

“Semantic technology is here to stay...”

Amit P. Sheth

Professor, Computer Science, University of Georgia
Director, Large Scale Distributed Information System Lab
Chief Technology Officer (CTO) – SEMAGIX Corporation



[Sheth](#) is an educator, researcher and entrepreneur. He joined the UGA and started the [LSDIS](#) lab in 1994. For nine years before that, he served in R&D groups at Bellcore, Unisys, and Honeywell. In August 1999, Sheth founded Taalee, Inc., a Venture Capital funded enterprise software and internet infrastructure startup based on the technology developed at the LSDIS lab. He managed Taalee as its CEO until June 2001. Following Taalee's acquisition/merger, he serves as the CTO and co-founder of [Semagix](#), Inc (formerly Voquette, Inc). His research has led to two companies, several commercial products and many deployed applications. He has published over 150 papers and articles, given over 130 invited talks and colloquia including 19 keynotes, (co)-organized/chaired ten conferences/workshops, served on 90 program committees, etc.

Miltiadis: Dear Prof Sheth, we are really honored for your kindness to provide us this interview. I would like to start by asking you what do you think for the so-called Next Generation Web Research. Do we have to wait for much more time before we will have an evidence of real world applications?

Amit: Thanks for talking to me. There are a number of exciting “next generation” technologies on the horizon that will drastically change how we see the Web, internet, computing and communication in general. Use of semantics (as in semantic technology or in a more focused perspective taken by some in SW community) is only one of the important component technologies. Here again, semantics is only being exploited by a majority of SW researchers for a relatively narrow purpose of automating simpler- or shall we say well defined- things, although some commercial semantic technologies are being used for addressing broader objectives. In the larger scheme of things, more automated travel reservation or scheduling is not something humans care so much about – I mean it is interesting for pedagogy but given that there are so many subjective elements in making a travel arrangement, not something that would be used in the way the problem is formulated. What really matter is the real purpose of any technology – e.g., for businesses it is gains in productivity and competitiveness, for a common man it is often about luxury, relaxation or entertainment, and so on. Here automation is only a part of the equation.

As for your question on evidence of real world applications – for semantic technology (as also relevant to the vision of the SW), applications are here already – paid for and deployed! Because of business considerations, most of these are Enterprise applications rather than pan-Web applications.

As for your question on evidence of real world applications – for semantic technology (as also relevant to the vision of the SW), applications are here already – paid for and deployed! Because of business considerations, most of these are Enterprise applications rather than pan-Web applications

Miltiadis: Dear Amit, you are one of the most active people in SW and the most interesting is your IS perspective. What is your opinion for the role of SW in IS research and vice versa? What do you answer to all those people who claim that OK... SW? But where is it?

Amit: Semantics has long been recognized to be very important in IS, Databases, AI, Linguistics and many other fields. From the IS/DB perspective, I remember talking about “So Far (Schematically) yet So Near (Semantically)” in 1992, but lots of smarter people have talked about semantics for some time. More recently however, two things have happened – one positive, one potentially not so positive. The positive thing is that we have now been able to engineer semantic technology

that supports large scale semantic applications, and use large populated ontologies to provide semantic underpinning. At the same time a questionable development is a rather overwhelming importance attached to “formal semantics”. Tom Gruber who brought limelight to the term ontology in early 1990s last year talked about limited value and success of formal ontologies (something he had worked on) and underscored the importance of semi-formal ontologies. I have further explored this theme, along with real world observations, in my [article](#) in Data Engineering Bulletin in December 2003.

“Underlying to my albeit personal view is that I must have the expressiveness in representation to meet my application requirements, rather than starting at the other end -- with computability and computing concerns—for example of an inferencing technique- and then set out to determine expressiveness and my modeling capabilities”.

Underlying to my albeit personal view is that I must have the expressiveness in representation to meet my application requirements, rather than starting at the other end -- with computability and computing concerns – for example of an inferencing technique- and then set out to determine expressiveness and my modeling capabilities. Thankfully, query processing works for a broad variety of data and knowledge representation, and we have learned to implement that efficiently. This is a key relevance of IS/DB for semantic technology. As for SW for IS, we get to look at challenging applications where multidisciplinary approaches are necessary and hence learn to leverage techniques from allied areas. For example, to develop solutions for some of the complex problems in integration analytics that we deal with now, one needs to bring together automatic classification using machine learning, NLP techniques for document processing, and IS/DB for semi-structured and structured data management and query processing.

Miltiadis: I have seen your excellent work in the Large Scale Distributed Information Systems (LSDIS) Lab at the University of Georgia. What do we have to expect from your research in the near future?

