ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

An Ontology for Gastrointestinal Endoscopy

Shahim Essaid

Oregon Health & Science University, Portland, OR, USA

Abstract. The field of gastrointestinal endoscopy can benefit from an ontology for the purposes of data coding and data integration. This paper presents early results of an effort to develop such an ontology based on the OBO Foundry principles and existing OBO Foundry ontologies. Initially, the ontology will be limited to representing entities and relations currently implicit or hard-coded in the user interface of an existing endoscopy reporting system. The ontology will also be mapped to an existing database to evaluate the feasibility of ontology-driven queries. The long-term goal is to evolve this ontology to an application-independent terminology and information model for the domain of gastrointestinal endoscopy.

Keywords: gastrointestinal endoscopy, knowledge representation, data integration

1 Introduction

The practice of gastrointestinal endoscopy produces a significant amount of structured data that is captured in endoscopy reports. To encourage consistency in data collection, the World Endoscopy Organization maintains the Minimal Standard Terminology for gastrointestinal endoscopy (MST) [1]. The MST specifies a minimal set of terms and data structures needed to encode the majority of endoscopy data. However, the domain knowledge represented in the MST (in the form of terms, relations, and data structures) is not in a computable format and it could benefit from an ontological and logical analysis, and reorganization.

Another important effort in the field of gastrointestinal endoscopy is the Clinical Outcomes Research Initiative (CORI) [2]. CORI was established to assess the utilization and effectiveness of endoscopy procedures in clinical care. To meet its goals, CORI has developed an endoscopy reporting software and a central data repository for endoscopy reports. The reporting software is being used nationwide and the data repository currently receives over 250,000 reports annually. The data repository is primarily used for research purposes and to report on practice patterns and clinical outcome measures.

The endoscopy reporting software developed by CORI was initially based on the MST's representation of endoscopy data but it evolved to include additional terminology and data elements. Also, efforts are in place to integrate date generated by commercial reporting systems into the CORI data repository. These efforts have highlighted the need for a shared, stable, and computable terminology and information model for the field of gastrointestinal endoscopy to facilitate data integration while maintaining clear and consistent semantics.

2 Motivation and Planned Development

Recent advances in the area of biomedical ontology [3-4] can provide a foundation for a more formal and logical representation of entities and data elements needed to represent endoscopy data. Also, the existence of standardized knowledge representation languages and related inferencing capabilities can enable sophisticated querying of logically represented data and knowledge [5]. These advances have motivated an effort within the CORI project to develop an ontology for the field of gastrointestinal endoscopy.

The ontology will follow the Open Biomedical Ontology (OBO) Foundry development principles [6] and reuse entities from existing OBO ontologies when appropriate. The BFO will serve as an upper level ontology and other ontologies (IAO, ogms, OBI, etc.) will be examined for middle level entities. Domain level entities will reference existing ontologies of anatomy, pathology, phenotypes, relations, and others when available. The ontology development project is hosted as a Google Code project [7].

3 Methods and Expected Difficulties

Development will start by identifying domain level terms and data elements hard-coded in the user interface of the existing CORI reporting software. These entities will initially form the main content of the ontology. This will decouple the domain knowledge from the application and allow for a more flexible evolution of the terminology and information model of the reporting software while still maintaining ontological and formal knowledge representation principles. The ontology will then be augmented by other entities from the MST, from free text entries in existing endoscopy reports, and from the endoscopy community. Also, as a proof of concept, the ontology will be mapped to the existing CORI data repository to evaluate the feasibility and benefit of ontology-driven queries compared to native SQL queries. The D2RQ Platform will be used for this part of the project [8].

A brief exploration of the user interface for the reporting software, and the MST, showed that difficult issues such epistemology vs. ontology, entities vs. statements, negation, and other related issues are frequent in clinical settings. Also, despite the relatively narrow focus of the practice of gastrointestinal endoscopy, endoscopy reports include information that ranges from current and past medical history, physical examination, visible endoscopy findings, and indirect findings through various imaging techniques. In addition to these various types of information, there is another epistemic layer that reflects the attitudes and judgment of clinicians in the form of assessments, diagnoses, plans, etc. To fully represent this information, an ontology will need a rich set of relationships that cover mereotopological, temporal, and modal relations, among others.

