Translational Medicine in Action: Linking and Visualizing a Network of Biomedical Research Scientists Using Nexus

Erich Bremer, Tammy DiPrima, Janos Hajagos
Stony Brook University School of Medicine
Stony Brook, NY US

Abstract: The goal of translational medicine is to translate basic science research into advances in clinical medicine. One way to meet this goal is to pair up basic scientists with clinical researchers who share common research interests. The challenge is that the terms used by each group do not perfectly align. To demonstrate the utility of using semantic web technology in translational medicine we applied it to interconnect clinical and basic scientists research interests. The research interests of SUNY REACH faculty were obtained from MeSH terms of publication data and are expressed in the VIVO ontology normalized to the UMLS. The VIVO ontology is part of the NIH funded VIVO project to interlink research scientists across different institutions. To explore novel interconnections in the network of research scientists the Nexus visualization environment was utilized. Nexus, a locally developed project, is a semantic web visualization tool built on the OpenSimulator platform. Nexus allows collaborative real time viewing and annotating of RDF data in a 3D environment.

The goal of SUNY REACH (Research Excellence in Academic Health) is to promote collaboration in biomedical research among SUNY’s geographically separated campuses and investigators. VIVO is an open-source semantic web developed by a consortium of academic universities (http://vivoweb.org), for the purpose of integrating and promoting a network of scientists. Shown below is the data in the VIVO ontology (right screenshot), and it rendered in the VIVO web application (left screenshot):

VIVO RDF can be access via URL or via SPARQL Endpoint

http://link.informatics.stonybrook.edu/umls/

UMLS - The central concept behind the UMLS (Unified Medical Language System) is the Concept Unique Identifier or CUI. The UMLS is a long term research projects of NIH’s National Library of Medicine. Every CUI has at least one representation as a string in one of the source vocabularies. The publicly-accessible English language portion of the 2010AA UMLS currently has over 43 different sources and 2,191,102 unique CUIs. In PubMed, MeSH terms are given for every indexed paper in Medicine. These MeSH terms can be normalized by the UMLS as MeSH is a source vocabulary of the UMLS. Every MeSH term identifier also has a unique UMLS CUI.

The first figure on the left shows a faceted browser displaying a URI. The middle figure the SPARQL Construct query to subset UMLS CUIs and Vivo research interests from the RDF repository. The figure on the right shows how the SUNY Reach investigator normalized MeSH research interests annotated with types from the UMLS Semantic Network.

Schematic of Data Flow

1. Extract & Transform
2. Load
3. Visualize

Nexus Design Methodology
Nexus uses a back-end server written in Java using Jena, Virtuoso, and a custom RDF ontology that controls visuals and layout aspects (nex:color, nex:xyz, nex:shiny, nex:nodesize, etc.). The flow of data from the back-end server to the front-end user interfaces is almost entirely a stream of triples. Front-end code processes the nex triples for instructions on how information should be presented and displayed for the user(s). Since 3D models are entirely represented in RDF, they can be persisted and shared by different front-end locations and/or triple stores. Nexus is still under heavy code revision. The sky and water are not necessary in OpenSimulator, it’s just there because it looks good.

Problems and Challenges
- Create a semantically driven web site of researchers → implement VIVO
- Import initial 534 SUNY scientists into system → convert CSV files to VIVO RDF
- Address lack of linkage on research interests → normalize data via PubMed MeSH/UMLS
- Visualize and study RDF data on SUNY Researchers → use Franz Gruff and in-house Nexus

Future Directions
- Expand # of SUNY Researchers in VIVO and maintain information
- Test how well extracted PubMed based research interests match self reported profile interests
- Consume RDF into other systems like standard web sites and reporting systems
- Develop HTML5/WebGL version of Nexus for greater flexibility
- Add and experiment with additional layout algorithms in Nexus

Erich Bremer, Associate Director Medical Informatics – erich.bremer@stonybrook.edu