Inside Bulletin

TOP NEWS
CFP for ECIS 2005
SW and IS track

AIS SIGSEMIS Bulletin
Editor
Miltiadis Lytras
EB members
Gottfried Vossen
Lina Zhou
Gerd Wagner
Ambjorn Naeve
William Grosky
York Sure
And all the SIGBoard
Members (see page 136)

EDITORIAL…

For the second time AIS SIGSEMIS Bulletin is on air. We would like to thank you for your warm receipt of the first issue which gives us a motive to continue our volunteer hard work towards more activities and services for our research community. Our portal site at www.sigsemis.org had **1650 unique visitors** (and same number of AIS SIGSEMIS Bulletin 1(1) downloads) from over **50 countries.** Inside this issue you can find many interesting things. A call for papers for our SIG Official Peer Reviewed Journal which will be published by IDEA Group (inaugural issue 1/2005). I invite you to consider this publication outlet as an interesting and high quality journal in which we will put all of our efforts in order to meet your high standards for quality and to communicate high impact research.

The special theme that is discussed in this issue is *Semantic Web Challenges for Knowledge Management: towards the Knowledge Web*. We would like to thank all the contributors of the short articles and especially the directors and Professors **Christoph Bussler and Dieter Fensel** of Digital Enterprise Research Institute (DERI) who contributed to the special section of DERI presentation. We do believe that you will be amazed by the quality of the work (done and in progress) at this Leading Institute.We would like to welcome our four new regular columnists: **Peter Alesso** *(Semantic Search Technology Column),* **Jessica Chen Burger** *(Semantic Web Technologies),* **Matteo Cristani** *(Methodologies for the Semantic Web Column), and* **Madhu Therani** *(Semantic Web Basics Column)* and to invite you reading their interesting columns.

In the second issue you will find two excellent interviews: Prof. *Christoph Bussler* from DERI and *Eric Miller*, Activity Lead of W3C Semantic Web Activity, comment on the current and the future role of the Semantic Web. Finally we would like to communicate one more interesting piece of news: Our SIGSEMIS track in ECIS-2005 (http://www.ecis2005.de/semantic.html), **Regensburg Germany,** for which you can find the relevant call for papers in Page 2.

We are looking forward to your active participation and collaboration to our initiative: SIG SEMIS is an open forum: We invite you to join us (http://www.aisnet.org/sigs.shtml) and to share your thoughts and perspectives.

We would like finally to wish all the best to the new President of AIS *Rick Watson* and to declare that we share and we are great supporters of his dream for Association for Information Systems (www.aisnet.org).

On behalf of SIG SEMIS Board,
Dr. Miltiadis D. Lytras
Athens University of Economics and Business
Department of Management Science and Technology
ELTRUN - The Research Center, URL://www.eltrun.gr

AIS SIGSEMIS ACTIVITIES

By SIG Board

ECIS 2005 Semantic Web and Information Systems Track
[http://www.ecis2005.de/semantic.html]

The 13th European Conference on Information Systems will be held in <u>Regensburg</u>, Germany. It is organized by the Institute for Management of Information Systems at the University of Regensburg. The ECIS will take place from May 26 to 28, the Doctoral Consortium from May 23 to 25.
[more info: http://www.ecis2005.de/index.html]

Track Chairs:
Gottfried Vossen, University of Münster, Germany
Miltiadis Lytras, Athens University of Economics and Business, Greece

Track Committee
Richard Benjamins (Intelligent Software Components), William Grosky (University of Michigan), Lakshmi S. Iyer (The University of North Carolina at Greensboro), Henry Kim (York University), Kinshuk (Massey University), Ralf Klischewski (University of Hamburg), Shiyong Lu (Wayne State University), Ambjorn Naeve (Royal Institute of Technology-KTH), Demetrios Sampson (University of Piraeus), Amit Sheth (CTO Semagix, University of Georgia), York Sure (University of Karlsruhe), Kim Veltman (Maastricht McLuhan Institute), Gerd Wagner (Eindhoven University of Technology), Lina Zhou (University of Maryland)

Call For Papers
The Semantic Web (SW) poses new challenges to Information Systems. A first observation concerning the current situation is that the field of SW is dominated by rather technical approaches exhibiting a lack of multidisciplinary contributions and insights. From this perspective this track attempts to fill this gap, with a special emphasis on demystifying the Semantic Web and revealing novel opportunities for value exploitation. With the common practice of considering the Semantic Web as a technology-driven phenomenon, we will contribute to a scientific debate, which reveals the practical implications and the research challenges of SW in the context of Information Systems. Our approach should go beyond the traditional research agenda of Information Systems and critical themes will be analyzed through a Semantic Web perspective in horizontal and vertical pillars. The main objective is to communicate high quality research findings in the leading-edge aspects of Semantic Web and Information Systems convergence. This statement distinguishes this track from traditional SW tracks: Traditionally, the Semantic Web is treated as a technological phenomenon with the main emphasis on technologies, languages and tools without similar attention given to theoretical constructions or linkages to multidisciplinary references: Our focus is on the Information Systems Discipline and we are working towards the delivery of the main implications that the Semantic Web brings to Information Systems and the Information/Knowledge Society.

Suggested topics:
- Semantic Web Issues, Challenges and Implications in each of the IS research streams
- Towards the development of the Knowledge society
- New Semantic Web enabled Tools for the citizen/ learner/ organization/ business
- New Semantic Web enabled Business Models
- New Semantic Web enabled Information systems and knowledge repositories
- Integration with other disciplines
- Intelligent Systems
- Standards
- Semantic enabled business intelligence
- Enterprise Application Integration
- Metadata-driven (bottom-up) versus ontology-driven (top-down) SW development

BEST PAPER will be invited for publication in AIS SIGSEMIS official peer reviewed journal: International Journal on Semantic Web and Information Systems.

An Interview with Eric Miller
Activity Lead for the W3C World Wide Web Consortium's Semantic Web Initiative

"Together, we make the Semantic Web a reality..."

Eric Miller

Semantic Web Activity Lead
W3C World Wide Web Consortium
200 Technology Square, NE43-350
Cambridge, MA 02139 USA
URL://http://www.w3.org/People/EM/
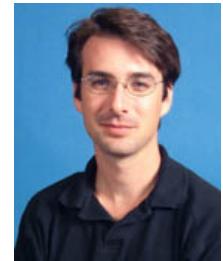
Miltiadis: Eric we are delighted you agreed to this interview. Let start by asking you to provide me your general idea on how W3C's SW Activity is going?

Eric: In short, fantastic! It's been a very exciting past few years for the Semantic Web Activity in terms of standards work, deployment and wide scale uptake of these technologies. There is much still to do, but the quality of the work, the passion of the people, and the benefits we're seeing for the Web make for an extremely exciting time.

> It's been a very exciting past few years for the Semantic Web Activity in terms of standards work, deployment and wide scale uptake of these technologies.

Miltiadis: What does your role as the Activity Lead for the W3C World Wide Web Consortium's Semantic Web Initiative involve?

Eric: My responsibilities include the architectural and technical leadership in the design and evolution of Semantic Web infrastructure. This involves working with W3C members so that both working groups in the Semantic Web activity, as well as other W3C activities, produce Web standards that support Semantic Web requirements. Additionally my responsibilities include fostering support among user and vendor communities for the Semantic Web by demonstrating the benefits and means of participating in the creation of a metadata-ready Web. And finally to establish liaisons with other technical standards bodies involved in Web-related technology to ensure compliance with existing Semantic Web standards and collect requirements for future W3C work.

Miltiadis: Your work in W3C concerning specifications is considered as of critical importance for the evolution of SW. In what phase are we now?

> *"We internally describe where we are at as 'Phase 2'. The focus of Phase 2 is 'wide scale deployment'.*

Eric: Good question. We internally describe where we are at as 'Phase 2'. The focus of Phase 2 is 'wide scale deployment'. The focus of the new working groups charted under the Activity, our education and outreach objectives along with our research and advanced development goals are all focused on taking the existing standards of XML, RDF and OWL and focusing on their effective deployment to enable the Semantic Web. Details on each of these are provided in some of your following questions.

Miltiadis: Can you demystify for SIGSEMIS members your organization and the work process of W3C SW initiative? How are your activities organized? How open is this SW initiative? How do you help this collaborative and intellectual process? How difficult is to manage all the communities of experts?

Eric: The goal of the Semantic Web initiative is as broad as that of the Web: to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. The W3C Semantic Web Activity has been established to serve a leadership role in both the design of specifications and the open, collaborative development of enabling technology. The Semantic Web Activity is focused on four areas to meet this goal.

> **" The W3C Semantic Web Activity has been established to serve a leadership role in both the design of specifications and the open, collaborative development of enabling technology. The Semantic Web Activity is focused on four areas to meet this goal".**

Enabling Standards: The Resource Description Framework (RDF) and the Web Ontology Language (OWL) are foundation standards for the Semantic Web. Additional standards however are necessary for using, querying and accessing data in a networked environment. Current working groups in Best Practices and Deployment and RDF Data Access are managed under the Activity and designed to facilitate Semantic Web development and ease the sharing of data located across distributed collections.

Education and Outreach: The Semantic Web lends itself to collaboration, teamwork, and cooperation. In addition to the Semantic Web Interest Group which serves as W3C's primary focal point for Semantic Web community discussion, there are a variety of domain specific communities who are using RDF/XML to publish their data on the Web. The Semantic Web Interest Group coordinates public implementation, shares deployment experiences of RDF, and helps promote the Semantic Web. These discussion groups provide valuable cross domain community input to current and potential future work associated with the Activity.

Liaison and Coordination: The Semantic Web Activity additionally managing the interrelationships and interdependencies among groups (both within W3C and outside the organization) focus on standards and technologies that relate to the goal of the Semantic Web.

Advanced Development: Just as the early deployment of the Web was accelerated by the availability of open-source code modules such as libwww, W3C is devoting resources to the creation and distribution of components to assist with deployment of the Semantic Web. We also use research funding to design and prototype pre-standards technologies that are part of a long-term vision. These technologies will eventually feed in to proposals for future Recommendation-track Working Groups. These advanced development initiatives are designed to work in collaboration with a large number of researchers and industrial partners to stimulate complementary areas of development that will help facilitate further deployment and future standards work.

