Empowering Translational Research using Semantic Web Technologies

Amit P. Sheth
Wright State University - Main Campus, amit@sc.edu

Follow this and additional works at: https://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

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.
Empowering Translational Research using Semantic Web Technologies

OCCBIO, June, 2008

Amit P. Sheth
Kno.e.sis Center, Wright State University
amit.sheth@wright.edu
• Translate discoveries that typically begin at “the bench” with basic research — in which scientists study disease at a molecular or cellular level — into progress to the clinical level, or the patient's “bedside.”

• Increasingly biology and biomedical science is a data driven discipline. Dealing with large amounts of distributed and (syntactically, structurally and semantically) heterogeneous data, and analysis of such data is a key challenge.

• Semantic Web is a Data Web where by associating meaning with data, it is easier to search, integrate and analyze data.
Translational Medicine ...  

...needs a connection

Hypothesis Validation  
Experiment design  
Predictions  
Personalized medicine

Biomedical Informatics

Semantic Web research aims at providing this connection!

Etiology  
Pathogenesis  
Clinical findings  
Diagnosis  
Prognosis  
Treatment

Genome  
Transcriptome  
Proteome  
Metabolome  
Physiome  ...

Genbank  
Uniprot

Pubmed  
Clinical Trials.gov

Medical Informatics  
Bioinformatics

More advanced capabilities for search, integration, analysis, linking to new insights and discoveries!
Evolution of the Web

1997

- Web of pages
  - text, manually created links
  - extensive navigation

2007

Web as an oracle / assistant / partner
- “ask to the Web”
- using semantics to leverage text + data + services + people

Web of databases
- dynamically generated pages
- web query interfaces

Web of services
- data = service = data, mashups
- ubiquitous computing

Web of people
- social networks, user-created content
- GeneRIF, Connotea
Outline

- Semantic Web – very brief intro
- Scenarios to demonstrate the applications
- Some of the Kno.e.sis capabilities in Semantic Web technologies and their applications to biomedicine and health care in collaboration with biomedical scientists and clinicians
Semantic Web Enablers and Techniques

- **Ontology**: Agreement with Common Vocabulary & Domain Knowledge; Schema + Knowledge base
- **Semantic Annotation (metadata Extraction)**: Manual, Semi-automatic (automatic with human verification), Automatic
- **Computation/reasoning**: disambiguation, semantics enabled search, integration, complex queries, analysis (paths, subgraph), pattern finding, mining, hypothesis validation, discovery, visualization
Maturing capabilities and ongoing research

- Text mining: Entity recognition, Relationship extraction
- Integrating text, experimental data, curated and multimedia data
- Clinical and Scientific Workflows with semantic web services
- Hypothesis driven retrieval of scientific literature, Undiscovered public knowledge
Opportunity: exploiting clinical and biomedical data

Collection of (heterogeneous) Documents
- text
- XML
- HTML pages
- Databases

Scientific Literature
- PubMed
  - 300 Documents Published Online each day

Health Information Services
- Elsevier iConsult

User-contributed Content (Informal)
- GeneRifs

NCBI Public Datasets
- Genome, Protein DBs
  - new sequences daily

Clinical Data
- Personal health history

Laboratory Data
- Lab tests, RTPCR, Mass spec

Search, browsing, complex query, integration, workflow, analysis, hypothesis validation, decision support.
Scenario 1:

- Status: In use **today**
- Where: Athens Heart Center
- What: Use of semantic Web technologies for **clinical decision support**

Also research on adaptive clinical processes
Operational since January 2006
Active Semantic EMR

Knowledge Enabled Information and Services Science
Agile Clinical Pathways

• Volatile nature of execution environments
  – May have an impact on multiple activities/ tasks in the workflow

• HF Pathway
  – New information about diseases, drugs becomes available
  – Affects treatment plans, drug-drug interactions

• Need to incorporate the new knowledge into execution
  – capture the constraints and relationships between different tasks activities
New knowledge about treatment found during the execution of the pathway

New knowledge about drugs, drug-drug interactions
Today, the Food and Drug Administration (FDA) is announcing to voluntarily withdraw <Entity id="122805" class="DrugOntology#prescription_drug_brandname">Bextra</Entity> ("Entify id="10288" class="DrugOntology#pre.valdecoxib" to the market. Pfizer has agreed to further discussions with the agency.

...This request is based on:
- Reports of serious and potentially life-threatening skin reactions, including deaths, in patients using Bextra. The risk of these reactions in individual patients is unpredictable, occurring in patients with and without a prior history of <Entity id="14280" class="DrugOntology#interaction_with_physical_condition">sulfur allergy</Entity>.

