

Semantic Web Information Systems: NSF IDM workgroup report on challenges and opportunities in Semantic Web

Amit Sheth, Isabel Cruz and IDM-SWIS Workgroup
<http://lsdis.cs.uga.edu/SWIS/>

Executive Summary

Semantic Web is emerging as an important new multidisciplinary area in Computer Science. Realizing the potential significant role of Information and Data Management (IDM) research community in enabling the Semantic Web, as well as potential impact on IDM research due to core Semantic Web concept of semantic annotation of all resources, the NSF-IDM program has in the past sponsored a symposium and a workshop. During NSF IDM PIs' meeting in May 2002, a workgroup further discussed a possible research agenda with IDM focus within the context of broader research related to Semantic Web. This research agenda was termed Semantic Web Information Systems (SWIS), which is summarized in this report.

Introduction

The Semantic Web concept is widely seen as the vision of the next generation of the Web. Some of its many views include:

- “The semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” [BHL01]
- “...next generation internet, where we will not only surf the web, but work the web.” [A01]

- “The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications. [W3C01]
- “The Semantic Web is a web of data, in some ways like a global database.” [B98a]

Many researchers view semantics as the necessary and the only way the current Web can continue to scale and support its growth. At least two reasons that advocate important role of the IDM community in enabling the Semantic Web are (a) semantics has been a recurrent theme of many IDM research topics including modeling, query processing, transaction management, etc., and (b) IDM research has highly successful track record in addressing scale, performance, reliability and robustness of information systems that involve large amounts of data—and Semantic Web will involve very large amounts of data and metadata.

Potential and likely relevance of IDM research to Semantic Web and vice-versa has in the past prompted the IDM Program Managers Drs. Maria Zemankova and Bhavani Thuraisingham to sponsor the International Semantic Web Working Symposium (SWWS) [<http://www.semanticweb.org/SWWS/report/>] and the Amicalola Workshop on

DB-IS Research for Semantic Web and Enterprises[http://lsdis.cs.uga.edu/semN_SF/]. Correspondingly, Isabel Cruz (who co-organized SWWS) and Amit Sheth (who co-organized Amicalola workshop) were asked to lead the workgroup discussions at NSF IDM PI's meeting in May 2002.

SWIS Workgroup: Suad Alagic, Kevin Ashley, Isabel Cruz (co-chair), Wenfei Fan, Michael Kifer, Phokion Kolaitis, Zoe Lacroix, Steve Liddle, Weiyi Meng, Fred Sadri, Amit Sheth (co-chair), Ellen Spertus, Susan Urban, Clement Yu

Summary of Discussions

Participants chose to characterize the research area at the intersection of IDM and Semantic Web as Semantic Web Information Systems (SWIS). Key observations of the workgroup are as follows:

- Web is increasingly more complex, and simultaneously there is information overload and starvation.
- SWIS uses explicit and machine processable semantics, which will allow orders of magnitude improvement in IDM.
- SWIS can significantly enhance in activities such as scientific discovery, business processes, national defense, healthcare, and education.
- IDM knows how to efficiently deal with very large amounts of diverse information (scale, heterogeneity).
- Fundamental IDM concepts/techniques of query processing/views, constraints, modeling, data consistency/transactions, etc. are highly relevant to the realization of SWIS.

Research Agenda

SWIS workgroup outlined the following non-exhaustive list of topics that present research challenges and promise of useful results:

- Languages for representing and reasoning with ontologies-- more expressive, deal with XML and other formats
- Constraints: more general/expressive formalisms, at different levels of the architecture, and their processing
- Semantic preserving/schema/data transformation (mapping, evolution), integration, versioning
- Query languages for heterogeneous data (structured/semi-structured/unstructured data), metadata and knowledge bases
- Logics for Semantic Web to support processing of uncertain, inexact, incomplete, temporal information
- Workflow processes (discovery, negotiation, quality, trust)
- User interfaces, interactivity
- Optimization, scalability, performance

Recommendations

SWIS group recommended that the time was right to initiate a focused initiative within IDM that involve basic research and experimentation on SWIS.

[B98] T. Berners-Lee, Semantic Web Road map, <http://www.w3.org/DesignIssues/Semantic.html>

[BHL01] T. Berners-Lee, J. Hendler, O. Lassila. The Semantic Web. Scientific American, 284 (5), 2001, pp. 28–37

[W3C01] W3C: Semantic Web Activity Statement, 2001, <http://www.w3.org/2001/sw/Activity>.

Appendix: Summary of Participants' Suggestions

Below is a partial list of key topics workgroup participants presented.

Steve Liddle

- computational feasibility, deductive DB, automatic markup, ontology management, approximate information

Phokion Kolaitis

- expressiveness vs tractability
- semantics of schema mapping, data integration
- schema mapping and query answering, incomplete databases

Kevin Ashley

- identification of conceptually important parts
- link analysis (learning from links)

Suad Alagic

- capturing semantics: metadata, class object, XML schemas
- semantics-preserving transformation/mappings

Clement Yu

- automation of ontology construction

Carlo Zaniolo

- ontology versioning, document change management
- data intensive web language

Fred Sadri

- uncertain, inaccurate data
- credibility or reliability of data

Michael Kifer

- more expressive language than description language

- need to go beyond XSB: research on deductive database and especially object-oriented deductive databases is very relevant (the past work on these subjects is already extensively used by the various semantic web researchers)
- need more than ontologies to be ready for birthing applications

Weiye Meng

- schema extraction (XML can help here)
- schema integration (ontology can help here)

Wenfei Fan

- ontology mapping/integration
 - schema equivalence, information capacity
 - schema-directed transformation/integration
 - integrity constraint propagation
- query language for ontology, metadata and data

Susan Urban

- discover and create semantic relationships between data and services
- queries based on data, services and relationships
- dynamic specification of processes (workflow with relaxed specifications)

Zoe Lacroix

- semantics of data sources/web resources and its use in query optimization
- soundness of queries/data using constraints