Miltiadis: Is it fair to say RDF is a huge success? From the standards perspective, its broad acceptance should be highly satisfying to you-right? However, where do you think it is in its adoption process/cycle? Is it possible to find out products that deal with storage, querying and computation issues that are industry strength?

> *"I'm quite pleased with the success of RDF, but there is a lot of work still to do. In particular, work of hiding a lot of the complexity to the end user is an important one".*

Eric: I'm quite pleased with the success of RDF, but there is a lot of work still to do. In particular, work of hiding a lot of the complexity to the end user is an important one. Example such as Adobe's XMP or Creative Commons on the content creation side of the process I find quite encouraging but this is only the beginning. There are far too many products and toolkits to list, but recent work from HP, Oracle, IBM along with start-ups such as Tucana and Network Inference are clear indicators that industrial strength Semantic Web products and toolkits are available now. .

Miltiadis: What is your prognosis on OWL? Clearly the researchers, esp. those from logic or AI background love it. What do you think has to happen for mainstream product developers and application developers to either use or benefit from having a standard such as OWL (when the objective is to develop/support semantic applications)? Can OWL be as successful or influential as RDF is destined to be?

Eric: I'm actually very encouraged by the uptake of OWL as well. The Semantic Web Activity worked very hard to try and provide a layered approach in functionality among RDF, RDF Schema and OWL. I believe this effort will be a big win for both RDF and OWL and both will benefit. While RDF provides simple, flexible description framework for the Web, OWL builds on RDF to provide richer descriptive capabilities within Web communities. While RDF enables discovery and reuse of data across domains, OWL enhances discovery and reuse of data within domains. I see RDF and OWL as complementary in nature and both successful in enabling the Semantic Web.

The main differences I see among the views of the leaders you mention however are not necessary with the vision of the Semantic Web, but rather with how to most effectively achieve the goal.

Miltiadis: Given your influential role as an activity lead, I am sure you have interacted and collaborated with lead thinkers of SW, like. Tim Berners-Lee, Jim Hendler, Amit Sheth, Ora Lassila, Ian Horrocks, and several others. Do you think there is a unified vision? Also, I would guess you carry a heavy load in terms of translating a vision into reality. Do you try to modulate the vision with practical considerations? Anything you wish to share about the vision versus where we are in the process of realizing it?

Eric: In general there are many views of the Semantic Web just as there are many views of the original Web. The main differences I see among the views of the leaders you mention however are not necessary with the vision of the Semantic Web, but rather with how to most effectively achieve the goal. The task of translating this vision and different perspectives into a reality is a certainly a difficult one. But the exciting aspect of this is that they (along with many many others) have all come together at W3C to actively working together to see the Semantic Web a reality. This is one very powerful indicator of success.

With respect to where the Semantic Web is at regarding process of realizing the vision, we've come a very long way since the original Activity was launched. If we think back to the phases associated with Web deployment

> **1. The Web was born at CERN**
> **2. Was first picked up by high energy physicists**
> **3. Then by academia at large**
> **4. Then by small businesses and start-ups**
> **5. Big business came only later!**

I'd suggest the Semantic Web is now at #4, and very quickly moving to #5. From my perspective, this is quite an impressive accomplishment and reflects the quality of work from all of the people involved.

Miltiadis: Eric, It is possible to identify for our members some of the hot topics in your research agenda in W3C SW initiative. Do you have a specific timeline for their achievement?

"If we are going to entrust more of our knowledge to the Semantic Web, we must be assured that the Web will respect many more of the social agreements that we enforce in the physical world".

Eric: There are many. Here, however, are but a few high level hot topics that are on the W3C's Semantic Web research agenda.
Creating a Policy Aware Infrastructure: *The development of a Policy Aware Infrastructure for the Web is required. The Semantic Web will only achieve its potential as an information space for the free flow of scientific and cultural information if its infrastructure supports a full range of fine-grained policy controls over the information contained in the Semantic Web. If we are going to entrust more of our knowledge to the Semantic Web, we must be assured that the Web will respect many more of the social agreements that we enforce in the physical world. For the Semantic Web includes not only freely available information, but also personal information and information available to a person or agent only as a result of its membership in groups. A policy-aware infrastructure -- one that gives information creators and users the types of control over information we have all become accustomed to in the physical world such as the ability to assert and exercise privacy and intellectual property rights -- will make the Semantic Web into a vibrant and humane environment for sharing knowledge and collaborating on wide range of intellectual enterprises.*

Ontological Evolution: An important goal of the Semantic Web is to address the problem that in the course of scientific (or any) endeavor, one changes the vocabularies one uses to organize, discover, and communicate. A given vocabulary may be refined, resulting in a need for migration from old to new. Communication between distinct groups using different vocabularies creates the need to create common vocabularies which optimally suit all involved. Semantic Web techniques should make this difficult process of creating new common vocabularies as easy as possible. The Semantic Web already removes confusion by giving each term a globally unique URI. OWL ontologies and rules languages allow relationships between old and new terms to be expressed. There is, however, little experience with the serious management of such evolution. The Semantic Web needs to incorporate versioning and provenance within its foundation. Human understanding changes and statements that we once thought were accurate are later described to be inaccurate. However, the original statement should not be deleted from our corpus of human knowledge. The Semantic Web should not be required to forget that a statement was once believed to be a true statement. Versioning is such a common approach to representing discrete states of understanding that it warrants explicit treatment in the Semantic Web.

> *"The Semantic Web needs to incorporate versioning and provenance within its foundation. Human understanding changes and statements that we once thought were accurate are later described to be inaccurate".*

Web of Trust: Trust in the human social context is based on constantly evolving and adapting information. Two parties may trust each other based on a history of mutual interaction, based on formal contracts that in turn rely on other established systems (e.g. legal and legislative), and based on risk analysis of a failure of any party to perform as agreed. A trust language for the Semantic Web that is capable of representing these complex and evolving relationships will be crucial to our future ability to build software that behaves more in the manner of an intelligent assistant than a rote rules processor.

Information Flow and Collaborative Life: Many tools used with collaborating groups today instrument the flow of data, information, and knowledge. One of the challenges we will meet is to strike a balance between requiring authors to do more at the outset to make information machine processable, insisting that everything the machine could use to answer a question be recognized and identified by the (human) questioner, and leaving large quantities of information inaccessible to the machine. There are many more, but I think I'll save these for another time. :)

Miltiadis: What skepticism you have seen in industry with respect to broader acceptance of SW related standards, techniques and tools? How do you help address theses? Based on your experience could you outline mechanisms that you have found as effective for the specification and the promotion of standards?

Eric: An interesting transformation that is occurring in many sectors of industry in the recognition that the data that is being created by applications is far more valuables than the applications themselves. Freeing the data from the applications that created them and managing this information directly relates to a strong return on investment. The predominant skepticism I hear is perhaps the most is 'if I have XML why do I need RDF'. It's interesting however to see some of the skepticism dissipates after organizations learn from experiences (often times painful ones) that agreement on syntactic conventions are often overly brittle and not adequate for the effective management of data.

Miltiadis: Several times all these specification activities for the huge mass of people are considered as intensive technical oriented processes with limited interest. Having in mind that people are interested in services and value adding processes, what W3C SW initiative is targeting now? Could you outline some services?

Eric: It's easy to only focus on the standards and miss the larger picture of what these standards are trying to enable. The key is to focus on what it means to various end users communities and what sort of benefits these technologies provide. I think simple services are the best at conveying this point. For example, imagine a service that auto-classifies a document and returns a set of topics in RDF about that resource. Next, imagine that that each of these resources is a news items. Merging the news item metadata with the subject classification metadata allows an end user to organize news items now by Subject rather than simply the channel in which he/she is subscribed. Very simple, very easy and with immediate benefit. There are hundreds of these kind of value added services that will help enable the Semantic Web.

> *It's easy to only focus on the standards and miss the larger picture of what these standards are trying to enable. The key is to focus on what it means to various end users communities and what sort of benefits these technologies provide.*

Miltiadis: Dear Eric, it is obvious that there would be different views about SW given its importance and scope. I would like to copy to you a small portion of an opinion of an interesting thinker John Sowa: " In 6 years (1998 to 2004) with ENORMOUS hype and funding, the semantic web has evolved from Tim BL's book to a few prototype applications, which are less advanced than technologies of the 1970s such as SQL, Prolog, and expert systems -- and they're doing it with XML, which is far less advanced than LISP, which was developed in the 1950s. This contrast does not give me a warm, hopeful feeling about the semantic web..." What would you answer to such a critique?

Eric: Part of what happens in development of Web applications is for technologies to inherently become less advanced, but dramatically scale up in terms of use and deployment. Web Services, for example, is less advanced than CORBA and OMG's related work, but it can be used on a widely deployed infrastructure, hence its success. The Semantic Web is similar in this regard in that its building from existing domain specific solutions and weaving these into the Web. The Web is inherently an open system where anyone can say anything about anything. RDF and OWL are declarative means of describing knowledge to be shared on the web. How an application chooses to manage this data (SQL and relational databases, Prolog, etc.) is done in the privacy of one computer. The key to remember however is the enormous benefit in being able to provide common protocols for accessing this information and means of writing this down in a common way that others can benefit from. The goal we're after is to making simple things simple and the complex things possible in an open Web.

> *"The key to remember however is the enormous benefit in being able to provide common protocols for accessing this information and means of writing this down in a common way that others can benefit from. The goal we're after is to making simple things simple and the complex things possible in an open Web".*

Miltiadis: And what about very different form of critique from those who feel overwhelming dependence on DL is a problem as few real-world programmer (i.e., those who do VB, C and Java) will ever use it and if they thy to use it, they will face the long standing performance and scalability problems

Eric: Software is layered. Programmers, who write VB, C, and Java, never directly use a whole load of complex services (network management, database access, numerical calculation) except through well defined interfaces. I expect a similar pattern for accessing Semantic Web libraries. In fact, we're already seeing this in many applications.

