Semantic Web Services for N-glycosylation Process

Satya S. Sahoo, Amit P. Sheth, William S. York, John A. Miller
LSDIS Lab and Complex Carbohydrate Research Center,
University of Georgia, Athens, GA

Abstract:

Glycomics is one of the many research efforts currently underway in the biosciences domain, which is characterized by high throughput data generated at multiple experimental stages. For example, analysis of N-glycosylation encompasses stages from cell-culture to peptide identification and quantification. Research groups across the world use diverse cell cultures, separation and spectroscopic techniques, and data identification, correlation and integration methodologies. Thus, data generated at different phases of the process by multiple groups are both structurally and functionally heterogeneous. Automatic semantic annotation of such experimental data with concepts defined in domain ontology can provide detailed process information, including the experimental environment that is critical for the comparison and analysis of such data, thus increasing the opportunity for rapid knowledge discovery. Semantic annotation of scientific data not only allow standard interpretation, but by taking advantage of the rich relationships among concepts in the ontology it is possible to derive mapping, inferences and correlations that may be too obscure for analysis and discovery by humans. Web service technology is a natural enabler for synergistic usage of computational tools developed at different labs using heterogeneous data.

As part of the Bioinformatics for Glycan Expression core of the NCRR Integrated Resource for Biomedical Glycomics, we have implemented an N-glycosylation process (**NGP**), a *Semantic Web Process* which uses Web services to expose computational tools for different phases and support interoperability. NGP uses two ontologies: **PropreO**, which captures knowledge of processes, resources, techniques and instruments, and **GlycO**, which captures glycomics domain semantics populated from knowledge from multiple sources. Ontologies are very deep (e.g., GlycO currently has 767 classes, is 11 level deep a place, and make extensive use of the OWL constraints) and are populate from multiple public knowledge sources. These ontologies are being used to annotate data with different formats, of different types (text, image as well as raw data) generated at different phases of NGP. The Web services, implemented for different phases of the N-glycosylation, are natural candidates, by functionality and context, for annotations by concepts from **PropreO**. These semantic Web services pave a path for semantically mediated service discovery, invocation and integration spanning multi-phase processes in Glycomics research in a 'meaningful' manner. The presentation will include a demonstration of the initial implementation of NGP using the METEOR-S system that supports full Semantic Web Process lifecycle (definition, annotation, publication, discovery, and composition). All resources and software are open-source and available form the LSDIS lab web site.

Short Biography: Satya S. Sahoo is a PhD student and a research assistant on the Bioinformatics for Glycan Expression project at the University of Georgia (UGA). Prof. [Amit Sheth](#) is the director of the Large Scale Distributed Information Systems (LSDIS) lab and a professor of Computer Science at UGA and co-founder/CTO of Semagix Inc. Prof. William York is an Associate professor of Biochemistry and Molecular Biology with the Complex Carbohydrate Research Center (CCRC) at UGA. Prof. John Miller is a professor of Computer Science at UGA.

Contact Information:
Prof. Amit Sheth, amit@cs.uga.edu, 706-254-3901 (m), 706-542-4771 (f).

A/V Requirements: LCD projector (optional wired or wireless internet access for live demo from our laptop)