However, the primary use cases described below can be met by limiting our initial development efforts to an is_a hierarchy and a basic set of qualitative mereotoplogical relationships. The initial version will also be limited to representing endoscopic findings (polyps, ulcers, foreign bodies, etc.), their anatomical locations, and their clinical descriptions, according to the OBO Foundry ontology development principles.

4 Primary Use Case

The ontology will primarily serve as an interface terminology that supports data entry and enables consistent coding of endoscopy reports. The ontology will also be used to explore the value of ontology driven data retrieval by executing ontology driven queries against the current CORI dataset. These initial use cases can be met by a limited set of entities and relations and the remaining domain knowledge will be added as need arises.

References

1. Minimal Standard Terminology, http://www.worldendo.org/mst.html
2. Clinical Outcome Research Initiative, http://www.cori.org/
3. Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. Brief Bioinform. 7(3):256–274 (2006)
4. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. 25(11):1251-1255 (2007)
5. Stevens R, Aranguren ME, Wolstencroft K, Sattler U, Drummond N, Horridge M, et al. Using OWL to model biological knowledge. International Journal of Human-Computer Studies. 65(7):583-594 (2007)
6. OBO Foundry Principles, http://www.obofoundry.org/crit.shtml
7. Gastrointestinal Endoscopy Ontology, http://gieo.googlecode.com/
8. The D2RQ Platform – Treating Non-RDF Databases as Virtual RDF Graphs, http://www4.wiwiss.fu-berlin.de/bizer/d2rq/

ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

Enriching the Ontology for Biomedical Investigations (OBI) to Improve its Suitability for Web Service Annotations

Chaitanya Guttula[1], Alok Dhamanaskar[1], Rui Wang[1], John A. Miller[1,3], Jessica C. Kissinger[1,2,3], Jie Zheng[4], Christian J. Stoeckert, Jr.[4]

[1]Department of Computer Science, [2]Department of Genetics, [3]Institute of Bioinformatics, University of Georgia, Athens, GA, USA
[4]Penn Center for Bioinformatics, Department of Genetics University of Pennsylvania, Philadelphia, PA, USA

Abstract. With the increasing development and use of ontologies in the biomedical domain, opportunities for their utilization in applications and workflows are being created. In this paper, we discuss how the Ontology for Biomedical Investigations (OBI) can be enriched to support annotation of Web services. The methodology includes designing ontology analysis diagrams for Web services and analyzing them to find the terms that need to be added to the ontology. The enriched ontology can then be used for annotating the Web services with the help of annotation tools like the one in the RadiantWeb tool-suite. Using annotated Web services to perform service discovery and make service suggestions provides a way to evaluate the validity of the annotations made and the terms added.

Keywords: ontology, OBI, biomedical, Web services, semantic annotations.

1 Introduction

In recent years, the number of tools and software applications available as Web services in the biomedical community has increased dramatically. Complex real world tasks generally require coordinated use of multiple Web services. It is a challenge to find those Web services that suit the users' needs or work effectively in Web service compositions. Semantic annotations of the Web services would facilitate Web service discovery and composition [1]. A Web service may be described using the Web Service Description Language (WSDL), which specifies the set of operations provided by the Web service, as well as details about these operations, including their inputs and outputs. Standardized annotations of a Web service include the semantics for the input, output and functionality of each of the service's operations. Bioinformatics Web services are used to analyze biomedical data and hence, need relevant terms for their annotation.

It is preferable to use an ontology that is compatible with other biomedical ontologies. Open Biological and Biomedical Ontologies [2] (OBO) compliant ontologies are interoperable with each other, because they share a common upper level ontology, the Basic Formal Ontology (BFO) [3], and a common set of relations, the Relation Ontology (RO) [4]. The Ontology for Biomedical Investigations (OBI) [5], a member of the OBO Library, is being developed to address the need for consistent description of biological and clinical investigations. OBI is a process oriented ontology that models a process with input, output and objective specifications and is suitable for supporting Web service annotations. This paper reports on our efforts to enrich OBI for the purpose of semantically annotating Web services to enhance its usability.