Miltiadis: A few days ago I had a conversation with a colleague in the university. He claimed that W3C for him is an organization with solid technical orientation. But in the other hand I do believe that you are really business oriented? What you think?

Eric: W3C started with a solid technical focus. As the Web evolved so did the W3C. Without loosing its technical focus, it has grown to gather expertise in business, social impacts of technologies, usability, legal and many others as needed. The most amazing thing to me is to see all of the experts at W3C interact as a team to lead the Web to its full potential. It really is a remarkable organization full of people around the world who care deeply about what they do.

> *In Chicago, it seems to me this was a turning point as everyone who attended realized the Web was not a fad, but rather something that was going to revolutionize how we communicate. The WWW2004 conference had a similar impact on me with regards to the Semantic Web. The technologies and toolkits are maturing.*

Miltiadis: *Besides standardization related activities, what else do you coordinate or direct? Very recently you had organized the Developer Day at WWW2004. Any takeaway from that? Any thing to show where we are in terms of maturing of the technology, adoption by the industry, contribution of academics in terms of making vision more realistic or practical, or providing tools and test suites for developers to use?*

Eric: *Besides the enabling standards related work, I also direct areas of Advanced Development work and Education and Outreach related to the W3C Semantic Web Activity. You mention WWW2004. Coordinating W3C Semantic Web panel and Developers day at WWW2004 in particular was an interesting experience this time around. The WWW2004 Web conference had a huge Semantic Web focus that permeated almost all aspects of the conference. The energy at the meeting, the collaboration occurring in the corners and though out the night, reminded me of the second Web conference in Chicago. In Chicago, it seems to me this was a turning point as everyone who attended realized the Web was not a fad, but rather something that was going to revolutionize how we communicate. The WWW2004 conference had a similar impact on me with regards to the Semantic Web. The technologies and toolkits are maturing. Semantic Web applications are becoming far more prevalent. Novel ideas for how these technologies may be used are happening on a daily basis. It was quite a week!*

Miltiadis: Undoubtedly a lot of things have to be done in order to realize the full potential of SW. Towards this direction Academia and Industry are looking for converging synergies. Could you share with us your experience from academia and industry collaboration in your SW initiative?

Eric: *I believe that fostering converging synergies between Academia and Industry is a key ingredient for realizing the full potential of the Semantic Web. Several synergistic projects have already confirmed this for me. Project SIMILE for example is a joint project conducted by the W3C, HP, MIT Libraries, and MIT CSAIL with a focus of applying Semantic Web technologies to the digital library and personal information management space. SWAD-Europe is another Academia and Industry supported project which aims to support W3C's Semantic Web initiative in Europe, providing targeted research, demonstrations and outreach to ensure Semantic Web technologies move into the mainstream of networked computing. The lessons learned from theses (and other) projects help shape future items for the Semantic Web and provide real world demonstrations of these technologies and open source code to help bootstrap a network effect. It's very much a symbiotic relationship that benefits all.*

> *"I believe that fostering converging synergies between Academia and Industry is a key ingredient for realizing the full potential of the Semantic Web. Several synergistic projects have already confirmed this for me"*

Additionally, we've recently announced a W3C Workshop on Semantic Web for Life Sciences to be held in October which I believe will be an important additional step in converging synergies in Academia and Industry sectors associated with Life Sciences with Semantic Web. It is in part through these collaborative and synergistic projects that the Semantic Web is enabled.

Miltiadis: You are also a Research Scientist at MIT's Computer Science and Artificial Intelligence Laboratory. What is the research culture there?

Eric: *The collaboration with individuals of MIT's Computer Science and Artificial Intelligence Laboratory have been second to none. There are many benefits of such collaboration, but one that immediately comes to mind is Haystack which is focusing on a universal information client based on RDF. This collaboration as been extremely enjoyable and I'm constantly impressed with the myriad of extremely creative ideas for how to use Semantic Web every time I get to sync up with this team. It's an incredible experience.*

Miltiadis: I interact with people from research institutes worldwide. A general conclusion is that they all share a very optimistic vision for the role of new technologies and but also have a list of challenges and possible pitfalls. What are the major problems that you see in the promotion of the SW and your outcomes in W3C?

Eric: I think the major problems we'll be facing will not be technical but rather social in nature. The Semantic Web enables a transparent society, which will benefit from a policy aware infrastructure.

Miltiadis: I noticed that SW initiative is included in The Technology & Society Domain of W3C. Could you outline some societal concerns of your initiative (e.g. disabled people, etc)?

Eric: The Semantic Web is indeed part of the Technology & Society Domain of W3C because we believe in both the technical aspects and social implications (privacy, security, access control, etc.) associated with this work.

Miltiadis: Eric, I think that several people would be interested in participating in your activities. Is there a formal procedure to join SW initiative? Do you have open calls for candidates in your working groups?

Eric: The Semantic Web Interest Group is open to all and a good place for people to go to share ideas, thoughts and experiences related to the Semantic Web. The enabling standards associated with the Activity are done in Working Groups. When a working group is formed there is a Call for Participation. Any W3C member can nominate a representative to work on a working group. I'd be happy to talk with any reader interested in learning more about the necessary steps for participating in such work.

> *"The enabling standards associated with the Activity are done in Working Groups. When a working group is formed there is a Call for Participation. Any W3C member can nominate a representative to work on a working group. I'd be happy to talk with any reader interested in learning more about the necessary steps for participating in such work".*

Miltiadis: Any thoughts you would care to share on the formation of the New Special Interest Group on Semantic Web and Information Systems on AIS?

Eric: I think this group formation is a very good idea and I whole-heartedly encourage this to happen. My only suggestion would be to have a goal for addressing community needs, focus on demonstrations that help people understand the benefits from these technologies and please, by all means, share your experiences with others on the Semantic Web Interest Group list! :) There are many other people outside AIS that I'm sure would be interested in your findings.

Miltiadis: Dear Eric thank you for your time. It was an excellent talk. Would you like to give parting thoughts to our readers?

Eric: I was involved at the early phases of the web. I feel that it has had a significant impact on me personally. If you are not involved, get involved! Together, we make the Semantic Web a reality.

NOTE: I would like to thank Prof. Amit Sheth for contacting me with Eric Miller and helping me in organizing the interview especially.

Eric Miller (http://www.w3.org/People/EM/) is the Activity Lead for the W3C World Wide Web Consortium's Semantic Web Initiative.

Eric's responsibilities include the architectural and technical leadership in the design and evolution of Semantic Web infrastructure. Responsibilities additionally include working with W3C Working Group members so that both working groups in the Semantic Web activity, as well as other W3C activities, produce Web standards that support Semantic Web requirements. Additionally, to build support among user and vendor communities for the Semantic Web by illustrating the benefits to those communities and means of participating in the creation of a metadata-ready Web. And finally to establish liaisons with other technical standards bodies involved in Web-related technology to ensure compliance with existing Semantic Web standards and collect requirements for future W3C work.

Before joining the W3C, Eric was a Senior Research Scientist at OCLC Online Computer Library Center, Inc. and the co-founder and Associate Director of the The Dublin Core Metadata Initiative, an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models.

Eric is a Research Scientist at MIT's Computer Science and Artificial Intelligence Laboratory.

Announcing our SIG sponsored International Journal: International Journal on Semantic Web and Information Systems

By SIG Board

Headline

In last week IDEA Group Publishing accepted our proposal to publish in print our SIG Sponsored Peer Reviewed International Journal entitled *International Journal on Semantic Web and Information Systems.* With a clear publication strategy and a solid interest in the promotion of Knowledge Society through semantic web, our journal will be our main communication channel for research outcomes. The cultivation of the Semantic Web Vision in Information Systems Research Community will require diversified means. The IJSWIS journal, through its capable editorial team and its direct linkages to IS community will substantially initiate scientific discussions on critical issues that correspond to the ultimate objective to express and exploit meaning through information systems. In the next pages we outline the purpose, the objectives and the topics covered by the new journal and we introduce the editorial team. In the 2nd issue of our newsletter more information will be provided concerning our Journal.

The overall mission of this journal

The International Journal on Semantic Web and Information Systems is an open forum aiming to cultivate the Semantic Web vision within the Information Systems research community. In the common practice of anticipating Semantic Web as a technology- driven phenomenon, we provide a scientific insight, which reveals the practical implications and the research challenges of SW in the context of Information Systems. Our approach goes beyond the traditional research agenda of Information Systems and critical themes are analyzed through a Semantic Web perspective in horizontal and vertical pillars.

The main idea is to communicate high quality research findings in the leading edge aspects of Semantic Web and Information Systems convergence. This statement distinguishes our journal and differentiates our publishing strategy from other publications: Traditionally Semantic Web is treated as a technological phenomenon with the main emphasis on technologies, languages and tools without similar attention given to theoretical constructions or linkages to multidisciplinary references: *Our focus is on the Information Systems Discipline and we are working towards the delivery of the main implications that the Semantic Web brings to Information Systems and the Information/Knowledge Society.*

*Figure 3. The motivation for the IJSWIS as a value chain*
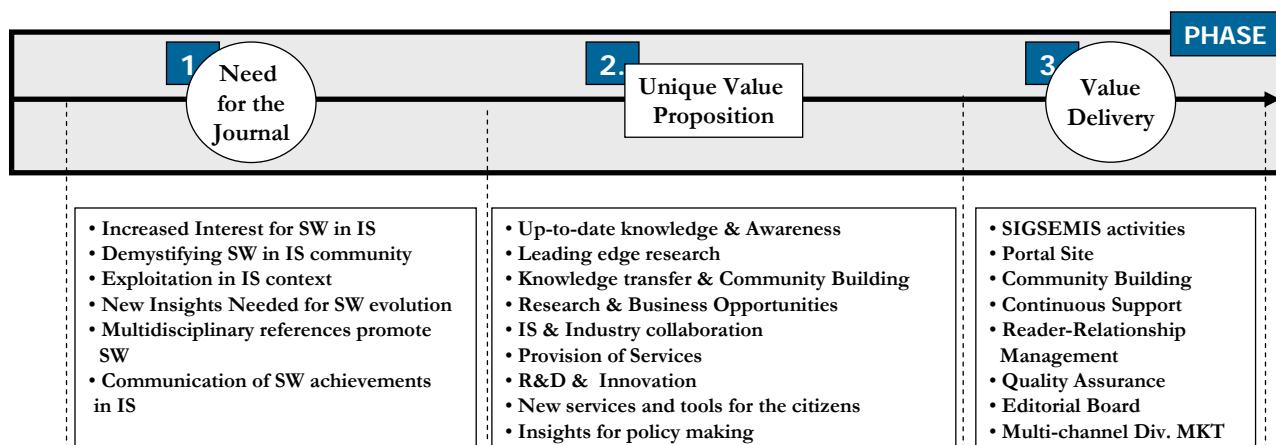
A new journal definitely has to answer three critical questions:

1. What is the need for the new journal?
2. What is its unique value proposition for relevant target audiences?
3. Which strategy will support the publication process towards the development of a branded, recognized and high ranked journal in the major scientific area that contributes?

