and enrich the UMLS and improve the domain and range of relations.

Monist Paper

This paper will cull the work done on assessing the Semantic Network from a formal-ontological point of view. The basic idea has been to assign each semantic type to exactly one of the following formal ontological categories of independent continuant, dependent continuant and occurrent. The goal then is to restrict the domain and range of a certain number of associative relations to one of these formal ontological categories. This idealized version of the semantic network would then serve as a foil against which to test the Semantic Network. It will also discuss the limitations of applying strict adherence to formal-ontological principles can provide a more

coherent and consistent means for the organization of biomedical information. One reason for this is that the Semantic Network is a “lightweight ontology,” since it lacks the resources to structure information in a format that would allow us to express many of these subtle ontological distinctions.

Service to the Profession

My time at the LHNCBC has provided me with ample opportunities to meet interesting people in the field of medical informatics and ontology. In addition I have been given opportunity to participate in a number of activities that have especially enriched my understanding of these fields.

KR-MED 2006 – Biomedical Ontology in Action

I served on the scientific program committee of KR-MED 2006, which is an international workshop that brings together researchers from a broad range of fields that are related to formal ontology and medical informatics. The goal of this year’s workshop is to show how current research can be brought to bear on the practical problems associated with the development of applications supported by these ontologies, i.e., to show biomedical ontology "in action".

Journal of Applied Ontology

I also served as a reviewer for several papers for the Journal of Applied Ontology.