V R I J E   U N I V E R S I T E I T   B R U S S E L

F A C U L T E I T   W E T E N S C H A P P E N
V A K G R O E P   I N F O R M A T I C A   E N   T O E G E P A S T E   I N F O R M A T I C A
S Y S T E M S   T E C H N O L O G Y   A N D   A P P L I C A T I O N S   R E S E A R C H   L A B

STAR Lab Technical Report

Amicalola Report: Database and Information Systems Research Challenges and Opportunities in Semantic Web and Enterprises

Amit Sheth and Robert Meersman

affiliation	University of Georgia, LSDIS Lab, Athens, GA, USA, amit@cs.uga.edu
keywords	semantic web, ontology, database technology
number	STAR-2002-13
date	12/11/2002 5:32
corresponding author	
status	published
reference	Sheth A. & Meersman R. (ed.), SIGMOD Record Special Issue on Semantic Web, Database Management and Information Systems 2000 31(4)

Pleinlaan 2, gebouw G-10, B-1050 Brussel
Phone: +32-2-629.3308 • Fax: +32-2-629.3525

Amicalola Report: Database and Information Systems Research Challenges and Opportunities in Semantic Web and Enterprises[1]

Amit Sheth
University of Georgia
Athens, GA, USA
amit@cs.uga.edu

Robert Meersman
Vrije Universiteit Brussel
Brussels, Belgium
meersman@vub.ac.be

http://lsdis.cs.uga.edu/SemNSF/

1 Executive Summary

This report describes opportunities for the DB/IS community to contribute to the advancement of the Semantic Web and the challenges or new research topics presented by the vision of the Semantic Web to the database and information systems (DB/IS) researchers. It is based on the NSF-OntoWeb Invitational Workshop on DB/IS Research for Semantic Web and Enterprises that was held during April 3-5, 2002 at the Amicalola Falls State Park in the northern Georgia mountains. Most of the workshop participants were industry R&D leaders or senior academics from the fields of database management and information systems who at various points in time have been deeply involved with semantics or interdisciplinary work in knowledge representation. Others included AI and database researchers who are active with Semantic Web related projects, those who have worked on semantic modeling and interoperability dealing with different domains (e.g., geographic) and/or media (video, images). This report could not have been produced without their generous contribution, which we explicitly acknowledge and for which we are once more very grateful.

Amicalola Working Group:

Organizers: Robert Meersman, Amit Sheth
Applications subgroup: Michael Bordie (Cooridnator), Umeshwar Dayal (Coordinator), Ramesh Jain. Frank Manola, Hans-Jorg Stork, Bhavani Thuraisingham
Ontology subgroup: Stefan Decker (Coordinator), Yahiko Kambayashi, Vipul Kashyap (Coordinator), Max Egenhofer, William Grosky, Michael Uschold
Web Services subgroup: Karl Aberer, Isabel Cruz, Dieter Fensel (Coordinator), Mike Huhns, Munindar Singh (Coordinator), Ling Liu, Rudi Studer

Participants identified significant past successes in DB/IS that are likely to play an important role in realizing the Semantic Web, especially by bringing this community's unique strengths in technical capabilities in semantic modeling, query processing, transactions and workflow systems. Equally important is this community's ability to develop technologies that are scalable, high performance, and robust that this area has proven success with. Although semantics is not a new topic to this community, the participants identified several new research challenges for DB/IS researchers that Semantic Web poses. Besides the broad vision of seeing the entire web as a global information system, and observing semantics as the primary enabler of scalability required for the next generation of the Web, this community also sees more immediate applications that benefit enterprise and e-commerce between a group of enterprises and industry through the scalability and productivity improvements semantics can bring. Several ideas in community building, outreach and funding initiatives were discussed.

[1] This workshop was sponsored by NSF CISE-IIS (Program Manager: Dr. Bhavani Thuraisingham), OntoWeb, University of Georgia Research Foundation, Inc. and the LSDIS Lab.