We will try in the next paragraphs to provide a thorough argumentation with respect to the above questions. Figures 3 & 4 depict the main points and summarize key issues.

The motivation for the new journal is derived from two facts: First of all, during the last years there has been a tremendous evolution of the Semantic Web, which has developed a great potential for its future role. Additionally, the IS research community requires a research forum and a publication outlet which will pursue the simplification and the promotion of SW for its specific characteristics. The convergence of SW and IS provides an excellent research context which is not covered from other initiatives that pay an enormous attention to themes related to Technology or Artificial Intelligence. ***The starting point for the justification of our journal is that the IS research community can play a much more active role in the research discussion of the Semantic Web and our journal intends to be the main and dominant dialogue channel for this process***.

Figure 4. The Unique Value Proposition of the journal
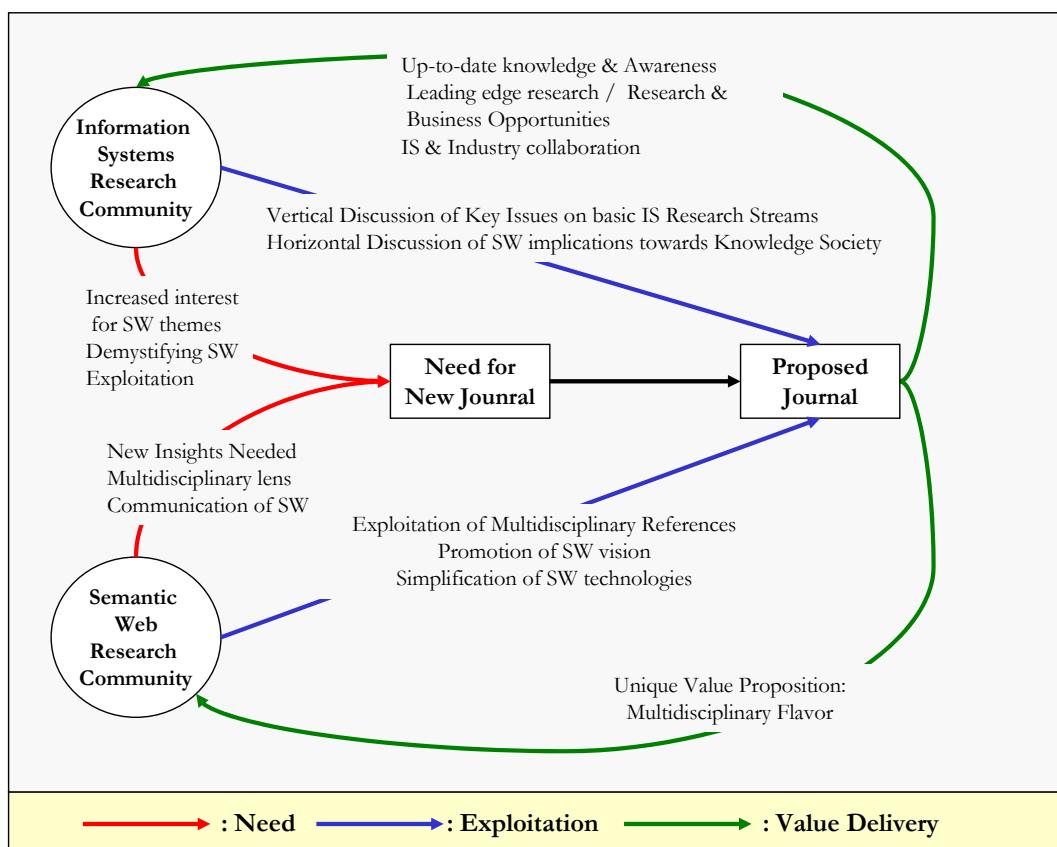
Our journal targets on the convergence of Semantic Web and Information systems and this strategically interesting synergy is going to be exploited for the provision of a unique value proposition to Academics, Industry and Government. High Quality Assurance, Readers/authors-Relationship Management, Fully Integrated IT support as well as communication of leading edge research and opportunities will contribute to a powerful value mix.

The third question refers to the critical strategy that will be diffused in several actions and practices. The key issue is the differentiation of our proposition in comparison with other initiatives. An integrated publication strategy is pursuing the high quality aspect of our proposition and an integrated communication marketing process works towards a 1-1 relationship management. The Association for Information Systems provides an excellent starting point for the building of a recognized and branded journal.

Our goal is to contribute in Theory, Practice and Methodology of Information Systems through the following integrated approach:

(i). **THEORY:** The development of theory in the convergence of IS and SW is organized on Key Themes such as:

- Information Systems Discipline
- E-business
- Knowledge Management
- E-learning
- Business Intelligence
- Organizational Learning
- Agents
- Adaptive Systems
- Enterprise Application Integration
- E-government
- Mobile and Wireless Technologies
- Decision Making
- Database systems
- Impact of XML
- Standards such as RDF and OWL
- Ontologies, their design and exploitation
- Process orientation
- HCI
- Multimedia
- Semantic interoperability
- Human-Machine semantic mediation

Our Goal and strategy is to prepare for all these themes, special issues that will discuss how Semantic Web poses new challenges and a new research agenda for each vertical IS theme and how it shows promise to deliver impact. Moreover the goal for each special issue will be to promote innovative propositions.

In knowledge management, the overall paradigm shift that is enabled by Semantic Web technology can be described as a transformation from knowledge-push to knowledge-pull. This is achieved by opening up previously hidden sources of information that can be cross-searched and combined in well-defined and machine-processable contexts as well as configured and controlled by each individual user.

This paradigm shift is manifesting itself as a fundamental transformation of many different fields, e.g. from teacher-centric to learner-centric e-learning, from producer-centric to consumer-centric e-business, from doctor-centric to patient-centric e-health and from authority-centric to citizen-centric e-government. Our journal will focus on making the implications of these structural changes visible and understandable to a non-technical audience, so that they can take a more active part in the discussion on how these new possibilities should be exploited in order to optimize the benefits for society as a whole.

(ii). **PRACTICE:** A core strategy in our journal is to pay close attention to the simplification of what Semantic Web means in practice. According to this, published articles will provide added value to readers by answering key questions for new semantic web enabled information systems.

(iii). **METHODOLOGY:** The research orientation of AIS as well as our SIG SEMIS research role requires the development of methodological guidelines for conducting research in Semantic Web related themes. Thus issues such as epistemology and Research Methods for Semantic Web and Information Systems research are of critical importance for our journal.

As a concluding remark our journal has a clear ultimate goal:

To provide awareness, new knowledge and significant insights within the IS research community for the current and the future role of Semantic Web, recognizing that technology is the facilitator and not the ultimate goal. Thus concepts, models, theoretical propositions as well as real world examples and innovative case studies bring forward the capabilities of the Semantic Web and the evolution of the New Generation Information Systems.

The overall scope of this journal
International Journal on Semantic Web and Information Systems promotes a knowledge transfer channel where academics, practitioners and researchers can discuss, analyze, criticize, synthesize, envisage, realize, communicate, elaborate and simplify the more than promising technology of the Semantic Web in the context of Information Systems. In the ancient Greeks' rhetoric semantic was the ultimate milestone in the quest of human mind to create and to communicate meaning.

Figure 5. The overall scope of the IJSWIS
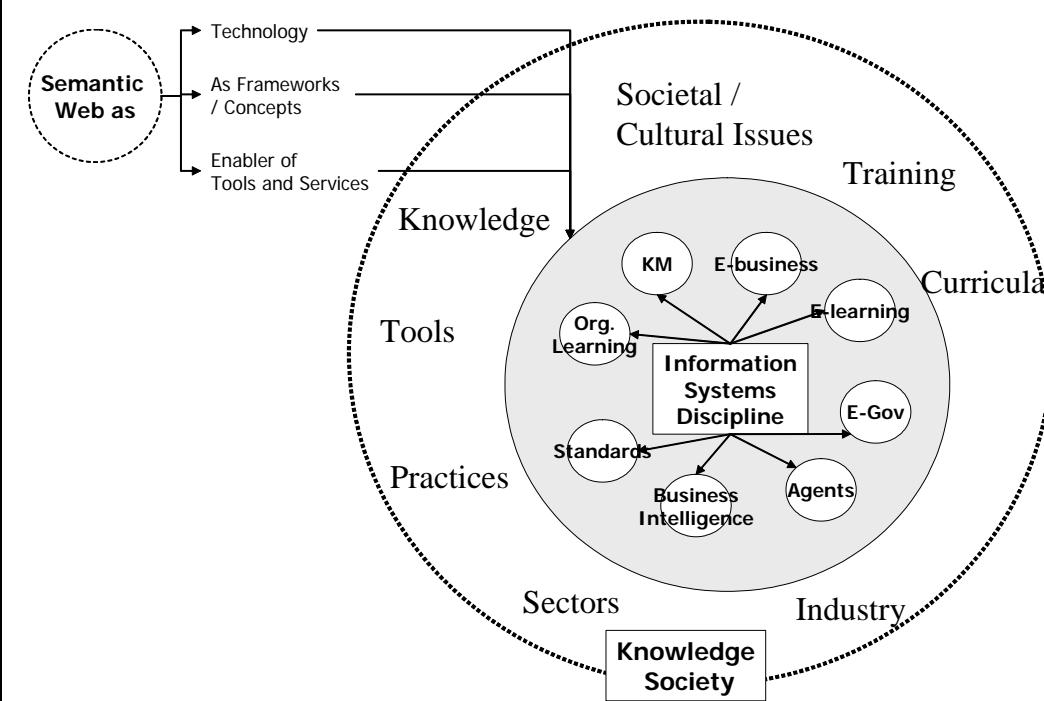
In the Knowledge Society the exploitation of knowledge requires integration, intelligence, flexibility as well as accuracy and reference layers that are held together by a simple and ultimate goal: To utilize available technologies towards the "effective" knowledge representation and retrieval. But this fact is the basis of Information Systems: From an ontology perspective we develop conceptualizations and from these we build systems, services and practices that require an effective mix of information technology, processes, people and applications. The scope of our journal is to discuss the Semantic Web as an indissoluble whole of Technologies, Frameworks, Concepts and Practices that enable tools and services capable of supporting new innovative, effective and feasible information systems.

