

Keynote talk: What can semantics do for Bioinformatics?

Amit Sheth

LSDIS lab at the University of Georgia and Semagix, Inc.

Bioinformatics is about providing biologists and those in allied disciplines with ability to exploit information - information that is increasingly distributed, heterogeneous and massive. Bioinformatics has already succeeded in utilizing database management, workflow and information retrieval technologies, which have provided syntactic search, heterogeneous data access and sharing, and limited forms of integration. A good amount of effort has also involved use of statistical and syntactic techniques to support the essential tasks of finding patterns, similarities, and matches to identify building block structures. Looking to the future, we can realize more exciting potential of bioinformatics if we have more automated ways for analysis leading to insight and discovery - to understand cellular components, molecular functions and biological processes, and more importantly complex interactions and interdependencies between them. And while lot of effort in the last decade focused on genes, next set of challenges involve more complex structures of protein and carbohydrates.

This talk focuses on semantics enabled bioinformatics. We outline increasing use of semantic techniques in bioinformatics for search, browsing and integration. Ontologies provide underpinning for most of today's semantic techniques and the Semantic Web research, and bioinformatics is one of the most aggressive adopters of ontologies among science and industry domains. We also investigate how we start to address more pertinent discovery and analysis questions such as: "how does the glycosylation pattern of the cell change when it undergoes a physiological change, such as differentiation or transformation to a cancer cell?" We will also weave a brief overview of the research and exciting commercial state of the art in the semantic technology (specifically ontology driven information systems).

Sample background material: Bioinformatics for Glycan Expression (an example research project) [1], SCORE [2]/ Semantic Enhancement Engine (an example technology) [3], Taalee Semantic Search [4], Semagix Freedom (an example commercial product) [5].

Speaker

Amit Sheth is a Professor at The University of Georgia (UGA) and CTO of Semagix, Inc. He started the LSDIS lab at UGA in 1994. Earlier he served in R&D groups at Bellcore, Unisys, and Honeywell. In August 1999, he founded his second company, Taalee based on the research in the LSDIS lab, and managed it as the CEO until June 2001. He currently serves as the CTO and a co-founder of Semagix, Inc, which emerged from a merger with Taalee. His R&D has led to three significant commercial products, several deployed applications, many invited talks including 18 keynotes, and over 150 publications [6].

References

- [1] Bioinformatics for Glycan Expression. <http://lsdis.cs.uga.edu/proj/glycomics/>
- [2] Amit Sheth, Clemens Bertram, David Avant, Brian Hammond, Krzysztof Kochut, and Yashodhan Warker. Managing Semantic Content for the Web. *IEEE Internet Computing*, pp. 80-87, Aug 2003. <http://lsdis.cs.uga.edu/lib/download/S+2002-SCORE-IC.pdf>
- [3] Brian Hammond, Amit Sheth and Krzysztof Kochut. Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content. In V. Kashyap and L. Shklar, editors, *Real World Semantic Web Applications*, 2002. <http://lsdis.cs.uga.edu/lib/download/HSK02-SEE.pdf>
- [4] Taalee Semantic Search. <http://www.streamingmediaworld.com/gen/reviews/searchassociation/index.html>
- [5] Semagix Freedom. <http://www.semagix.com/>
- [6] Amit Sheth homepage. <http://lsdis.cs.uga.edu/~amit>