2 Background

The Semantic Web concept was widely adopted as a vision, a challenge, and, by some a necessity. Many elaborations have been provided, including:

- The Semantic Web is a computer system, a distributed machine which should function so as to perform socially useful tasks. [B98b]
- "The Web of data (and connections) with meaning in the sense that a computer program can learn enough about what data means to process it." [B99]
- "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]

For the purposes of this report, we focus on the unique distinction between the current web and the Semantic Web. The current Web is sometimes referred to as an "eyeball Web" where all interpretation of accessed information occurs, literally, in the eye of its beholder, viz. a human. On the Semantic Web interpretation will be *primarily* done by software agents: every information-dependent resource, includingenterprises, information services, application services, and devices, need to become augmented with machine processable descriptions to support the finding, reasoning about (e.g., which service is best), and using (e.g., executing or manipulating) the resource. The idea is that self-descriptions of data and other techniques would allow context-understanding programs to selectively find what users want, or for programs to work on behalf of humans and organizations to make them more scalable, efficient and productive.

None of the above definitions of or perspectives in Semantic Web exclude significant role of database and information systems (DB/IS), quite to the contrary.. Semantics has indeed been an important undercurrent in database areas of modeling, query processing and transactions. Yet as observed at a CoopIS panel [CoopIS01] and in the background on the Amicalola workshop[Agenda] , most recent workshops and conferences have had limited involvement and participation by the DB/IS research community.

Arguably a significant majority of Semantic Web researchers today come from AI. This has allowed the early research in Semantic Web to benefit from the strength of past AI research, which includes skills in knowledge modeling and representation languages. There are some significant differences in the way how different research seem to be viewing approaches and mechanisms to achieve a Semantic Web.

One distinction stands out. Database has for long realized the value of data independence, and has distinguished between schema and data. This has been the key to the scalability, efficiency, and robustness of data management solutions. By the desire to annotate each resource, the Semantic Web vision calls for creation of the equivalent of a massive new distributed database of metadata (annotations), whose size can be of the same order of magnitude as data itself and of which the complexity will likely exceed that of the data itself. This should clearly be viewed as the opportunity for DB/IS to contribute synergistically with other disciplines to make the Semantic Web a reality.

Thus the workshop's agenda was to discuss what DB/IS can do for the Semantic Web and to identify new research challenges for the DB/IS research community in the process of achieving the vision of Semantic Web. In the process, the

Amicalola workshop complements and continues the work of other workshops which studied the relationships of the Semantic Web vision with various disciplines [S02], including AI sub-communities such as knowledge representation [E02] and machine learning.

As we noted earlier, semantics has been part of various methods and techniques in database management, including (but not limited to) modeling, query processing, transaction management. However, emerging Semantic Web changes the thinking about semantics at two levels:

- semantic annotation of all resources changes the scale at which the techniques need to exploit semantics, and
- broader form/type of semantics, as in domain semantics, which opens new research opportunities in applying them.

3 Workshop Overview

The workshops consisted of three activities. The first day involved short presentations by most the participants (presentations and position papers appear on the workshop web site and the proceedings, respectively). The second day consisted of workgroup discussions. Three workgroups were formed by the participants on the topics of ontology, web services and application pull. This division was likened to that in medical field of anatomy, physiology, and pathology, respectively. The third half day consisted of review of workgroup results, an exercise in discussing the role of DB/IS in enabling and making Semantic Web successful, and the new challenges the emerging area of Semantic Web poses for DB/IS research. (A table of the results from this last activity is appended at the end of this Report.) Let us briefly review output of each of the workgroups, followed by the review of the relationships between DB/IS and Semantic Web.