...Date created: <Regexp type="date">April 7, 2005</Regexp>
Diabetes mellitus adversely affects the outcomes in patients with myocardial infarction (MI), due in part to the exacerbation of left ventricular (LV) remodeling. Although angiotensin II type 1 receptor blocker (ARB) has been demonstrated to be effective in the treatment of heart failure, information about the potential benefits of ARB on advanced LV failure associated with diabetes is lacking. To induce diabetes, male mice were injected intraperitoneally with streptozotocin (200 mg/kg). At 2 weeks, anterior MI was created by ligating the left coronary artery. These animals received treatment with olmesartan (0.1 mg/kg/day; n = 50) or vehicle (n = 51) for 4 weeks. Diabetes worsened the survival and exaggerated echocardiographic LV dilatation and dysfunction in MI. Treatment of diabetic MI mice with olmesartan significantly improved the survival rate (42% versus 27%, P < 0.05) without affecting blood glucose, arterial blood pressure, or infarct size. It also attenuated LV dysfunction in diabetic MI. Likewise, olmesartan attenuated myocyte hypertrophy, interstitial fibrosis, and the number of apoptotic cells in the noninfarcted LV from diabetic MI. Post-MI LV remodeling and failure in diabetes were ameliorated by ARB, providing further evidence that angiotensin II plays a pivotal role in the exacerbated heart failure after diabetic MI.

Angiotensin II type 1 receptor blocker attenuates exacerbated left ventricular remodeling and failure in diabetes-associated myocardial infarction., Matsusaka H, et. al.
Scenario 2

• **Status:** Completed research
• **Where:** NIH/NIDA
• **What:** Understanding the genetic basis of nicotine dependence.
• **How:** Semantic Web technologies (especially RDF, OWL, and SPARQL) support information integration and make it easy to create semantic mashups (semantically integrated resources).
Genome and pathway information integration

- Reactome
  - pathway
  - protein
  - pmid

- KEGG
  - pathway
  - protein
  - pmid

- HumanCyc
  - pathway
  - protein
  - pmid

- Entrez Gene

- GeneOntology
  - GO ID

- HomoloGene
  - HomoloGene ID
Knowledge Enabled Information and Services Science

BioPAX ontology

Entrez Knowledge Model (EKoM)
Biological Significance:

- Understand the role of genes in nicotine addiction
- Treatment of drug addiction based on genetic factors
- Identify important genes and use for pharmaceutical productions
Scenario 3

- **Status**: Completed research
- **Where**: NIH
- **What**: queries across integrated data sources
  - Enriching data with ontologies for integration, querying, and automation
  - Ontologies beyond vocabularies: the power of relationships
Link between glycosyltransferase activity and congenital muscular dystrophy?

Adapted from: Olivier Bodenreider, presentation at HCLS Workshop, WWW07
In a Web pages world...

Gene ID: 9215

- **has_associated_disease**: Congenital muscular dystrophy, type 1D
- **has_molecular_function**: Acetylglucosaminyl-transferase activity

Adapted from: Olivier Bodenreider, presentation at HCLS Workshop, WWW07
With the semantically enhanced data

```
SELECT DISTINCT ?t ?g ?d {
    ?t is_a GO:0016757 .
    ?g has_molecular_function ?t .
    ?g has_associated_phenotype ?b2 .
    ?b2 has_textual_description ?d .
    FILTER (?d, "muscular dystrophy", "i") .
    FILTER (?d, "congenital", "i") .
} .
```

From medinfo paper.
Adapted from: Olivier Bodenreider, presentation at HCLS Workshop, WWW07
Scenario 4

• Status: Research prototype and in progress
• Where: UGA
• What:
  – Semantic Problem Solving Environment (PSE) for Trypanosoma cruzi (Chagas Disease)
    • Workflow with Semantic Annotation of Experimental Data already in use
    • Knowledge driven query formulation
Knowledge driven query formulation

Complex queries can also include:
- on-the-fly Web services execution to retrieve additional data
- inference rules to make implicit knowledge explicit
Scenario 5

- When: Research in progress
- Where: Cincinatti Children’s Hospital Medical Center, AFRL
- What: scientific literature mining
  - Dealing with unstructured information
  - Extracting knowledge from text
  - Complex entity recognition
  - Relationship extraction
An excessive endogenous or exogenous stimulation by estrogen induces adenomatous hyperplasia of the endometrium.

- Entities (MeSH terms) in sentences occur in modified forms
- "adenomatous" modifies "hyperplasia"
- "An excessive endogenous or exogenous stimulation" modifies "estrogen"
- Entities can also occur as composites of 2 or more other entities
  - "adenomatous hyperplasia" and "endometrium" occur as "adenomatous hyperplasia of the endometrium"
• What can we do with the extracted knowledge?

• **Semantic browser demo**
Conclusion

Semantic web technologies can help with:

– Fusion of data: semi-structured, structured, experimental, literature, multimedia

– Analysis and mining of data, extraction, annotation, capture provenance of data through annotation, workflows with SWS

– Querying of data at different levels of granularity, complex queries, knowledge-driven query interface

– Perform inference across data sets
• Researchers: Satya Sahoo, Cartic Ramakrishnan, Pablo Mendes and Kno.e.sis team

• Collaborators: Athens Heart Center (Dr. Agrawal), CCHMC (Bruce Aronow), NLM (Olivier Bodenreider), CCRC-UGA (Will York), UGA (Tarleton), Bioinformatics-WSU (Raymer)

• Funding: NIH/NCRR, NIH/NLBHI, NSF

http://knoesis.org
References


- Demos at: http://knoesis.wright.edu/library/demos/