Figure 5, provides a graphical overview of the overall scope of IJSEMIS. Three areas are in a continuous interchange and provide value through dynamic flows and exchanges, namely the Information Systems Discipline, the IS Research Streams and the emerging Knowledge Society.

The inner circle provides a main research objective for the AIS Special Interest Group on Semantic Web and Information systems. The extensive discussion of issues and the production of new knowledge is related to the semantic web implications in main IS research streams such as E-business, Knowledge Management, E-

learning, Business Intelligence, Organizational Learning, Agents, Adaptive Systems, Enterprise Application Integration, E-government, Decision Making, HCI, Multimedia etc. We will encourage special issue proposals for most of <u>the above 3 themes</u> and we believe that the capacity of the SIG Board members as well as the contribution of excellent Guest Editors will provide a competitive advantage for our journal. This internal web provides a solid orientation for our publishing strategy but represents only the basic layer of our scope and potential contribution.

The outer circle represents another interesting aspect of the journal's overall scope. Our vision for the cultivation of the Semantic Web vision in IS, is closely related to the so-called Knowledge Society, where transparent technology provides new tools, services for the citizens, learners, disabled people as well as for businesses, organizations and government.

From these perspectives several more pillars outline the scope of our journal:
- Presentation of new semantic web enabled tools depicting the power and the capabilities of the Semantic Web.
- Sectoral/Industry Analysis of Semantic Web enabled IS (in Tourism, Knowledge Intensive Organizations, Commerce, etc).
- Discussion of innovative training activities concerning SW (community building, knowledge transfer) as well as thorough discussion on the critical theme for the inclusion of SW in IS curricula.
- Forum for epistemological, societal and cultural issues that are affected by the Semantic Web.
- Discussion of leading edge research in knowledge representation and retrieval onfive levels: Artifact, Individual, Team, Organization, Network

The above description indicates the multidisciplinary flavor of our approach. We don't believe that the SW is a solid AI issue nor that it is a "Database"-driven phenomenon. In our perception, the Semantic Web is a milestone on the road towards the ultimate human quest for efficient knowledge heritage. From philosophy we derive the axioms of dialectic and we will work hard to develop a multidisciplinary and evolutionary journal of high quality.

Possible topics to be covered by this journal

The Main themes covered in the journal include:

Figure 6. Main topics in the IJSWIS

i. Semantic Web Issues, challenges and Implications in each of the IS research streams (some indicative are presented in figure….
ii. Real world applications towards the development of the Knowledge society
iii. New Semantic Web enabled Tools for the citizen/ learner/ organization/ business
iv. New Semantic Web enabled Business Models
v. New Semantic Web enabled Information systems
vi. Integration with other disciplines
vii. Intelligent Systems
viii. Standards
ix. Semantic enabled business intelligence
x. Enterprise Application Integration
xi. Metadata-driven (bottom-up) versus ontology-driven (top-down) SW development
xii. From e-Government to e-Democracy.
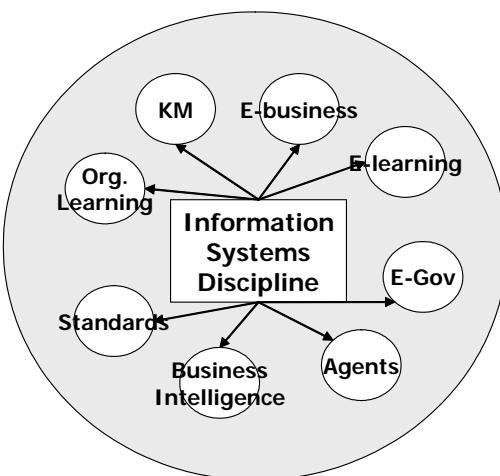
Paper Submission guidelines and Evaluation process

The integrated Publication strategy is depicted in figure… Four pillars and a stream of enablers collaborate towards a high quality journal.

The first step in the process is the securing of **Qualitative Contributions**. For this purpose, 4 main streams of activities will facilitate a push strategy:

- Personal invitations to top academics and researchers will be used in the first issues in order to communicate our journal to top researchers and potential contributors.
- Special issues from SIGSEMIS Minitracks in international conferences of IS (ECIS, AMCIS, ICIS).
- Relationship building with Competence - Research centers and Leading IT companies in the semantic web field.
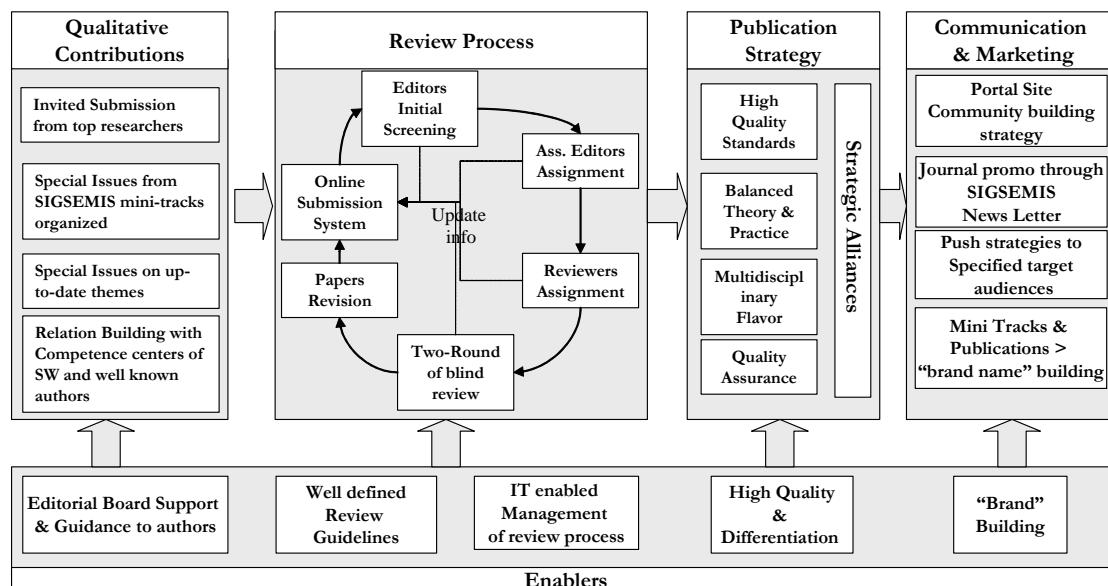
Figure 7. The integrated Publication Strategy of the IJSWIS

The Review Process will be another critical part of our strategy for competitive advantage and value adding services. A life cycle with distinct phases and an integrated support from an integrated IT infrastructure will secure a neutral, and constructive review process. All submissions will be made through the integrated Review System and will be forwarded to one of the editors. An initial screening will be used in order associate editors with relevant papers according to their main themes.

Associate Editors will use the system to assign 3 reviewers, which through a two round blind review will provide constructive comments. Associate editors in close collaboration with the editor(s) will decide on the issue of final acceptance.

Publication Strategy: The SIGSEMIS board consists of high quality academics and an excellent advisory board, which will support the journal in every stage of its preparation and development. The publication strategy has been crafted from a close understanding of the demands in the IS community for accurate and leading edge research on Semantic Web. The main emphasis is paid on quality assurance and a balanced discussion of theoretical and practical discussion of SW for IS.

Communication and Marketing: Several communication channels and a diversified marketing mix will be used in order to reach the relevant readership segments. Four main "vehicles" will be the carriers of our value proposition. First of all the Portal Site at www.sigsemis.org will support a community building strategy for all the readership audiences as well as authors and key people in the SW. A significant role in the promotion of the journal will be played by the SIGSEMIS Σigma Newsletter that will not only advertise the forthcoming

journal issues but will also host interviews with forthcoming contributors aiming to foster the interest in upcoming journal issues. Finally several other activities of SIGSEMIS such as Mini Tracks organizing and book editing and publishing will pursue the "brand" building process for our journal.

The contributions will fall into 6 different categories:
Table 1. Contribution Types

Contribution Type	Length	Key Objective	Evaluation factors
Full Research Papers	*4000-5000*	*Presentation of Research outcomes*	*Theoretical Background (20%)* *Significance of propositions (40%)* *Quality of writing (20%)* *Discussion of implications (20%)*
Research in Progress papers	*3000-3500*	*Outlining interesting future research outlets*	*Theoretical Background (30%)* *Methodology outlined (30%)* *Research Problem Description (20%)* *Quality of writing (20%)*
Literature Review Papers	*5000-7000*	*Intensive Critiques of literature / Gaps for possible research*	*Theoretical Background (40%)* *Critical thinking (20%)* *Discussion of gaps in theory (20%)* *Quality of writing (20%)*
Case Studies	*4000-5000*	*Discussion of real world implementations*	*Research Issues (30%)* *Promotion of theory & Practice (30%)* *Discussion of outcomes (20%)* *Quality of writing (20%)*
Critique of Clusters of SW projects	*5000-7000*	*Evaluation of outcomes*	*Methodologies used (50%)* *Discussion of Performance Gaps (30%)* ***Quality of writing (20%)***
Visioning papers	*4000-6000*	*Crafting roadmaps for the future*	*Innovation (50%)* *Theory and Technology exploitation (20%)* *Quality of writing (20%)*

Danny Ayers, http://dannyayers.com

The Missing Webs

It isn't too controversial to suggest that the Internet could be a lot more useful than it is today. Where there is difference in opinion is in the best approach to take to improve things. One possible route forward is that of the W3C's Semantic Web initiative. Here I will try to express why I believe it is the most constructive route forward. I'll start with a list of aspects of the Web that could be said to be lacking.