3.1 Application Pull

There was significant agreement, especially among the industry participants, that a future Semantic Web promises significant benefits to businesses. Semantics was seen as a required contribution to the efficiency of the world (e-)economy, in at least three concrete ways.. First, by generically improving the efficiency (e.g., reduce the cost) of business, government, and personal processes currently on or planned for the web through the creation of easily accessible, standardized, *meaningful* interfaces with and descriptions of systems and data. Second, semantics are required to address the challenges posed by the growth and sophistication of the web. Machine processable semantics is seen as the critical elements of a scalable solution to deal with the current and anticipated growth of the web and to deal with the expected vast number and sophistication of the services available over the web. Third, semantics are required to exploit the unique opportunities that the Semantic Web will offer such as converting all relevant processes (e.g., tax preparation, supply chains management) from incomplete (e.g., using only accessible information) and discrete (e.g., compute once and again when ever the solution becomes grossly sub-optimal) to comprehensive (e.g., using all relevant information) and continuous (i.e., tax planning and preparation are an integral part of the life of a person or organization such that every financial event can be considered in real-time).

Focusing as well on the process perspective, rather than only on data, the Application Pull subgroup observed that web services are for real, that their organizations have started to prototype using them, and the promise of semantic composition of processes (as in workflows) hold huge promise to business productivity and efficiency.

Key discussion areas and conclusions of this WG included the following:

(a) applications ranging from individual applications (e.g., continuous tax preparation), B2B (e.g., supply-chain) and scientific/ engineering research could benefit from Semantic Web R&D, with corresponding beneficiaries varying from individuals, organizations and society

(b) Semantic Web should and can lead to significant benefits that include lower barriers to entry, adaptability or dynamic behavior (to support changing situation), supporting continuous activity, and various improvements (timeliness, accuracy, transparency, etc.)

(c) Challenges to realizing Semantic Web's potential to applications include design/specification of upper ontologies and domain ontologies with broader acceptance, support for ontology management activities (create, search, select, maintain, map/integrate), etc.

Significant parts of the discussion involved outlining a real possibility of obtaining an order of magnitude or higher improvement in key business applications such as supply-chain management if even limited part of Semantic Web vision is realized, as well as in noticing that at some companies, Web Services and their use/support for semantics can be seen as initial forays towards Semantic Web applications. This WG felt that the Semantic Web vision is more than a research initiative, and that there are plentiful real-world applications that can benefit as aspects of the Semantic Web vision are realized.

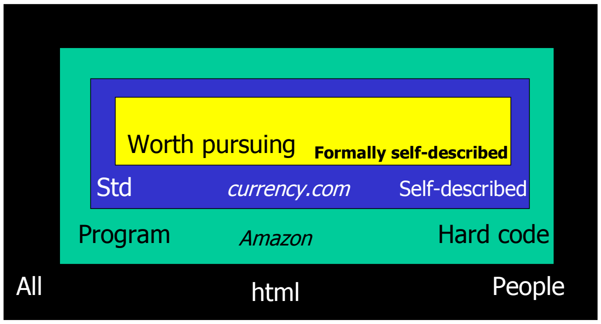
3.2 Ontology

This subgroup focused on the role of database management in support for ontology engineering and management. The subgroup discussed many aspects of ontology lifecycle (ontology search, match, merge/refine, maintenance, creation, modification/versioning, requirements analysis, evaluation, learning, consistency checking, deployment).[It then identified potential role of known database research and technology in addressing various step in ontology lifecycle, as well as identified distinctions between assumptions and focus of database research with respect to unique features and requirements of such methods and techniques in supporting various step of ontology lifecycle. A selection of the items of research identified in this WG includes:

- Inference v/s Query Rewriting/ Processing for Semantic Integration (e..g., RichPerson = (AND Person (> Salary 100))
- Distributed Inferences and Loss of Information when supporting relationships other than equality
- Query Languages for combining metadata and data queries
- Graph-based data models and query languages
- Schema Correspondences/Mappings
- Intensional Answers (when answers are descriptions, e.g. (AND Person (> Salary 100)) instead of a list of all rich people)
- Semantic Associations (identification of meaningful or contextually relevant relationships between classes and instances)

3.3 Web Services

There was, perhaps predictably ,significant interest in web services at Amicalola Falls, with overlapping and complementary discussions in this subgroup as well as Application Pull subgroup). It quickly identified that Semantic Web Services (SWS)--the Web Services that are "formally self-described"-- to be of primary research interest and of critical importance to Semantic Web. The role of P2P (peer-to-peer protocols) as a possible new way of organizing WS-based systems was discussed, as well as moving from natural language (as in textual description of Web Services) to tags to domain ontologies was described as a way to provide increasing level of semantics.