Amit: At the LSDIS lab, our research on semantic technology and SW have two foci. One area of focus is on techniques for knowledge discovery and semantic analytics, where we are learning to analyze a large amount of information in the form of populated ontologies and automatically extracted semantic metadata from heterogeneous content. An intermediate outcome of this work has been the development of SWETO, an ontology testbed with a large public-use populated ontology for evaluating SW tools and benchmarking semantic application.

Our second area of focus is on Semantic Web Services and Processes, where we have taken a comprehensive view of semantics (describing data, function, execution and QoS semantics) to improve all steps in the Web Process lifecycle (annotation, discovery, publication, composition, orchestration). Intermediate outcomes from this research has been WSDL-S (for semantic annotation of WSDL), and tools for semantic annotation, publication and discovery of Web services, with more to come. Domain specific knowledge is critical understanding and applying semantics, so the LSDIS Lab has collaborations in the areas of bioinformatics, geographic information systems, and homeland security.

Miltiadis: You are also the Chief Technical Officer of SEMAGIX, one of the most significant companies in SW applications. Would you like to share with us your experience related to SEMAGIX?

Amit: In August 1999, I licensed a technology resulting from research at LSDIS to start Taalee and managed it as a CEO for two years. Taalee is now Semagix after its acquisition/merger and renaming. Most companies that we thought could be our competitors have disappeared for one reason or the other, and new ones have come up. Although the bubble-bust was not easy for anyone, us included, strength and uniqueness of our technology was the primary reason for our survival. Now the stability in our development team has allowed us to engineer a terrific technology. Our CEO Larry Levy’s ability to match the semantic technology and Freedom product with real market needs and his ability in signing up marquee customers in these markets have landed us several Fortune 500 customers as well as opportunities for developing incredibly exciting and demanding semantic solutions.

Although the bubble-bust was not easy for anyone, us included, strength and uniqueness of our technology was the primary reason for our survival. Now the stability in our development team has allowed us to engineer a terrific technology.

Miltiadis: Could you provide us some sound cases of SEMAGIX applications? How does industry (medical, banking sector, etc) reflects on new semantic web applications and tools?

Amit: Semagix has focused on Financial Services, Media & Entertainment, Government/Defense, and Pharmaceuticals. Most stand-out applications of Semagix Freedom are in semantic analytics involving analysis of heterogeneous documents and databases, although semantic integration and semantic search are often involved too. Primary enabling technical capabilities utilized in these solutions involve creation of large populated domain ontologies (from multiple knowledge sources and using techniques for disambiguation), automatic semantic metadata extraction, use of main memory techniques for demanding computation such as discovering complex relationships, and being able to do all these with structured, semi-structured and unstructured data. I believe some of the domains such as bioinformatics and pharmaceuticals are inherently well suited for SW applications as domain knowledge plays an important role.

Miltiadis: Dear Amit, I know that you are a visionary with an active participation in the field. Would it be possible to outline a few (lets say 3-5) hot research areas in the field of Next Generation Web Research?

Amit: Its hard to pick a few, but here I have a few favorite ones. In research arena, these include: increasing automatic extraction/annotation of newer forms of digital media, including streaming media, broadcast TV, and sensor generated data streams; complementing semantic or thematic metadata (and corresponding domain ontologies) with spatial and temporal metadata and ontologies, and providing comprehensive spatio-temporal thematic reasoning; and extending semantics description of static aspects (such as data input and output) of resources or Web services to descriptions related to functional and execution behavior and quality of service, along with increasing semantic support for dynamic nature of Web Processes. In commercial and application arena, some of the favorites include automated literature search and mining for pharmaceutical R&D; business intelligence applications of opinion and brand management for marketing; and increasing use of semantics in Web search especially as more major players compete with Google.

Miltiadis: Would you like to share your vision for semantic web for the next decade? Would it be possible to share with us a few examples of real world SW enabled tools and applications for citizens that you do believe that we will see in the next decade?

Amit: More exiting and important goals of the next generation are improving the human experience and enriching the living, and I can now see a possibility of a major shift from focus on computing to improving human experience – not only with better ability to use heterogeneous content and apply knowledge, but also to incorporate perception and pervasive computing. My view is an amalgamation of what I have seen on “experiential computing” by Ramesh Jain, “computing with words” by Lotfi Zadeh, and “humanist computing” by Jonathan Rossiter. For those focused on semantics and IS, we still need to address the difficult and fundamental problem of identifying entities (from unstructured text), semantic disambiguation and discovering (potentially fuzzy, inexact or probabilistic) relationships. And while formal representation and techniques certainly have a role, we need to find much better way for involving humans – much more than in human interfaces and visualization issues-- in any approach supporting semantics and knowledge management

“My view is an amalgamation of what I have seen on “experiential computing” by Ramesh Jain, “computing with words” by Lotfi Zadeh, and “humanist computing” by Jonathan Rossiter.

