Report on Semantic Web for Health Care and Life Sciences Workshop

Huajun Chen, Kei Cheung
Zhejiang University, Yale University
huajunsir@zju.edu.cn
kei.cheung@yale.edu

Michel Dumontier
Carleton University
michel_dumontier@carleton.ca

Eric Prud'hommeaux
World Wide Web Consortium
eric@w3.org

Alan Ruttenberg
Science Commons
alanruttenberg@gmail.com

Susie Stephens
Eli Lilly and Company
stephens_susie_m@lilly.com

Yimin Wang
University of Karlsruhe
ywa@aifb.uni-karlsruhe.de

ABSTRACT
The Semantic Web for Health Care and Life Sciences Workshop will be held in Beijing, China, on April 22, 2008. The goal of the workshop is to foster the development and advancement in the use of Semantic Web technologies to facilitate collaboration, research and development, and innovation adoption in the domains of Health Care and Life Sciences, We also encourage the participation of all research communities in this event, with enhanced participation from Asia due to the location of the event. The workshop consists of two invited keynote talks, eight peer-reviewed presentations, and one panel discussion.

Categories and Subject Descriptors

General Terms: Algorithms, Design, Experimentation, Human Factors,

Keywords: Semantic Web, Health Care, Life Sciences, Biomedical Informatics, Ontology, Data Integration.

1. INTRODUCTION
Biomedical researchers need to be able to ask questions that span many heterogeneous data sources in order to make well-informed decisions. For this to be achieved, knowledge about drugs, patients, diseases, genes, proteins, and pathways must be effectively integrated. Yet the integration of disparate biomedical data continues to be a significant challenge.

Many organizations are now exploring the use of Semantic Web technologies in the hope of easing the effort required for data integration. The benefits promised by the Semantic Web include aggregation of heterogeneous data using explicit semantics, simplified annotation, web-based sharing of findings, expression of rich and well-defined models for data aggregation and search, easier reuse of data in unanticipated ways, and the application of logic to infer additional insights.

W3C has established the Semantic Web for Health Care and Life Sciences Interest Group (HCLSIG) to help organizations in their adoption and application of Semantic Web technologies. The HCLSIG is chartered to develop and support the use of Semantic Web technologies to improve collaboration, research and development, and innovation adoption in the domains of Health Care and Life Sciences. Towards this goal, we are organizing a workshop on the Semantic Web for Health Care and Life Sciences in conjunction with WWW2008. The workshop is located in Beijing, the capital of mainland China. In addition to its cultural, historical, and political significance, Beijing is one of the Asian regions that is experiencing great economic and technological growth. It is truly exciting for this workshop to take place in Beijing at this time.

2. CONTENT
We have accepted eight paper submissions for presentation at the workshop. Among the accepted Semantic Web papers, three were from Asia, three papers were from Europe, and two from North America. The workshop will help to promote Semantic Web research and collaboration in the HCLS context across different cultural backgrounds. The accepted papers cover a broad range of topics including web ontology modeling, data integration, data mining, ontology storage, scientific collaboration, and community interaction.

Michel Dumontier et al report their endeavor in modeling the pharmacogenomics of depression. They present a knowledgebase designed with Semantic Web technologies capable of capturing essential aspects including genes, gene variant, SNPs, drugs, measures and outcomes, as well gene-drug interactions and drug treatments. The knowledge base draws upon OWL 1.1's role chains to facilitate question answering and pharmacogenomics knowledge discovery.

Yuxin Mao et al present a case study of ontology engineering as well as its applications in traditional Chinese medicine with consideration for the special characteristics of the Chinese language. In addition, a sub-ontology model is proposed to support dynamic ontology reuse. The engineering effort results in an ontology of 20,000 classes and 100,000 instances.

Alexander Garcia et al focus on using Semantic Web technologies to capture and model neuron-radiological knowledge in a head injury scenario. They further investigate how the proposed ontology can be utilized to facilitate information retrieval from a pre-existing database of medical neurological images.

Andrew Newman et al introduce a project called BioMANTA that focuses on utilizing Semantic Web technologies together with a

WWW 2008, April 21–25, 2008, Beijing, China.
ACM 978-1-60558-085-2/08/04.

scale-out architecture to provide efficient, scalable RDF storage and inferencing. The fully-developed system provides efficient analysis, querying, and reasoning about protein-protein interaction data.

