Semantically Annotated RESTful Services for Large-scale Metabolomics Data Analysis

Ashwin Manjunatha  
*Wright State University - Main Campus*

Paul E. Anderson  
*Wright State University - Main Campus*

Satya S. Sahoo  
*Wright State University - Main Campus*

Ajith H. Ranabahu  
*Wright State University - Main Campus*

Michael L. Raymer  
*Wright State University - Main Campus*, michael.raymer@wright.edu

See next page for additional authors

Follow this and additional works at: [http://corescholar.libraries.wright.edu/knoesis](http://corescholar.libraries.wright.edu/knoesis)

Part of the Bioinformatics Commons, Communication Technology and New Media Commons, Databases and Information Systems Commons, OS and Networks Commons, and the Science and Technology Studies Commons

Repository Citation

[http://corescholar.libraries.wright.edu/knoesis/96](http://corescholar.libraries.wright.edu/knoesis/96)
Authors
Ashwin Manjunatha, Paul E. Anderson, Satya S. Sahoo, Ajith H. Ranabahu, Michael L. Raymer, and Amit P. Sheth

This presentation is available at CORE Scholar: http://corescholar.libraries.wright.edu/knoesis/96
Semantically Annotated RESTful Services for Large-scale Metabolomics Data Analysis

Ashwin Manjunatha, Paul Anderson, Satya Sahoo, Ajith Ranabahu, Michael Raymer, Amit Sheth
Kno.e.sis Center, Wright State University
{ashwin,paul,satya,ajith,raymer,amit}@knoesis.org

1. Introduction

Metabolomics:
The term metabolomics is defined as a comprehensive analysis in which metabolites of a biological system are identified and quantified. Any technique that can quantify metabolites can be used for metabolomics, but there are two primary techniques seen in the literature; nuclear magnetic resonance (NMR) and mass spectrometry with a prior on-line separation step such as high performance liquid chromatography (HPLC) or gas chromatography (GC).

NMR:
Nuclear magnetic resonance (NMR) spectroscopy is an experimental technique that exploits the properties of an atom’s nucleus. It can be used to obtain information about the concentration and structure of molecules. NMR studies magnetic nuclei by applying a static magnetic field followed by applying a second oscillating magnetic field. Specifically, only nuclei with an odd number of protons or neutrons can be measured using NMR. However, the two most common atoms studied are 1H and 13C.

NMR spectrometer:
Toxicology is the branch of pharmacology that deals with poisons and their effects on plant, animal and human life.

2. What is the problem?

- Large data sets
- Standard post-instrumental processing
- Quantification of spectral features
- Normalization
- Scaling
- Multivariate statistical modeling
- All computationally intensive processes
- Variety of algorithms for each step

Need a robust and flexible analysis platform

A common solution for flexibility

Move to a Service based Architecture!
- Provide Web Services for each algorithm
- Assemble workflows as required!

Taverna – an open source family of tools for designing and executing workflows

http://www.taverna.org.uk/

3. How about Scalability?

Computing Cloud
Sharing hardware resources, software and information are provided to computers and other devices on demand.
- Many vendors

Hadoop
An open-source software framework for reliable, scalable, distributed computing.

[http://hadoop.apache.org]
- Uses the map-reduce computational paradigm
- Runs on Computing Clouds

Use Apache Hadoop on Computing Clouds to run processes in parallel. Applicable to many common mathematical operations such as summing and averaging.

4. Annotation and SA-REST

SA-REST
W3C submission on Semantic Annotation of RESTful services [3].

Three basic properties
- domain-ref: mark the top level domain of a document e.g. Nucleotides
- sem-ref: mark the domain of a linked document
- sem-class: mark the meaning of a selected word

Adding metadata to point to richer models

Ontology

5. Advantages of Annotation

Faceted Search

Technique for accessing a collection of information represented using a faceted classification, allowing users to explore by filtering available information.

When annotated with richer models, the indexing software can easily create faceted indexes to support a fine grained search. Even the regular keyword search can be improved.

1. Query by concept – not by keyword

Search for “NCI:FASTA” instead just FASTA. Yields documents that indicate the term FASTA as defined by the NCI Thesaurus.

2. Filter by multiple facets

Issue queries indicating many facets, say “type: soap-binding:java include:NCI:FASTA” to look for service descriptions that are SOAP services with Java bindings including mentions about NCI:FASTA.

Semi-Automated Composition

When service interface documents are annotated service compositions can be done more intelligently.

1. A composition tool can warn the creator of incomparable connections: Output of Service A cannot be input to Service B!

2. Supplement transformations by suggesting matching elements: Create transformations or suggest the difficulty of transformation to the human (see Mediatability[1])

6. What is the bottom line for the Biologist?

Better Search for Biological Web Services

Search can be searched with more precise terms and concepts. Search by ontology concept and add facets to make precise filtering.

Convenience in Creating Workflows

Find and mash services together with ease. The tools can suggest the degree of match and also create data mappings. The workflows can be made graphically and then executed by just a point and click. There is no need to download, install and configure a number of applications.

Faster processing and result generation

The backend services can be Cloud based providing results much faster than any single computer.

No need for heavy in-house computing facilities

Use services that are hosted on clouds and avoid the equipment costs and all the hassle of hardware maintenance. Pay per use pricing model is convenient for sporadic usage.

7. Tools

Firefox Plug-in
Annotate web pages inside the browser and submit them to the index

Indexing/Search framework
1. Built using the technology made for faceted classification of Web APIs [2].
2. Multiple Apache Lucene indexes in the back-end

8. References

3. SA-REST: Semantic Annotation of Web Resources, W3C member submission by Wright State University http://www.w3.org/Submission/SA-REST/