I consider myself to be among the very few lucky guys who has an opportunity to simultaneously work with entire span of research, prototyping, technology transfer, commercialization and real world application deployment.

Miltiadis: Do you think that the emergence of wireless technologies will influence our traditional approach of designing IS? Do you feel that we will soon be talking about a “ubiquitous semantic WEB”?

Amit: Two of the interesting issues that arise with wireless data applications are location awareness and increasing value of metadata due to the inability to transfer and display large amount of text. So wireless is about as different as interactive TV, streaming media or sensor data. Supporting these will provide new applications where semantics and metadata will play increasing role compared to text on terminals of wired computers.

Miltiadis: Amit, your career path in Industry and Academia is amazing. What can you glean from these experiences? What is in other words the added value on you from industry and academia?

Amit: I consider myself to be among the very few lucky guys who has an opportunity to simultaneously work with entire span of research, prototyping, technology transfer, commercialization and real world application deployment. At LSDIS, I can work with colleagues and a large group of PhD students on long term and conceptual research which allows me to

collaborate with industry and provide inputs to standards activities. We have twice licensed technology resulting from our research, leading to start ups, including Semagix (earlier Taalee). At Semagix, I get to work with smart engineers – some of whom are LSDIS alumni-- to develop a leading product in SW and architect customer specific solutions. On the same day I can work on research papers and prototypes, as well as deal with challenges of a deployment at a Fortune 500 customer. It has been incredibly exciting.

Miltiadis: Information Systems affect everybody's life. Nowadays we all discuss about the Knowledge Society? Is this utopia? What Knowledge Society means from you?

Amit: No, it is fairly certain that nothing we are seeing is a utopia. We all have tendency to get unduly excited with every new trend and fad, and after a field matures, we find out that instead of them being major life-changing technology or science, they are a step towards a continuing evolution. As to your question about the role of knowledge, we are now able to automate processes involving large knowledge bases. An average domain ontology developed by Semagix is populated by well over a million entity instances and well over a million relationship instances, typically created by integrating parts from multiple knowledge sources (some of which may be created and maintained by humans). Such a populated ontology at a center of a semantic technology is well suited for many interesting applications.

Miltiadis: Dear Amit, I asked from a student to prepare for you a profile for your work and career. In fact I asked him to summarize you (if I may say so) in few pages. He concluded his exercise by stating: "Amit Sheth delivers trust in his eyes... So I would ask you to share with as your work and life style.

Amit: I am a type A person-- Students at LSDIS and employees at Semagix do find me quite demanding ☺☺ Life style is pretty hectic, especially while directing both the LSDIS lab and Semagix R&D, sometime compromising social and family life. On the other had, when dealing with a challenging problem, such as developing a scalable technology that support and utilize semantics, one needs to have patience.

I have a few principles or guiding rules. First my own motto and guidance to my students is "learning to learn", as learning a skill or technology is not sufficient. New issues and challenges constantly arise, new techniques and technologies become relevant, so being able to learn new things, being adaptive is crucial if one wants to continue to be relevant. Second, I see opportunities as an educator, a researcher, and an entrepreneur (something I have adapted from Ramesh Jain). Correspondingly, I select activities to achieve impacts through training my PhD students to become among the best in our area of work, seek to achieve research impact through quality (as reflected in citations) rather than quantity (as reflected in number of publications), and try to achieve impact of research in terms of commercialized technology, deployed applications, and increased commerce.

**First my own motto and guidance to my students is "learning to learn", ...
"Second, I see opportunities as an educator, a researcher, and an entrepreneur"...
"I select activities to achieve impacts through training my PhD students to become among the best in our area of work, seek to achieve research impact through quality (as reflected in citations) rather than quantity (as reflected in number of publications).."**

Miltiadis: You have organized several workshops, tracks and special issues on SW related themes. Do you plan anything for the near future? We heard that you would have a leading role in new International Journal on Semantic Web and Information Systems of AIS SIGSEMIS.