The Web Services subgroup noted that compared to the issues that deal with data, web services are more challenging in matters such as modeling, organizing collections, discovery and comparison, distribution and replication, access and composition, fulfillment (contracts, coordination versus transactions, compliance), and quality aspects more general than correctness or precision, compliance). They are also more dynamic and have more difficult characterization of security and trust. This discussion led to the following research challenges for realizing SWS in future:

- Conversational (state-based, event-based, history-based) web services
- Interoperability, composition and translation of web services
- Representations for services: programmatic self-description
- Commitments, contracts, negotiation
- Discovery, location, binding
- Compliance,
- Cooperation
- Transactional workflow: rollback, roll-forward, semantic exception handling, recovery
- Trustworthy service (discovery, provisioning, composition, description)
- Security; privacy v/s personalization
- Quality of Service, w.r.t. various aspects Esoteric and advanced issue

Workshop included presentations and some discussion on areas that are related to semantic web—multi-model semantics, context-aware computing, semantics to pragmatics, experiential computing. Although these may not yet be identified as one of the core areas of Semantic Web, they may become critical new areas in their own right.

In summary, the current web supports virtually every type of human endeavor and these uses are growing dramatically in coverage, sophistication, and adoption. Semantics is viewed as the most important enabler to continue this with better scalability and productivity. DB/IS research has the potential to assume an increasingly important role in making the Semantic Web happen for business and scientific uses, significantly impacting how the technology and the Web supports individuals, organizations and society at large.

4 Next Steps

4.1 Outreach and Community Building

This workshop has already been followed by another workgroup on Semantic Web at the NSF-IDM PI's workshop in May [S+02]. Additionally, the organizers reviewed the results at the panel "Research Directions for the Semantic Web" organized by Rudi Studer at OntoWeb3 in Sardinia, Italy in June 2002.

We also expect improved interactions between relevant communities by involvement of prominent DB/IS researchers in specific Semantic Web activities such as the ISWC Conference, and increased participation in the Semantic Web tracks that are being appended to a number of relevant recurring events such as the WWW Conference.

A number of networks and resources supporting the emerging Semantic Web community are or have been set up, probably the most famous one at present being the OntoWeb Thematic Network of the EU (under its 5th Framework Program), http://www.ontoweb.org., in which 180 partners (with more that 50% from industry) are actively collaborating to gather, represent and disseminate knowledge about relevant technologies, methods and tools. As the Semantic Web grows, we expect such initiatives to multiply and spread to other networks supporting a variety of interested communities, at least as Special Interest Groups.

4.2 Nourishment and Sustenance

With the basis provided by Semantic Web Working Symposium, this workshop and NSF-IDM work group, NSF-IIS is currently evaluating the possibility of initiating a program that can sponsor research in this area.

A number of initiatives are also envisaged in Europe notably as part of the planned 6th Framework Program, due to start in 2003 and in which the Semantic Web will be the cornerstone in more than one Key Action of its Work Program (http://www.cordis.lu). A number of concretely focused calls were already done as part of the 5tyh Framework, and a number of projects are under way or starting up as this Report appears (ibid.).

5 Conclusions

The success and potential of the web is leading to the possibility that every information resource, person, organization, and many of the activities relating to them will be located on or be driven by the Web. This poses the opportunity of qualitatively improved interactions but also quantitatively changes the scale and scope of already well-understood challenges in computer science. The simple extrapolation of the current Web (e.g., simply more resources) requires qualitatively improved solutions to problems of interaction between resources, currently called interoperation, integration, and collaboration. The sole, scalable solution involves improving the automation of interactions, which in turn can occur only with access to enhanced "meaning" of all resources and the ability of software agents on the Web to deal with this enhanced meaning.