Yu Tong et al investigate the feasibility of utilizing the enhanced integration capability enabled by Semantic Web technologies to map and analyze various biomedical complex networks. A new methodology named semantic graph mining is proposed accordingly, which uses the semantic graph model to integrate graph mining and ontology reasoning for better analyzing biomedical complex networks. As part of the biomedical use case, the Gene Ontology is characterized and visualized based on resource importance calculation; and also a global herb-drug interaction network is mapped through semantic integration of legacy relational databases for discerning interesting patterns such as frequent sub-graphs and community structures.

Matthias Samwald et al present the "Entrez Neuron", a neuron-centric interface that allows for keyword-based queries against a coherent repository of OWL ontologies about neuronal structure, physiology, mathematical models and microscopy images. The returned query results are organized hierarchically according to brain architecture. Where possible, the application makes use of entities from the Open Biomedical Ontologies (OBO) and the 'HCLS knowledgebase developed by the W3C Interest Group for Health Care and Life Science. It also makes use of the emerging RDFa standard to embed ontologies and semantic annotations within its HTML based user interface.

Sudeshna Das et al describe a scientific collaboration framework (SCF) with semantic underpinnings that is based on the popular content management system Drupal. The framework is designed to support interdisciplinary scientists in publishing, sharing and discussing content such as articles, perspectives, interviews and news items, as well as provide personal biographies and research interests – the basics of any online community. These web materials can then be linked to external knowledge repositories of life science entities such as genes, antibodies, cell-lines or model organisms.

Matthias Löbe et al outline how Semantic Web technology can play a simple but crucial role in organizing contact management for clinical experts. They designed a lightweight ontology that contains only the main concepts of clinical trials. Several relational database schemata were then mapped to this ontology. As a result, data quality was significantly improved with regard to syntax uniformity, completeness, consistency, and refresh period. Based on this data, they built an application for patients to search for experts depending on their indication and location.

3. KEYNOTE TALKS

We invited two keynote speakers: Greg Tucker-Kellogg (Chief Technology Officer, Lilly Center for Drug Discovery), and Mark Wilkinson (Professor of Medical Genetics, University of British Columbia). Dr. Tucker-Kellogg will address the challenges in managing information within the pharmaceutical industry, and for the role that the Semantic Web may play in this effort. Dr. Wilkinson will describe his work in bringing together Web Services and the Semantic Web.

4. PANEL DISCUSSION

Our workshop features a panel discussion. Experts in the field will serve as panelists to discuss topical issues at the intersection of Semantic Web and HCLS research. Panelists will be invited to discuss how to Asia is well positioned and stands to benefit from this combined approach. Audience participation will provide additional insight and discussion with panelist members. Details of the panel discussion are being planned at the time of writing this report.

5. NEXT STEP

This workshop serves as a start to broaden the participation of Asian research communities in Semantic Web for Health Care and Life Sciences. In the future, more efforts will be required to increase the interest of using Semantic Web in different HCLS sectors in Asia. In addition, we need to identify HCLS use cases for the Semantic Web communities in the East and in the West to work together.

6. ACKNOWLEDGMENTS

We are very grateful for the support and endorsement by our program committee members. They are Dean Allemang, Christopher Baker, Olivier Bodenreider, Keith Boyce, Cory Brouwer, Bill Bug, Helen Chen, William Cheung, Tim Clark, Sudeshna Das, Don Doherty, Lee Feigenbaum, Andrew Gibson, Carole Goble, Yufeng Guo, Amarnath Gupta, Volker Haarslev, Peter Haase, William Hayes, Yusheng He, Ivan Herman, Vipul Kashyap, Marijke Keet, Phil Lord, M. Scott Marshall, Robin McEntire, Parsa Mirhaji, Mark Musen, Fabian Neuhaus, Eric Neumann, Chimezie Ogbuji, Elgar Pichler, Matthew Pocock, Daniel Rubin, Matthias Samwald, Susanna Sansone, Michael Schroeder, Nigam Shah, Amit Sheth, Robert Stevens, Rudi Studer, Adam West, Katy Wostencroft, Xuezhong Zhou. We would also like to give particular thanks to the WWW workshop organizer Irwin King for his great assistance in the process of organizing this workshop.