2 Enrichment of OBI to Support Service Annotations

Ontologies used for annotations should provide terms that correspond to key aspects of a Web service description. If the required terms are not available in OBI, we add them. However, terms are reused where possible. We begin the process of enriching OBI by creating a generic model for Web services and further refine it to model specific types of Web services. Constructing a generic model involves creating an ontology analysis diagram, which shows relationships between different top level terms for Web services, including the objective of a Web service and its operations.

A sample generic model can be viewed at: http://mango.ctegd.uga.edu/jkissingLab/SWS/Ws annotation/resources/GenericCmap.jpg

Modeling a Web service at a more specific level requires a detailed analysis of the Web service's operations in terms of its inputs, outputs and objective specification. The outcome of this can be seen for example in the ontology analysis diagram for ClustalW, available at: http://mango.ctegd.uga.edu/jkissingLab/SWS/Wsannotation/resources/clustalCmap.jpg.

We are modeling several Web services, including ClustalW and BLAST Web services currently available at the European Bioinformatics Institute (EBI). The ClustalW Web service was studied and its inputs and outputs were summarized in a spreadsheet along with their definitions. We finally determine the terms and the possible positions where they can be added to the ontology on the basis of the above-mentioned ontology analysis diagram. Once a term is fully described by specifying its set of restrictions (e.g., objective specification for Web service operations), we come up with a logical definition for the term.

Using Protégé, we have added the new terms in the OWL file that is available at: obi.svn.sourceforge.net/svnroot/obi/trunk/src/ontology/branches/webService.owl (Figure 1). A description logic reasoner (e.g., HermiT) is used to check for consistency of the added terms, as well as to infer the correct placements of the terms in the ontology's hierarchy. A request has been sent to the OBI issue tracker and the terms are currently pending approval.

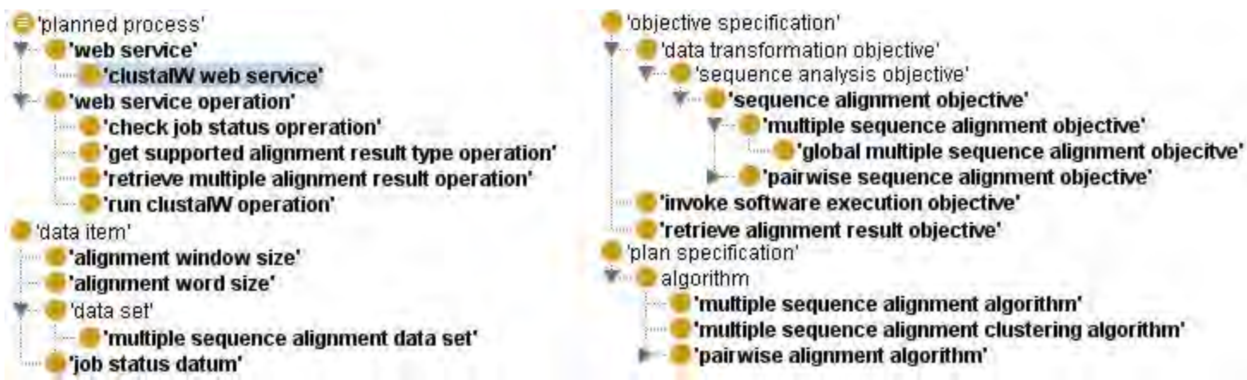

Figure 1. Ontology Hierarchy of terms added to OBI used for clustalW annotation, the terms in bold are the newly added terms