Amit: Berner-Lee's coining of SW was quite timely. It was time to take help of semantics for bringing more value to the huge amount of information Web grew to contain. But semantics has been investigated in IS and DB for quite some time; in 70s and 80s in the context of semantic data models, and in 80s and 90s in the context of schema and database integration (e.g., in 1993-94, Vipul Kashyap and I talked about Semantic Information Brokering involving Web and non-Web content). AI and some other fields have embraced SW readily. So I am a bit surprised that IS and DB communities have not been very enthusiastic about SW. We had an excellent workshop on [DB/IS Reseach for SW and Enterprise in Amicalola](#) to discuss just this. And in the Data Engineering Bulletin [article](#) I tried to respond to some of the hesitations and caveats that DB and IS communities have expressed regarding SW and have perhaps kept them from embracing it. May be the timing is better for SIGSEMIS to build a more vigorous community, but I hope that you will involve both formal and non-formal approaches, and both automated and human-in-the-loop approaches.

Miltiadis: Amit, I would like to share with us your thoughts for the future of SW. Is it one more bubble? What are the critical milestones for its evolution?

Amit: When thinking about SW, I focus on semantics rather than Web – semantics has been around for a long time, and Web is one context of its application. If SW is narrowly defined, ruling out anything that does not involve formal representation and inferencing, then it may not be in vogue too long. That is why, we included enterprise in our Amicalola workshop name, and that's why Ramesh Jain and I have named the book series we just started "Semantic Web and Beyond: Computing for Human Experience."

Miltiadis: A few days ago I have a conversation with a colleague in the university. He wanted to learn more concerning Semantic Web and Next Generation Web Research. And basically he put me a dilemma: "Technologies or Theories?". What is your advice to a newcomer in the field? Many student members of our SIG would be interested on it.

Amit: Fundamental concepts, multidisciplinary approach (not limiting to just one of AI, KM, IR, DB, or IS), and smart engineering that scale with enterprise and Web scale data and large knowledgebases.

Miltiadis: Amit, you teach several semantic web courses on your university. Do you think that IS curriculum in the next years will incorporate such courses and in which direction?

Given the high level of student interest, I would venture to predict that this topic will be increasingly added to graduate curricula. Textbooks on SW and semantic information systems will likely be published this year, which should help.

Amit: Absolutely. I think I was the second person to teach a course on Semantic Web (and the first one that took a multidisciplinary view as the first one had a distinct AI perspective; Google: Semantic Web Course). I have already taught it thrice, and increasing number of faculty has offered a course on this topic world wide. Recently, I introduced a course on Semantic Web Services and Processes. Given the high level of student interest, I would venture to predict that this topic will be increasingly added to graduate curricula. Textbooks on SW and semantic information systems will likely be published this year, which should help.

Miltiadis: I would like to comment your perception for the role that IS community can play in SW. At the current moment the field is dominated by researchers from the AI and DB field. Do you think that can be a convergence?

Amit: This has got to be a multi-disciplinary approach. For example Semagix Freedom uses to a varying degree techniques generally considered to be part of KR, DB, IR, and lexical analysis.

Miltiadis: How do you find the formation of the New Special Interest Group on Semantic Web and Information Systems on AIS? What role do you expect from SIG SEMIS in the forthcoming years?

Amit: It all has to do with the enthusiasm of the community and adaptive direction by its leaders. You have invested considerable energy and have managed to get broad participation in a short time. Some of us would be delighted to be available as guide or resources, but it is on the shoulder of your team.

Miltiadis: Dear Amit thank you for your time. It was an excellent talk. Would you like to state something to our readers?

Amit: We live in an exciting time. It is now possible to engineer solutions of real world value – I know this as enterprise customers are paying for these semantic solutions. The problem remains challenging – a lot has to be done and can be done. Semantic technology is here to stay.

When thinking about SW, I focus on semantics rather than Web – semantics has been around for a long time, and Web is one context of its application. If SW is narrowly defined, ruling out anything that does not involve formal representation and inferencing, then it may not be in vogue too long.

We live in an exciting time. It is now possible to engineer solutions of real world value – I know this as enterprise customers are paying for these semantic solutions. The problem remains challenging – a lot has to be done and can be done. Semantic technology is here to stay.

Short Vita

Dr. Amit Sheth is a Professor of Computer Science and the director of the Large Scale Distributed Information Systems (LSDIS) Lab, at the University of Georgia. His primary current research interests include

- **Semantic Interoperability, Semantic Web and Global Information Systems** with applications to Digital Libraries, Video and Digital Media applications utilizing Broadband, and E-commerce, and
- **Enterprise Integration** with emphasis on Semantic Web Services, their discovery and composition; Multi-organizational Business Processes with emphasis on exception handling, and adaptation.

