

ABSTRACT

The adoption of the Semantic Web by life sciences provides exciting opportunities for exploring novel ways to conduct biomedical research. Particularly, the approval of the Web Ontology Language (OWL) by the World Wide Web Consortium (W3C) has provided a global standard for shared representation of biomedical knowledge in the form of ontologies. However, though there are numerous examples of bio-ontologies being used to describe "what is" (i.e., a universal view of reality), there is a dearth of examples where ontologies are used to describe "what might be" (i.e., a hypothetical view of reality).

This thesis proposes that, to achieve scientific rigor, it is important to consider approaches for explicitly representing subjective knowledge in a particular domain - in particular, we examine phenotypic classifications in the clinical domain. We provide supporting evidence, that the OWL is suitable for formally representing these subjective perspectives. Envisioning ontologies as hypothetical, contextual and subjective is a notable departure from the commonly-held view in the life-sciences, where ontologies (ostensibly) represent some "universal truth". We support these arguments with both empirical and quantitative studies. We demonstrate that, when expressed in OWL, many phenotypic classification systems can be accurately modeled *in silico*. We then demonstrate that these knowledge-models can be "personalized", and show that such models can enable the automated analysis of data in a transparent manner. This results in more rigorous clinical research, while simultaneously allowing the clinicians to maintain their role as the final arbiters of decisions.

Finally, we investigate methodologies that might facilitate the encoding and sharing of personalized expert-knowledge by non-knowledge-engineers - a necessary step in making these ideas useful to the clinical community. The knowledge-acquisition bottleneck is the primary barrier to the widespread use of ontologies in life sciences. Thus, we investigate data-driven methodologies to automatically extract knowledge from existing data-systems, and show that it is possible to "boot-strap" the construction of knowledge models through various data-mining algorithms.

Taken together, these studies begin to reveal a path toward Web-embedded, *in silico* clinical research, where knowledge is explicit, transparent, personalized, modular, globally-shared, re-used, and dynamically applied to the interpretation and analysis of clinical datasets.

BIOGRAPHICAL NOTES

Born: September, 20, 1978, Iran

Academic Studies: B.Sc., Sharif University of Technology, 2001
M.Sc., University of Tarbiat Modares, 2003
M.Sc., University of Sheffield, 2005

Current Position: PhD. candidate

GRADUATE STUDIES

Field of Study: The application of Semantic Web Technologies in Clinical Research.

SELECTED PUBLICATIONS

Soroush Samadian, Benjamin M. Good, Bruce McManus, Mark D. Wilkinson (2012) A data-driven approach to automatic discovery of prescription drugs in cardiovascular risk management. *Bio-Ontologies 2012*

Soroush Samadian, B. McManus, and M. D. Wilkinson, "Extending and encoding existing biological terminologies and datasets for use in the reasoned Semantic Web.," *Journal of biomedical semantics*, vol. 3, no. 1, p. 6, Jul. 2012.

M. D. Wilkinson, L. McCarthy, B. Vandervalk, D. Withers, E. Kawas, and S. Samadian, "SADI, SHARE, and the *in silico* scientific method.," *BMC bioinformatics*, vol. 11 Suppl 1, no. Suppl 12, p. S7, Jan. 2010.

SELECTED PRESENTATIONS

A data-driven approach to automatic discovery of prescription drugs in cardiovascular risk management. *Bio-Ontologies 2012*

Data and Knowledge Modeling in Support of Clinical Research, Young Investigators forum, 2009,

SUPERVISORY COMMITTEE

Dr. Bruce McManus
Dr. Wyeth Wasserman
Dr. Raymond Ng
Dr. Mark Wilkinson



THE UNIVERSITY OF BRITISH COLUMBIA

PROGRAMME

The Final Oral Examination
For the Degree of

DOCTOR OF PHILOSOPHY
Bioinformatics

SOROUSH SAMADIAN

B.Sc., Sharif University of Technology, 2001
M.Sc., University of Tarbiat Modares, 2003
M.Sc., University of Sheffield, 2005

Friday, June 14, 2013, 3:00 pm
Room B141, Buchanan Building, 1866 Main Mall
Latecomers will not be admitted

**“Constructing and Applying Semantic Models of Clinical Phenotypes
to Support Web-embedded Clinical Research”**

EXAMINING COMMITTEE

Chair:

Dr. David Owen (Pathology and Laboratory Medicine)

Supervisory Committee:

Dr. Wyeth Wasserman (Medical Genetics)

Dr. Raymond Ng (Computer Science)

Dr. Mark Wilkinson (Center for Plant Biotechnology and Genomics,
Polytechnic University of Madrid)

University Examiners:

Dr. Ho Kendall (Emergency Medicine)

Dr. Scott J. Tebbutt (Respiratory Medicine)

External Examiner:

Dr. William Ghali

Community Health Sciences Department

University of Calgary

Calgary, AB