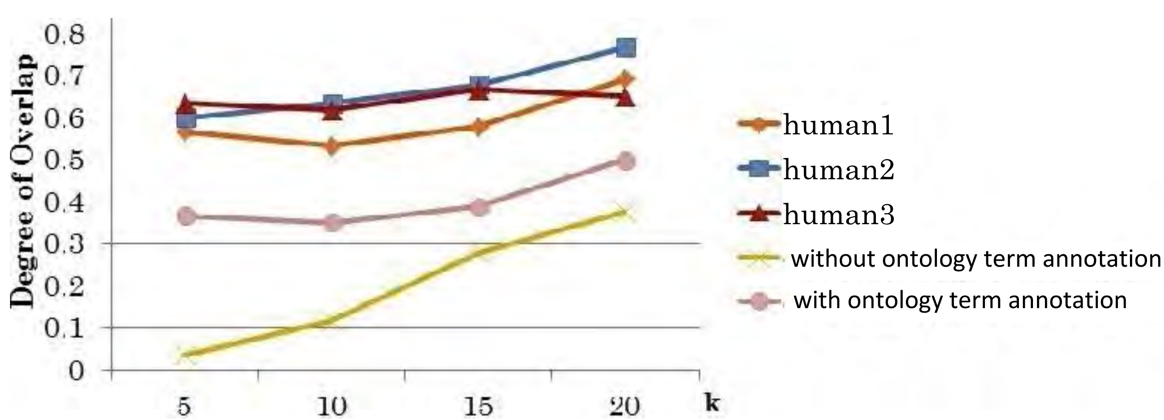

Figure 2. Comparing Annotation Cases.

3 Evaluations of Web Service Annotations Using OBI

Considering the increasing number of available Web services in Biomedical domain, manual annotation and composition of Web services is a tedious task. The RadiantWeb tool-suite is a Web application with a simple drag-and-drop user interface that includes tools for annotation, discovery and suggestion of Web services. The purpose of annotations is to provide formalized documentation that can be read by humans and processed by machines.

To illustrate the use of annotations, we considered a common scenario encountered by biologists, that of discovering more information about a particular protein sequence and its evolutionary relationship to other protein sequences. In order to find this information, we had to design a workflow consisting of multiple Web service operations. The workflow mainly utilized popular bioinformatics programs such as BLAST and ClustalW. The RadiantWeb Tool Suite was used to ease the process of providing annotations and creating the workflow. Figure 2 depicts the effectiveness of annotations, which confirms that the annotated Web services perform better for service discovery and suggestions than unannotated ones. A more detailed evaluation of effectiveness of annotations can be found in [7].

4 Discussion: Related Work & Conclusions

To the best of our knowledge the only other major effort in the biomedical domain focused on creating or enriching ontologies for the purpose of semantically annotating Web services is the EMBRACE Data and Methods (EDAM) ontology [6]. EDAM covers several but not all of the terms required for the annotation of Web services in this domain. For example, for BLAST Web Services missing terms include 'low complexity sequence filter', 'number of top combinations', 'pairwise alignment sensitivity'. Also, EDAM is a work in progress with several of its properties having ranges/restrictions specified as undefined, in addition to it not being OBO compliant, meaning it has lesser compatibility with other biomedical ontologies that are OBO compliant.

In this paper, we apply a systematic methodology for enriching existing biomedical ontologies (OBI in our case), so that they can support semantic annotation of Web services in this domain. We have enriched OBI with terms required for the annotation of BLAST & ClustalW and will continue working on other Web services. Tools such as RadiantWeb can be used to make the process of annotating Web services quick and easy. Our preliminary evaluation made it clear that discovery and service suggestions with annotated Web services yield better results than with unannotated Web services [7].

Acknowledgements

Funding for this study was provided by NIH R01 GM093132.

References

1. Phillip Lord, Sean Bechhofer, et al.: Applying Semantic Web Services to Bioinformatics. LCNS, vol. 3298, pp. 350-364 (2004).
2. Barry Smith, Michael Ashburner, et al.: The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration. Nature Biotechnology vol. 25, pp. 1251-1255 (2007).
3. http://www.ifomis.org/bfo Accessed Apr 30, 2011.
4. http://www.obofoundry.org/ro/ Accessed Apr 30, 2011.
5. Brinkman RR, Courtot M. et al.: Modeling Biomedical Experimental Processes with OBI. Journal of Biomedical Semantics, Suppl 1:S7 (Jun 2010).
6. http://edamontology.sourceforge.net/ Accessed Apr 30, 2011.
7. Rui Wang, Chaitanya Guttula, Maryam Panahiazar, Haseeb Yousaf, John A. Miller, Eileen T. Kraemer and Jessica C. Kissinger: Web Service Composition using Service Suggestions. In: Proceedings of the 2011 IEEE International Workshop on Formal Methods in Services and Cloud Computing, Washington, DC (July 2011).