Navigable Web

There is a huge amount of information on the Web, but it's is of limited use without it being possible to access that information with ease. Compared to traditional systems the current Web is closer to a filesystem than a relational database. We have a means of storing and labelling the documents, what we don't have is any built-in technique for indexing and searching them. Catalogue-styled portals do help, and search engines like Google are extremely good at finding a needle in a haystack.

The hierarchies of catalogue portals and Google point to ways in which information can be more efficiently retrieved. Many portals are built from taxonomic hierarchies, in effect metadata-based navigation. Google's PageRank system uses the metadata of hyperlinks and the implicit metadata of the statistical occurrence of words to build its indexes. Portals general use rigid taxonomies, which have the drawback of inflexibility, and Google reveals the imperfections of a loose, statistical approach. The structural data of portals can be seen as hard facts, and although Google may take advantage of machine learning-style numeric techniques, those are also applied to (suitably organised) linkage facts.

Data Web

The current Web is primarily a very large number of hyperlinked documents. Whether they're written in loose HTML or more controlled XHTML format, these documents are designed for human reading. The intended path of use goes directly from the organised bits of data through a renderer to the end user. The Web is currently closer to a microfiche repository with an optical viewer than a knowledge representation system. But much of the information on the world's computers isn't in this form; it exists as chunks of information relating to real-world or abstract notions and the relationships between these chunks. As a generalization it could be called relational data, and in fact much of it is stored in quasi-relational SQL databases. But on the current Web data like this is only usually available through very narrow, human-oriented interfaces. A lot of data held by companies and other organizations will be commercially or politically sensitive, and would need to be kept private. But a considerable proportion of it could be made available more widely to general benefit. Given a framework that supports differing levels of access control, data could be published anywhere in between the private/public extremes.

Trusted Web

If information relating to the source of information can be reliably managed, then this opens up potential in several directions. Being sure of aspects like 'who asserted' and 'when' related to facts enables any conclusions inferred from statements based on those facts to carry some of that assurance.

Dynamic Web

A visitor from another planet might be forgiven for thinking that computers are solely communication devices. Apart from infrastructure wiring, the Web barely acknowledges that computers are good for *computing*. To take the computing model beyond the isolated mainframe or desktop PC requires integration of software across organization and even application boundaries.

Transparent Web

The Web Service approach of passing messages between systems offers a partial solution to making the Web more dynamic. For example, material contained in relational databases can be exposed, so their information becomes as available as that of published documents. But as already noted, the interface tends to be narrow. A database of a hundred tables, a thousand columns and a million rows may appear on the Web as a single node through which queries have to be tunnelled. For efficient interaction between end users and services and between services, a level of transparency is needed in which parcels don't have to be opened to discover their contents.

Ubiquitous, User-Friendly Web

Currently most access to the Web takes place through PCs or laptops. There has been some extension into smaller mobile devices, as well as TV-based systems. Wireless has also helped to break some physical restrictions. But still these are still relatively specialized interfaces; access is far from being on hand everywhere to everyone. In terms of user-friendliness, the Web is generally accessed through a HTML-oriented browser. This usually means read-only access, in very limited single mode of interaction. It lags far behind what is expected of "fat" desktop PC applications. Ubiquity and user-friendliness are key to humanity getting the maximum benefit of the Web, for people to have their abilities augmented at individual and societal levels.

Unified Web

The connectivity of the Web occurs at the level of hyperlinks, in effect the only shared languages are fairly low-level protocols. For the Web to be really useful, more sophisticated connectivity is needed. This requires language to describe the entities involved and the relationships between them. Given the scale and diversity of information sources, whatever language is used must be applicable in a very generic way. The only languages that are likely to fit the bill are mathematical, and the prime contenders are understandable in terms of first-order logic.

Evolving Web

Having suggested areas where the current Web is weak, it's worth stating something that is obvious, but often isn't taken into account. The Web as a whole is a dynamic system, it is developing year on year, day to day.
Most approaches to a future Web assume is that its information can be marshalled through the use of common languages, common specifications, standards. This certainly makes a lot of sense, for communication to take place common languages are needed. But an associated assumption is that good maintenance of these standards by organizations will spawn interoperable systems. This overlooks the fact that the world is full of hard-pressed developers looking for quick solutions to immediate problems. Having a solid specification doesn't mean it will necessarily be followed, especially when following it would mean more work in the short term for only hypothetical gains long term. But quick-and-dirty solutions aren't entirely a bad thing.

Emergent Web

Many of these loosely specified, limited application formats and protocols as avenues for progress are in themselves likely to be cul-de-sacs. But they appear in the global environment, and as it has been put in the Linux development community: "Given enough eyeballs, all bugs are shallow". In the context of innovation, a lot creativity is available, and a lot of different paths can be searched simultaneously. The Web offers an environment in which preliminary results can be rapidly shared. The "publish early, publish often" philosophy of Open Source encourages this.
As a means for the whole development community to explore and discover new application domains, the spontaneous emergence of *shared* quick-and-dirty solutions is probably second to none. Alongside the sharing of the techniques, down to the level of code, there is also a positive feedback loop for systems and protocols that encourage sharing, reuse and interoperability. Individual companies may build software Cathedrals, but for them to succeed commercially they must still be viable alongside the Bazaar.
A key aspect of the global environment is natural selection - techniques that are useful, ideas that *work*, these are the ones with the best chance of survival. To work in a diverse, globally distributed system, open protocols

have an advantage, and flexible application infrastructures are likely to gain greater adoption than single-purpose stovepipe architectures.

The Missing Languages

So, we have a list of failings of the current Web, and a few avenues that might offer hope for overcoming them. The keystone is a language or languages that can assist in the creation of the 'missing Webs', yet doesn't run counter to the current quasi-biological, evolutionary Web.

The Triumph of Syntax

Web technologies to date have been dominated by syntax, which isn't altogether surprising as the initial technologies; HTTP and HTML are primarily defined in terms of syntax. In recent years there has been an explosion of XML based document and data languages, and a case has been made that the "bits-over-the-wire" are the key to the success of any Web-like systems. At this point in time, to many developers the syntax, specifically XML syntax, is a core technology. Looking over at the Semantic Web stack of technologies, right there in the lower layers is XML.

RDF, Semantic Velcro

The first of the "missing" Webs listed above was the Navigable Web. There's a need to make the existing information on the Web more easily available, to make it easier to find things. Google uses a bunch of smart algorithms to generate its indexes. But this all happens post-publication, and involves heuristics so will always fall short of the optimal. The most efficient approach to finding things is not to lose them in the first place. That depends on having machine-readable indexes.

But the current document Web is woven as velour - it's soft and human-friendly. Material needed for indexing, designed for machine consumption only forms a small part of the fabric. However, when most documents or pieces of data are created or acquired, there is a significant amount of contextual metadata available as well. If an article is typed on a home computer, that computer should also be aware of information like the name of the author, the date and so on. When a digital photograph is taken the camera should have information about its settings. Looking ahead, a GPS-enabled camera will know where the photograph was taken, that information should still be available when the photo is published on the Web. All these different kinds of information can be easily expressed in RDF. Such statements can allow machines to get their hooks into the softer data, to index it or otherwise process it without human intervention. In other words, RDF can be used for *Crocheting the Semantic Web*.

RDF alone doesn't solve many problems, but combined with existing technologies, and extended through languages like OWL it is at the heart of a very powerful framework. This framework isn't a magic bullet, there are still many difficult problems in scope. But at least it makes possible a consistent means of approaching those problems.

Knowledge Management gets Personal

Earlier it was suggested that grassroots development was a good mechanism for exploring new application domains. What the grassroots generally delivers are single-purpose applications with implementations, ad hoc specifications of make-do formats and protocols, and a bunch of data using these. What's been notable in recent years is that at some level these often share the foundations of the Semantic Web technologies: XML (possibly with namespaces) and often URIs and HTTP-based transports. The material will often have been developed completely without the Semantic Web in mind, but that doesn't matter. It's usually possible to define the application and data model in RDF, and apply syntactic transformation to the data to produce RDF/XML. The data from many modern applications is a single RDF schema and a single XSLT stylesheet away from being first-class Semantic Web data.

A case in point is syndication and its languages: RSS ("RDF Site Summary" and variations) and more recently Atom. The first version of RSS (0.9) was defined in RDF/XML, though its purpose leaned towards filling in little boxes in an online application from Netscape (my.netscape.com). An XML-only version followed (0.91, 2.0), along with an improved RDF version (1.0). There has been an explosion in the deployment of these various formats, due in no small part to the increased popularity of blogs, personal sites in the form of chronological logs. Many news-oriented sites also publish data in these formats. Specialist tools for generating and reading this kind of data have emerged. The "aggregator" or newsreader is a form of viewer that enables the user to comfortably keep track of dozens of sites on a daily basis. This avoids the inconvenience of manual

navigation necessary with traditional Web browsers. Atom is a new in-development XML-based format and HTTP-based protocol set which aims to use what has been learnt from its predecessors and offer a more modern, unified system.

The material generated by many syndication systems may be "simple" XML, but the domain model corresponds directly to that of the RDF model of the data provided by RSS 1.0 (an OWL expression of Atom is in development). One syntactic transformation later, you have data for the Semantic Web. But perhaps more importantly, the RSS system has shown the benefit of what's been called microcontent - each syndication "feed" publishes a series of items, each of which is comprised of a small chunk of metadata describing a resource, typically a piece of in-line content. This kind of system has demonstrated its utility without Semantic Web technologies, but the system can be assimilated into the Semantic Web with no force whatsoever. The material published in feeds can be read just as easily by inference-capable, general-purpose RDF/OWL tools as the simple newsreaders currently on the market. The information can be utilised in concert with data from any other sources that have an RDF expression.