We see Semantic Web as a long term and fundamental research direction for DB/IS which requires vigorous research program. It has unique challenges in such issues as scalability, performance and robustness that DB/IS has successfully tackled in the past, yet Semantic Web poses unique new challenges for research. Amicacola group believes that both a significant funding program targeted at DB/IS and collaboration with allied disciplines should be part of a research agenda.

References

[A01] J. Andersen, The Semantic Web Tutorial, XML 2001, Finland. http://xmlfin.ecraft.fi/archive/xml2001/1

[Agenda] http://lsdis.cs.uga.edu/SemNSF/SemWebWorkshopAgenda.htm

[B98a] T. Berners-Lee, Semantic Web Road map, http://www.w3.org/DesignIssues/Semantic.html

[B98b] T. Berners-Lee, *Interpretation and Semantics on the Semantic Web*, 1998 http://www.w3.org/DesignIssues/Interpretation.html

[Be99] Tim Berners-Lee, Weaving the Web, Harper, 1999.

[CoopIS01] Panel on "Semantic Web: Rehash or Research Goldmine?" D. Fensel, R. Meersman, J. Mylopoulous and A. Sheth, CooPIS, Trento, Italy, 2001.

[E02] J. Euzenat, Report from the NSF-EU Workshop „Research Challenges and Perspectives of the Semantic Web", Sophia-Antipolis, October 2001, http://www.ercim.org/EU-NSF/semweb.html

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

[S02] R. Studer, "Research Directions for the Semantic Web, " (Panel Introduction), OntoWeb3, Sardinia, Italy, June 2002.

[S+02] A. Sheth, et. al. Semantic Web Information Systems: NSF IDM workgroup report on challenges and opportunities in Semantic Web, July 2002.

Appendix: Compilation of the Amicalola Working Group's collective perception on the (bidirectional) interaction between the SW and the DB/IS research

DB / IS subcommunity	How is it relevant to research on the SW	How may the SW stimulate research in this community
DB theory	Type theory, Complexity, theory of concurrency	Ontology axiomatics and theory; formal semantics; semantics for incomplete, inconsistent and evolving representations
Data(base) semantics	Everything; in particular ontology language development; constraints; data structures	Ontology modeling; formal semantics of web services
Normalization/ design	Not specifically as such; some work on Non-First Normal Form	Requirement for formal properties for ontology organization; perhaps ontology design guidelines or “semantic normal forms”; conflict resolution; redundancy checks in general
Data modeling	reuse/extend/map DM formalisms, techniques and methods e.g. EER, ORM, UML for ontology (content) specification and design	semantic data modeling; ontology content creation techniques and methods; complex ontological relationships; domain models
View integration	Ontology alignment, translation, object identities, updateable views…; model mappings	see Federated DBs; ontology support for view and application integration; ontology composition and update
Schema integration	apply to autonomously designed schemas; global schemas as pre-ontologies? conflict detection	Ontology alignment; new kinds of models will pose new kinds of problems
Deductive DB/Datalog	Learn from its failure, query processing and F-logic	how to handle different complexity levels efficiently
Multimedia DB	Image ontologies; semantic indexing; similarity-based search	Image-based ontologies?
Temporal/Spatial DB	GIS semantics and archiving; histories data management;	requirement to model temporal knowledge as first class citizen in ontologies; spatial, temporal modeling in upper ontologies; versioning of GIS becomes critical issue
Document DB	Digital libraries, unstructured data; standards for digital library resource descriptions to beused on the SW	Lack of a priori global model presents a research challenge
OO DB	Object-oriented and object-based models for ontologies, extensible databases; modeling of object behavior; build OODB into Java	management of large collections of object-, behavior- and resource identifiers
Visual DB	Visualization for the SW, visual queries; ontology visualization	semantic upgrades of image databases to be used as visual ontologies
XML/Web DB	Most relevant, caching	Size and semantics; XML shortcomings for semantics definition
Distributed DB	everything	trust/privacy/compliance issues in distributed DBMS; design/dynamic tailoring of DDBMS underlying web services
Constraint DB	Constraint enforcement as semantics mechanism; semantics-based query processing	Non-closed world assumption issues