Sheth joined the UGA and started the LSDIS lab in 1994. In August 1999, Sheth founded Taalee, Inc., a Venture Capital funded enterprise software and internet infrastructure startup based on the technology developed at the LSDIS lab. He managed Taalee as its CEO until June 2001 when it was acquired by [Voquette](#), Inc. Subsequently he has been the CTO/co-Founder for Voquette, Inc. and Semagix (resulting from merger of Voquette and another company). He has also founded another high-tech company – Infocsm, Inc. He has provided consulting to a wide variety of companies involving real world solutions to the real world problems. His research has led to several commercial products and applications (including AdaptX/Harness from Bellcore based on InfoHarness project, METEOR EAppS from Infocsm based on the METEOR project, and Freedom from [Semagix](#) or Semantic Engine from Taalee based on SCORE technology and VideoAnywhere project), as well as two patents ([6,311,194](#), [5,241,675](#)). From 1985 to 1994, he served in R&D groups at Bellcore (now Telcordia Technologies), Unisys, and Honeywell.

Prof. Sheth has given over 130 invited talks, tutorials and professional courses, including nineteen keynote and plenary invited talks at international conferences and meetings,. He has over 150 publications, including two outstanding conference papers, and some of the most cited papers in federated databases, workflow management and semantic interoperability. He is on the editorial board of five journals. Recently he has served as a Program co-chair/Co-organizer of 2004 IEEE International Conference on Services Computing, [NSF-EU Workshop on DB/IS Issues for Semantic Web and Enterprises](#), [IEEE Digital Library Conference](#) and [WWW10 Workshop on Semantic Web](#). Earlier he has led six international conferences and workshop as a General/Program (Co-)Chair in the area of information system cooperation/interoperability, workflow management, and parallel and distributed information systems. He has also served twice as an ACM Lecturer, was one of the two international advisors of Japan's TARA program, and has served on over 75 program and organization committees. He received his B.E.(Hons.,1981) from B.I.T.S., Pilani, India and M.S.(1983) and Ph.D.(1985) from the Ohio State University.

Prof. Sheth enjoys mixing activities that span from conceptual research to entrepreneurship involving commercial products and real-world applications. His teaching activities have included introduction of innovative courses on Semantic Web, Semantic Enterprise, Enterprise Integration and Global Information Systems. These courses combine latest academic research with commercial state of the art research, as well as conceptual frameworks to hands on projects with practical applications. He is also involved in developing an interdisciplinary program in Masters of Internet Technology and has been on the steering committee of New Media Institute at UGA..

LSDIS lab (<http://lstdis.cs.uga.edu>) maintains very active collaboration with industry, and has been awarded six significant projects in the above mentioned areas with funding of over \$4.5 million by NIST, NRL, NSF and industry. Along with colleagues that jointly direct some of the projects-- Profs. Krys Kochut, John Miller, and I. Budak Arpinar, the lab has usually comprised of several research assistants and visitors.

Special Issue Theme: Semantic Web Challenges for Information Systems Research Community

Table of Contents

Lytras M. *, *Special Issue Editorial*

* ELTRUN, E-Business Center, Department of Management Science and Technology, Athens University of Economics and Business, Greece

Benjamins R.¹, **Contreras J.**¹, **Corcho O.**¹, **Gómez-Pérez A.**², *Six Challenges for the Semantic Web*

¹Intelligent Software Components, S.A., Spain

²Facultad de Informática. Universidad Politécnica de Madrid., Spain

Chebotko A., **Shiyong Lu.**, **Fotouhi F.**³, *Challenges for Information Systems Towards The Semantic Web*

³All Located at Department of Computer Science, Wayne State University, USA

Nickerson J.⁴, *Ideas on the challenge of the semantic web*

⁴Stevens Institute of Technology, USA

Hüsemann B., **Vossen G.**⁵, *Ontology-Driven Multimedia Object Management for Private Users –Overview and Research Issues*

⁵Both Located at European Research Center for Information Systems, University of Muenster, Germany

Sicilia M.⁶, *The Role of the Semantic Web in the Automation of e-Learning, Computer Science Department University of Alcalá, Alcalá de Henares (Madrid), Spain*

Klischewski R.⁷, *Semantic Web for e-Government – a Research Agenda*

⁷ Informatics Department, Hamburg University, Germany

Veltman K.⁸, *Challenges for the Semantic Web and Information Systems from Culture*

⁸ Maastricht McLuhan Institute (MMI), European Centre for Digital Culture, Knowledge Organization and Learning Technology, Maastricht, The Netherlands