Syndication is only one application area, but it demonstrates how there's a direct, clear path between current Web technologies and the Semantic Web. Any new system based on the existing Web is likely to be suitable for deployment on the Semantic Web. The RSS aggregator allows the end user to manage a large flow of information. Most such tools currently available are single-purpose applications - they can let you read data from RSS feeds, possibly directly from (HTML) Web pages. But the general approach of metadata-tagged snippets of information offers a way forward for Personal Information Management. RDF vocabularies are already available for a vast number of application areas. Down in the grassroots of RSS a key enabler is the FOAF ("friend-of-a-friend") vocabulary. This allows a service or individual to create Personal Profile Documents which can serve as machine-readable glue for interpersonal networks. Calendaring/event-based applications are the target of the RDF Calendar group, which has developed an RDF version of the iCalendar standard.

Users and Ubiquity

A major benefit of the kinds of data mentioned here is that they are all amenable to appearing at different scales. The Web itself may be the big master database, but a local cache on a server can contain the equivalent of millions of database records. A mobile device may only contain, say, a handful of names and addresses. But no matter which scale is looked at, the graph/triples data model of RDF hold, the relationships between resources can remain consistent. Although technologies under development by the W3C and others for new user interfaces, accessibility and device independence on the Web have little direct relation to RDF, many are closely associated with the lower levels of the Semantic Web stack, i.e. XML and URIs. Such technologies already have the hooks for them to take advantage of the Semantic Web.

In line with the grassroots RSS developments, the advance of meaningful, global machine-machine communications enabled by Semantic Web technologies will allow selective aggregation of information of relevance to the individual. Local (and not so local) indexing and searching built on logical foundations promises a paradigm shift to Personal Knowledge Management.

What Else is Missing?

Returning to the list of missing Webs, the "Data Web" is directly covered by RDF and associated technologies. Data can be made available as XML over HTTP and/or it may be expressed and described using RDF. The "Trusted Web" is rather more difficult, and requires the addition of infrastructure to perform digital signing and verification of data. Again this is somewhere that W3C has efforts on the go without direct connection to Semantic Web technologies : XML Signature, Encryption. However, from the point of view of a logical framework, such efforts are inextricably linked to the rest of the Web project. The "Dynamic Web" requires that services be discoverable and describable - Web Services languages like WSDL and UDDI in part cover this. But questions like *which* services are needed for a particular task, this leads into requirements that go beyond what Web Service wiring diagrams can offer. For the operation of this and the "Data Web", rather more than the syntax layer of XML is needed.

Introducing Semantics

Certainly without format-level consistency there can be no communication between systems. But however necessary, syntax isn't sufficient. Without shared languages or models to carry the semantics, there can be no

meaningful communication. So what form should these semantics take? For starters, since we're talking World Wide Web, what should the world model be like?

Open and Shut
A big decision is whether the world model should be open or closed. If closed, we're in effect saying our system knows all it needs to know. If open, what's not specified is unknown. The real world is full of unknowns, but this does not sit altogether comfortably alongside the closed world of traditional relational DBs or the negation by failure of Prolog. A key aspect of the Semantic Web vision is the open world assumption. My own opinion is that any system built to operate on a global scale probably must be open in this way. To use a nearby analogy, historically a problem of large-scale hypertext systems was maintaining link integrity. In the mess of the Internet this would fall somewhere between unfeasible and impossible. Get rid of the need for something at the end of a link, the problem goes away. The open world assumption is a kind of 404 for the Semantic Web. This doesn't throw out the possibility of using closed-world reasoning on Semantic Web data, as a locally closed world can be defined if required, by harvesting or ringfencing the statements of interest. But conversely filling a relational database with nulls will hamper its capabilities, and query responses can be a more accurate "don't know" rather than a rigid "false".

Shortest Path
It has been suggested by some that the logics behind OWL and RDF are not powerful enough for the semantic Web. The semantic Web is in the (possible future) eye of the beholder, and there's plenty of room for argument. But the W3C's Semantic Web technologies overlap with the current Web and do offer some degree of logical sophistication. We could redesign SQL- or Prolog-based systems to work with the current Web, or design a whole new Web Version 2.0 from scratch but this would be no small effort. What's more, we don't need to.

Dan Brickley of the W3C describes the situation succinctly:

Traveller: I'd like to find my way to "Semantic Web", please.
Bystander: Well... I wouldn't start from here.

In other words (also Dan's, paraphrased), we get more than we lose by constraining our ambitions to fit with existing works in progress: namespaces, signature, accessibility, mobile, speech recognition and so on.

Once we have arrived at "Semantic Web" in the RDF and OWL sense, that is only the start of another journey. From there it should be much easier for other, possibly more sophisticated logics and formalisms to find their way from the research community in deployed industrial standards.

The Semantic Web is Legion
An opinion held by some is that the development of these technologies will exclude other alternatives. There is an element of truth in this, as specifications recommended by the W3C are likely to have an impact on the way resources are deployed in the computer industry, and are in themselves the result of resource deployment by members of the consortium. But although RDF and OWL in themselves have fairly tightly demarcated pieces of the first-order logic territory (in the sense that the logics may be formally defined, and data and inference appropriately constrained), the lower layers of the Semantic Web layer cake don't mandate this. Work on Topic Maps has points of correspondence to RDF, and XML Topic Maps are built on the exact same foundations of URIs and XML. Many systems already use SQL databases alongside RDF without any conflict. There is a subset of RDF/OWL that overlaps with the logic of relational databases, and another near to logic programming. Extension can happen further up the abstraction layers - a system based on a rule-oriented approach can operate alongside the Description Logics of OWL. The general approach taken by the Semantic Web initiative is very open, and highly inclusive. At a minimum applications can use RDF as a simple description-oriented data language, applying whatever additional semantics they choose. Having some form of Web-friendly logic-based language in place is a big step forward. If anything, the presence of *some* logic on the Web will help bootstrap other techniques.

Transparency and Unification
Transparency is enabled by description of things, whatever they may be. Once a thing is described then it can be reasoned about, it becomes a part of the knowledge base. On the Web scale, we're talking in terms of a large knowledge base, and they bring problems of their own. Merging of large knowledge bases is a difficult

problem, but merging is only necessary when the data sets are disjoint in the first place. The layered approach of Semantic Web technologies allows heterogeneous data to be treated at varying level of homogeneity. Its a little paradoxical - individual parts of the systems may be disjoint, but the system as a whole can be unified. The parts of the system can be processed by the same logical framework from the digital equivalent of microscopic and macroscopic scales.

Building Bridges

The Semantic Web initiative, building a set of logic-oriented technologies with the aim of creating a better Web, receives most criticism from two different directions. The first comes primarily from members of the academic community, researchers familiar with the past 30 years of developments in computable logic. The criticism from this direction is along the lines that the logic of RDF and OWL isn't enough, and that the whole project is misguided in using inefficient syntaxes (such as XML), in fact using syntax-oriented techniques as part of the solution. There is no doubt that there is awareness of technology-adoption issues amongst these critics, some of whom still have trouble with popular perceptions of their work and the AI permafrost. In the opinion of a large number of people working on the problem, what the Web lacks is some kind of base in formal logic. Data on its own isn't much use. However the existence of the current Web, and the rapid development of XML-friendly applications should be taken into consideration when looking forward. The Semantic Web languages are best seen as the first step in adding logic to the Web, not the finished article. In passing it's worth noting that a partial solution to the old, difficult problem of knowledge acquisition has emerged. Experts and non-experts alike are freely contributing to the global knowledgebase in Weblogs. The use of minimal metadata in the form of RSS makes this material available for machine processing. There are millions of RSS feeds available, and millions of FOAF profiles describing the individuals that created that data. As the Semantic Web builds on the same foundations as the existing Web, data and even whole systems designed for the current Web can be repurposed for the Semantic Web with little effort.

The other direction from which criticism appears is that of the in-the-trenches XML/Web development community. Many developers consider the apparent complexity of Semantic Web technologies a barrier, especially when there is no apparent immediate benefit. But gradually it is dawning on the community that metadata can be very useful, and that declarative, logic-based techniques are more appropriate for dealing with application business logic than hard-coded rules. Interoperability is a major problem in XML technologies. Introduce a common language framework with ontological management of domain-specific terms and many of those interop problems are alleviated. One of the weaker points in the XML-is-enough argument is that though XML is excellent for managing hierarchy-based information, the graph model of RDF is a much better fit for the real world. Additionally XML tends towards per-domain vocabularies, with little or no opportunity for combination of data sets. RDF in XML on the other hand inherits the benefits of XML's relatively simple syntax, but allows vocabularies that may have been developed completely independently to be used together.

So in conclusion, there is a lot missing from the current Web. Those gaps can be filled in part using a logic-based framework. However, for any such framework to gain widespread adoption it must leverage the progress that has already been made on the Web, and in practice that means compatibility with syntax-oriented HTML and XML systems. These, and related Web technologies can help fill in more of the gaps. The future Web will undoubtedly be a compromise between the desires of the theorists and the demands of the practitioners (not a difficult compromise for the many individuals already working on theory and practice). A set of technologies has been designed with exactly these contrasting requirements in mind, those of the W3C's Semantic Web initiative. But the project is ongoing, and the W3C specifications are really just a toolkit for developers. There are plenty of problems remaining, but there is now a suitable framework on which to build. Let's get building.

Special Issue Theme: Semantic Web Challenges for Knowledge Management

Table of Contents

Efficient Knowledge Management by Extending the Semantic Web with Local Completeness Reasoning

Abir Qasem and Jeff Heflin
Dept. of Computer Science & Engineering
Lehigh University
19 Memorial Drive West
Bethlehem, PA 18015
{qasem , heflin}@cse.lehigh.edu

Introduction

In this post-industrial "information economy" more and more organizations are realizing that having actionable information gives them an invaluable competitive advantage. The term "intellectual asset" has been coined to reflect the significance of this type of information. The field of knowledge management (KM) provides tools, techniques and processes for the most effective use of an organization's intellectual assets [Davies 00]. But the advent of the Web and its subsequent ubiquity has fueled a rapid growth in information volume that has not slowed down. In fact, in addition to the more traditional web pages, we now have information from diverse sources like databases, sensors, web services and even intelligent agents. This diversity of data sources combined with the trends in division of labor in modern companies lead to a knowledge space that is highly distributed and ever changing. The traditional KM tools assumed a centralized knowledge repository and therefore are not suitable for this distributed knowledge medium [van Elst et. al. 03]. New KM tools are needed that integrate the knowledge sources dispersed across web into a coherent corpus of interrelated information.