Transaction modeling	loosening of ACID properties	Web services, Extended distributed transaction models; non-CWA issues; smart user profiling
Transaction processing	limits of what can/must be transactional	ACID properties of Web services; semantic support for very long transactions
Mobile DB	not directly; "mobile" is a platform issue	context-aware computing; device location-independent semantics; mobility issues raised/enabled by the (Semantic) Web
Main memory DB	Semantic caching	possibly semantic caching i.e. using application semantics or context
Parallel DB	unclear at present; straightforward reuse/apply (e.g. parallel queries, transactions, …) in certain niches	Not clear at present Web SoA; parallel architectures for ontology servers?
DB machines		Not clear at present Web SoA
DB security	A lot, e.g., access control	trust and privacy, QoS; dynamically changing and conflicting security requirements
Federated DB	Autonomy; approaches for integrating heterogeneous data sources, in particular web information sources; mediator/wrapper-based architectures	www = huge federated DB; develop more powerful (scalable) approaches for ontology alignment and integration; heterogeneous sources may have different credibility; service composition
Query processing	high applicability; e.g. "smart" query enhancement	
Query optimization	high applicability; e.g. use domain-knowledge to optimize query execution and rewriting	
Information retrieval	broad applicability of techniques and theory;	
DB interoperability	Everything; esp. see federated DBs; see schema integration	Semantic aspects of interoperability; see federated DBs; quality of interoperation
DB versioning	Link maintenance; ontology versioning	Annotations, ontology modeling, versioning of instance data
Metadata		Annotations, ontology modeling, versioning
Mediation/Middleware	Web services will benefit	P2P, collaboration, new market for mediating components
DB warehousing	DW architectures for decision support; improve e.g. web service efficiency; see the (S)Web as a giant DW	Smart data warehousing; share/compose application semantics; ontology behind "real" data
Data(base) mining	web mining; clustering; learning; information extraction profiles	mining from text; exploit semantics in mining; derive semantics inductively from query results on "real" data including exceptions; machine learning
Database architectures and DBMS	DBMS (components) as web service(s); add semantics to every function/module in a DBMS's architectures	Ontology support in data dictionaries; new, more flexible DB architectures for better SW support and processing on the web
Web-IS architectures	fitting enterprise IS (components) into the SW; Web IS; also see DBMS architectures	New architectures and design principles for Web IS
Functional	design of web services; functional	Decomposition and composition of web

modeling	modeling that deals explicitly with a domain's semantics	services; event modeling
IS in organizations	looser coupling required, provide potential for organizations to morph into the SW; see also workflow modeling	serving new organizations of business, community and government with emergent SW-based IS technology
Web-IS applications		smart (ontology-driven) SW portals and search engines ("Google++"-type); SW-based "direct marketing"-style systems; smart user profiling
IS workflow modeling	exception handling in long (business) transactions; workflows as "the" paradigm for "programming" the SW	unreliability of components; unavailability of services
IS methodologies	ontology lifecycle issues; as IS components become more intelligent, work shifts to self-organization	New thinking required! E.g. Web IS in enterprises; how must business processes change to deal with existence of the SW; develop/maintain SW-based systems for user community unknown a priori
CASE tools	ontology management systems	
User interfaces	new applications of design principles for GUIs	New and complex requirements and methods, immersive environments
DB application architectures		Web application service
AI-and-DB	knowledge representation, inference	
Uncharted territory 1		Sensor input and stream data management
Uncharted territory 2	In general, most algorithms in DM are poor when they are applied to access, report etc data on the web. Domain semantics in such requests need to be exploited, where however "centralized" solutions (where resources need to notify potential requestors) will not be scalable.	