



How semantic technology helps fighting infectious diseases: Biological systems approach to understand microbial pathogens

Erich Gombocz¹, James Candlin²

¹IO Informatics, Inc, Berkeley, CA

²SAGE N-Research, Milpitas, CA

Session:

Use Cases

Brief Description:

Semantic integration of microbial data with public resources creates a knowledgebase with functionally categorized proteins and disease-related pathways. Biological signatures are identified and applied to help prevent outbreaks.

Full Description / Abstract:

- Bacterial and viral-caused infectious diseases account for major health threats, yet rapid detection, discovery of causal relationships and prevention remains challenging. A wealth of information (and even complete microbial genomes) exists, but integrated use in context of biological systems to understand mechanisms in infectious diseases is just starting.
- Semantic technology excels in its ability to contextually integrate multiple resources and create biological signatures via semantic (SPARQL) queries.
- This methodology identifies peptides from different microorganism with common disease causing mechanisms, and applies resulting patterns as marker-based decision support to identify and categorize microbial threats.
- Semantic integration of experimental data with public resources (PATRIC, ICTV, VIDA, Viral ORFeome, miRBase, BioCyc, KEGG) creates a knowledgebase with functionally categorized proteins, bacterial or virus-specific genes and their related pathways. Semantically constructed biological signatures apply SPARQL within Applied Semantic Knowledgebases (ASK) to assist in outbreak detection and provide insights into microbial threats.

Additional Information (References, books, articles, publications)

E. Gombocz: "Semantic cross-domain integration: The intersection of research, public, and clinical data; creating applicable knowledge for decision support in patient-centric healthcare", *NCBO Webinar Series* Stanford, CA, May 4, 2011

T.N. Plasterer, R. Stanley, E. Gombocz: "Correlation Network Analysis and Knowledge Integration" In: *"Applied Statistics for Network Biology: Methods in Systems Biology"*, M. Dehmer, F. Emmert-Streib, A. Graber, A. Salvador (Eds.), Wiley-VCH, Weinheim, ISBN: 978-3-527-32750-8 (2011)

E. Gombocz, R. Stanley, J. Eshleman, C. Rockey: "Semantic Data Integration: Answer to Complexity in Translational Research", Poster at *Bio-IT World Conference 2011*
World Trade Center, Boston, MA, April 12-14, 2011

Speaker Biography:



Dr. Erich Gombocz has over 30 years of experience in Life Science research, laboratory automation and data management in scientific and distributed systems environments, plus more than 25 years programming experience in instrumentation control, user interface, database design, scientific analysis, and on-line laboratory automation as well as being developer of innovative software algorithms and architecture.

From 2000-2003, as Chief Science Officer of Biosentients, Erich contributed substantially to design and implementation of the company's informatics technology, including its systems architecture, data structure and analytical modules. Together with Bob Stanley he founded IO Informatics in 2003 focusing on semantic knowledge applications.

Dr. Gombocz has published over 60 scientific publications and holds currently more than 40 biotechnology- and software-related US and international patents. He is an international expert in separation science and bioinformatics, a member of several professional organizations, and serves on the editorial board of a number of scientific journals.

Significance:

This presentation demonstrates a use case in which the application of semantic technology is pivotal - not only to meaningfully integrate diverse resources, but also through the creation of biological signatures for complex processes via SPARQL, leading to direct applicable decision support with major consequences for the prevention of epidemic infectious diseases.