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], and Christian J. Stoeckert, Jr.[4]

[1]Dept. of Computer Science, [2]Dept. of Genetics, [3]Institute of Bioinformatics
University of Georgia, Athens, GA 30602
{guttula@cs., alok@cs., wang@cs., jam@cs., jkissing@}uga.edu
[4]Penn Center for Bioinformatics, Dept. of Genetics
University of Pennsylvania, Philadelphia, PA 19104
{jiezheng, stoeckrt}@pcbi.upenn.edu

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 provide for 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 `mango.ctegd.uga.edu/jkissingLab/SWS/Wsannotation/resources/GenericCmap.jpg`.

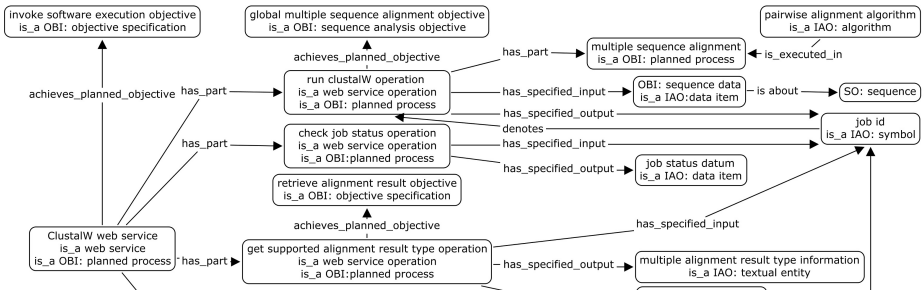

Figure 1: Part of the Ontology Analysis Diagram for ClustalW. The Complete Version Is Available at: `mango.ctegd.uga.edu/jkissingLab/SWS/Wsannotation/resources/clustalCmap.jpg`

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 in Figure 1 which is a part of ontology analysis diagram for ClustalW. 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. Based on the ontology analysis diagram (Figure 1), we finally determine

the terms and the possible positions where they can be added to the ontology. 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.

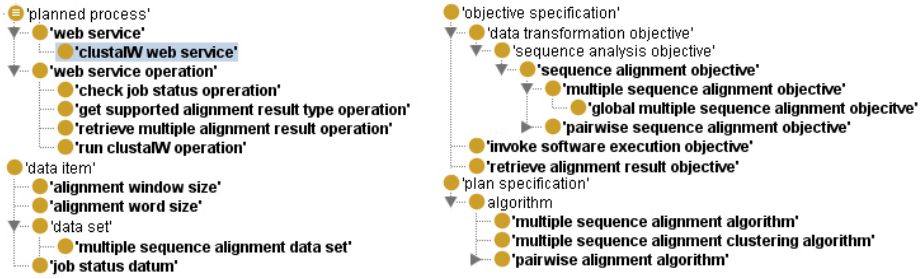

Figure 2: Ontology Hierarchy of terms added to OBI used for clustalW annotation, the terms in bold are the newly added terms

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 2). 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.

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.

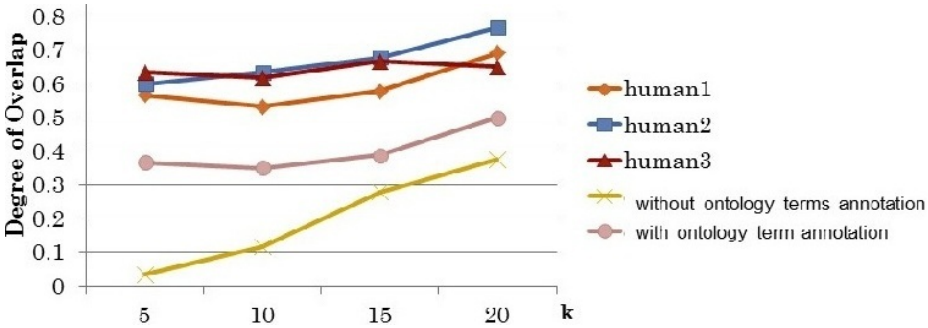

Figure 3: Comparing Annotation Cases.

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 3 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 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).