Brief Bio and (PR)²: Problems & Pitches – Rants & Raves by Janos G. Hajagos

Self Introduction

Dr. Hajagos (Senior Programmer/Analyst) has been instrumental in bringing advanced medical informatics approaches to the Hospital and Health Sciences Center of Stony Brook University. He is an active participant in the World Wide Web Consortium’s Healthcare Life Sciences Interest Group (W3 HCLS IG). Dr. Hajagos has also brought his extensive knowledge of semantic ontology to the organization and ingestion of mined data for our VIVO implementation. Dr. Hajagos is the principal data analyst in the New York State Department of Health Modernization of Medicaid Initiatives through a SUNY contract. He is the co-principal investigator on the VIVO mini-grant: Integrating the UMLS Ontology into VIVO for Linking Biomedical Scientists. He received his Ph.D. in Ecology and Evolution from Stony Brook University in 2005.

Publications


Figure or Logo for Nametag:
General Questions

1) What is your main interest in attending the workshop?
   My main interest in attending the workshop is to learn how national researcher network is evolving and what are the future directions that the participants see VIVO moving towards. Also, how the work in the Stony Brook mini-grant could be used to further enhance the future of a national researcher network.

2) What three features or functions of researcher networking (i.e., tools and services related to assisting researchers with finding people, resources, data, projects, and scholarly works) are most critical to adoption?
   Semantic technology is enabling the power and flexibility of the VIVO application and ontology. However, it is a relatively new technology and requires a whole perspective change for programmers and developers moving from a relational data model to an RDF data model. The semantic web tools are not at the same level of maturity as those for traditional relational data tools. For finding resources across research networks agreed upon standard for linking. The three critical features for adoption are:
   
   1. Developer Acceptance of the RDF model
   2. Well matured tools for working with RDF data
   3. Agreement on URI's to interconnect VIVO implementations

   My feeling is that if we build it they will come.

3) What three features of researcher networking are most critical to success after adoption, or sustainability?
   1. Keeping information updated on a continuous basis from automated and manual entry.

   Nothing will frustrate users more if they make changes to their VIVO profile and they get overwritten by an automated process.

   3. Finer grained control of display and content of the VIVO profile

   A populated instance of the VIVO ontology makes a lot of data available but if it is not flexible in how the data is displayed this might be a future challenge. Many users do not understand the difference between data and presentation. This can pose a challenge as VIVO builds out its ontology. As VIVO profiles become richer they might become too complex and costly for complete display. I think there will need development of tools to create abridged/executive summaries of a profile to fit personal and institutional needs.

4) Are you aware of especially innovative approaches to any of these features or functions?
   For issue of maintainability, I think an approach is to keep manually curated data in a separate named graph from data pulled in an automated method on the backend. There will be a need to develop a set of rules (maybe SPIN) that will integrate information together for the VIVO public display. In such data integration provenance of the data/named graphs will become important.
Another approach is directly in the VIVO ontology is to add annotations for manual entry. The feasibility of direct integration into the ontology would need to be explored.

I have been experimenting with a templating system for specifying flexible display of semantic web data. This approach is used in the faceted browser that is hosted at google code http://code.google.com/p/spyder-web/. By directly connecting a templating system to the SPARQL endpoint it would be possible to build abridged views of the data.

5) What features of researcher networking are most important to you as a researcher, for your own use?

I am interested in the ability to connect research interests across multiple institutional repositories and to find novel research interests between researchers. At the first level is to find links between researchers when a clinical researcher uses a different vocabulary for the same concept as a basic science researcher. At the second level it to links like researchers work based on a biological phenomena, such as, a metabolic pathway, published in a publically accessible database.

6) Are you or your group working on any of these features?

Yes, as part of the mini-grant we are integrating the UMLS ontology by utilizing concept unique identifiers URIs to integrate the normalize research interest for biomedical researchers. We are building software tools and publishing data sets to make this work.

More importantly we are promoting that the publication of more scientific research data in a normalized form will enable the real novel research network. We are participating in the HCLS IG working group task forces and beginning collaboration with outside researchers.

7) What would you like to learn / achieve at the workshop?

I want to learn from different VIVO implementers how they are working toward the goal of the national researcher network. My hope is that as a group we can brainstorm and find future direction for enabling the utilization of semantic web technology to improve the efficiency and effectiveness of science.