The Semantic Web offers a more suitable platform for information integration then the traditional Web. Since the data has well-defined meaning, software, instead of humans can be used to harvest information and subsequent knowledge from a wide variety of sources. [Davies et. al. 03] postulate that it will significantly improve the acquisition, storage and retrieval of organizational knowledge. They propose an architecture for KM in the Semantic Web that addresses all aspects of KM lifecycle, namely acquisition, representation, maintenance and use.

In the use phase, efficiency of knowledge retrieval is of paramount importance. This is difficult to attain in the Web's distributed knowledge space because information from several diverse sources that have different capabilities and communication protocols need to be pulled together in a timely fashion for it to be useful to the organization. Also, due to the large quantity of information sources, we cannot afford to query all of them. However, if we have compact meta-level descriptions of each source, then we can determine which set of sources need to be accessed, resulting in more efficient queries. In addition to providing information on relevant sources, source descriptions may also be used to indicate when accessing a source would be redundant. Such a redundancy might occur when two different sources have overlapping content. Reasoning about overlap is important for efficient KM.

In this work we use a formalism to characterize this overlap so that we can reason with it. Following work in the field of information integration [Friedman and Weld 97], our formalism is based on Local Closed World reasoning. We have augmented the W3C web ontology language (OWL) to allow us to express LCW statements. We have then described information sources using this augmented language and developed a prototype system that reasons with overlapping information and provides an integrated knowledge space to the user.

The rest of the paper is organized as follows: In Section 2 we provide pointers to related work, in Section 3 we describe our theoretical work and a simple prototype system that we have built, and in Section 4 we provide our conclusion and areas for future work.

Background

To be able to reason with overlap, we need to characterize and exploit source overlap. Addressing the source overlap problem is based on the idea of Local Closed World (LCW) information. LCW as proposed by [Golden et. al. 94] is a formalism for obtaining closed-world information on subsets of information that is known to be complete (LCW is described in more details in section 3). This formalism still allows other information to be treated as unknown. [Levy 96] extended this formalism to obtain complete answers from databases that have incomplete information.

All of this work however, assumes a priori knowledge of the local completeness information for each information source, which is an invalid assumption in the case of the Web. The Semantic Web on the other hand provides interesting possibilities for content providers to advertise the completeness of their sources. [Heflin and Munoz-Avila 02] have demonstrated this in a plan generation problem. They have shown how a planner can exploit LCW information encoded in SHOE [Heflin 98], another Web ontology language developed at University of Maryland, to complete a plan. Our paper builds on this work. We extend OWL to express source completeness and develop a system that reasons with that characterization and selects appropriate sources with respect to a query. Our system is based on the concept of "mediators" proposed by [Wiederhold 1992]; it is a system that is capable of integrating multiple sources in order to answer questions for another system.

Expressing Completeness on the Semantic Web

LCW information is used to specify the subsets of the information in a knowledge base that are known to be complete, while other information can still be treated as unknown. LCW information is given as meta-level sentences of the form LCW(ϕ), where ϕ is a first order logic sentence that contains one or more variables. If a sentence matches a substitution for ϕ then it is either already entailed by the knowledge base or it is false. In this sense, the sentence ϕ provides a scope for the relative completeness of the knowledge base. Note, that this information is local in the sense that it is local to the knowledge base that it describes.

However, these first order logic (FOL) formulas cannot be directly adapted to the Semantic Web. OWL, the de facto standard for the Semantic Web is closer to Description Logic (DL) rather than FOL. To represent LCW using OWL one has to express the formulas in DL. Unlike FOL, which allows us to refer to an object, DL only has notation to express definitions and properties of classes of objects. Classes provide an abstraction mechanism for grouping objects with similar characteristics. OWL classes are described through "class descriptions". Hence we have to express LCW for a class description, which will mean that we have LCW over all the instances of that class.

We use two meta-level statements to characterize a source's ability to provide complete information with respect to a query. They express the relevance information and LCW information about the contents of each information source. They are represented by formulas of the form LCW (ϕ) and REL(ϕ). For a knowledge source i, LCW_i (ϕ) indicates that for all x, if i does not entail that x is of type ϕ, then x is in the complement of ϕ. Formally, we can say $\forall x \ i \models type(x, \phi) \Rightarrow x \in \phi'$. REL_i (ϕ) on the other hand indicates that there exists an instance o in i such that i is of type ϕ. Formally, $\exists o \ instance(o) \wedge in (o,i) \wedge type(i, \phi)$.

We propose that an OWL document can use new properties lcw: isCompleteFor and lcw: isRelevantFor to state that it has complete <u>or</u> relevant information on some subset of information respectively. These properties are in a new namespace identified by the lcw prefix, and has rdf:Resource in its domain and owl:Class in its range. As such, it can be applied to any resource. The following examples show how to apply these properties to represent LCW. REL statements are expressed in a similar way.

We use the following to represent LCW (p (x, c)) on source s.

```
<rdf:Description rdf:about="s">
    <lcw:isCompleteFor>
       <owl:Restriction>
          <owl:onProperty rdf:resource="#p" />
          <owl:hasValue rdf:resource="#c" />
       </owl:Restriction>
    </lcw:isCompleteFor>
</rdf:Description>
```

Here isCompleteFor is applied to an individual of a class that has at least one of its property values equal to the resource c.

It is somewhat difficult to represent complete information on an object's values for a specific property. So to represent LCW (p (c, x)), we create an anonymous property that is the inverse of p, and restrict the value of the inverse (essentially restricting the value of the subject of p). To represent LCW (p (x,y)) we simply identify any individual with a property p.

We have built a prototype system, the Semantic Web Mediator, which identifies appropriate information sources with respect to a query. Its knowledge base contains two kinds of meta-information on the queries. It stores completeness information (i.e. which sources have all possible information about a query) and it stores relevance information (i.e. which sources have some information about a query). Using REL information, it can be determined which sources are relevant to a specific query. The LCW information makes it possible to prune redundant sources from this set. The knowledge base itself is initialized from OWL files that contain explicit description of sources' ability to provide completeness information.

Conclusions and outlook

In our work we have adapted LCW, a formalism commonly used to find relevant answers from an incomplete database, to characterize redundant information on the Semantic Web. We postulate that this representation will increase efficiency of KM on the Semantic Web. We have built a proof of concept system to explore the feasibility of this concept.

We are now in the process of building a more complete system. We plan for the following in recent future:

   a) Provide support for complex queries and data sources that commit to heterogeneous ontologies
   b) Allow for dynamic update of the system's knowledge base

References

[Davies et. al. 03] Davies, J.; Fensel, D.; Van Hermelen, F. 2003. Towards The Semantic Web: Ontology Driven Knowledge Management. John Wiley & Sons, NJ.

[Davies 00] Davies, J. 2000. *Supporting Virtual Communities of Practice*, In Roy, R. (ed.), Industrial Knowledge Management, Springer Verlag.

[Friedman and Weld 97] Friedman, M. and Weld, D. 1997. Efficiently Executing Information Gathering Plans. In *Proc. of IJCAI-97*.

[Golden et. al. 94] Golden, K.; Etzioni O.; and Weld, D. 1994. Omnipresence Without Omniscience: Efficient Sensor Managment for Planning. *In proc. of AAAI-9*4.

[Heflin 98] Heflin, J.; Hendler J.; and Luke S. Reading Between the Lines: Using SHOE to Discover Implicit Knowledge from the Web. 1998. In *AI and Information Integration. Papers from the1998 Workshop. WS-98-14.* AAAI Press, Menlo Park, CA, 1998. pp. 51-57.

[Heflin and Munoz-Avila 02] Heflin, J. and Munoz-Avila, H. 2003. LCW-Based Agent Planning for the Semantic Web. In *Ontologies and the Semantic Web. Papers from the 2002 AAAI Workshop WS-02-11*. AAAI Press, Menlo Park, CA, 2002. pp. 63-70.

[Levy 96] Levy, A. 1996. Obtaining Complete Answers from Incomplete Databases. In *Proceedings of the 22'nd VLDB Conference.*

[OWL 04] *OWL Web Ontology Language Guide*, retrieved March 15th, 2004 from http:// http://www.w3.org/TR/owl-guide/

[van Elst et. al. 03]. van Elst, L., Dignum V., Abecker, A 2003. Agent Mediated Knowledge Management, Springer Verlag.

[Wiederhold 92] Wiederhold, G. 1992. Mediators in the Architecture of Future Information Systems. *IEEE Compute*r.

The Road Ahead to Competency-Based Learning Activity Selection: A Semantic Web Perspective

Miguel-Angel Sicilia
Computer Science Department, University of Alcalá
Alcalá de Henares (Madrid) Spain
msicilia@uah.es

Introduction

From an organizational perspective, *e-Learning* can be considered an important component of the *Knowledge Management* (KM) function, as described by Wild, Griggs, and Downing (2002). In fact, even some architectural guidelines for this integrated view have been described elsewhere (Metaxiotis, Psarras and Papastefanatos, 2002), and the use of reusable *learning objects* in that context has also been analyzed recently (Lytras, Pouloudi and Poulymenakou, 2002). This perspective puts an emphasis on Web technology-based *learning activities* inside the organization as enablers of knowledge acquisition activities. In consequence, e-Learning becomes part of a more complex organizational *conduct*, in which lacks of required competencies trigger the search for appropriate contents or activities (i.e. learning objects), in an attempt to acquire knowledge and abilities that fulfil the contingent or strategic need. The diagram in Figure 1 depicts an abstract, simplified account for Learning Organizations that connect competency management with reuse of learning objects.
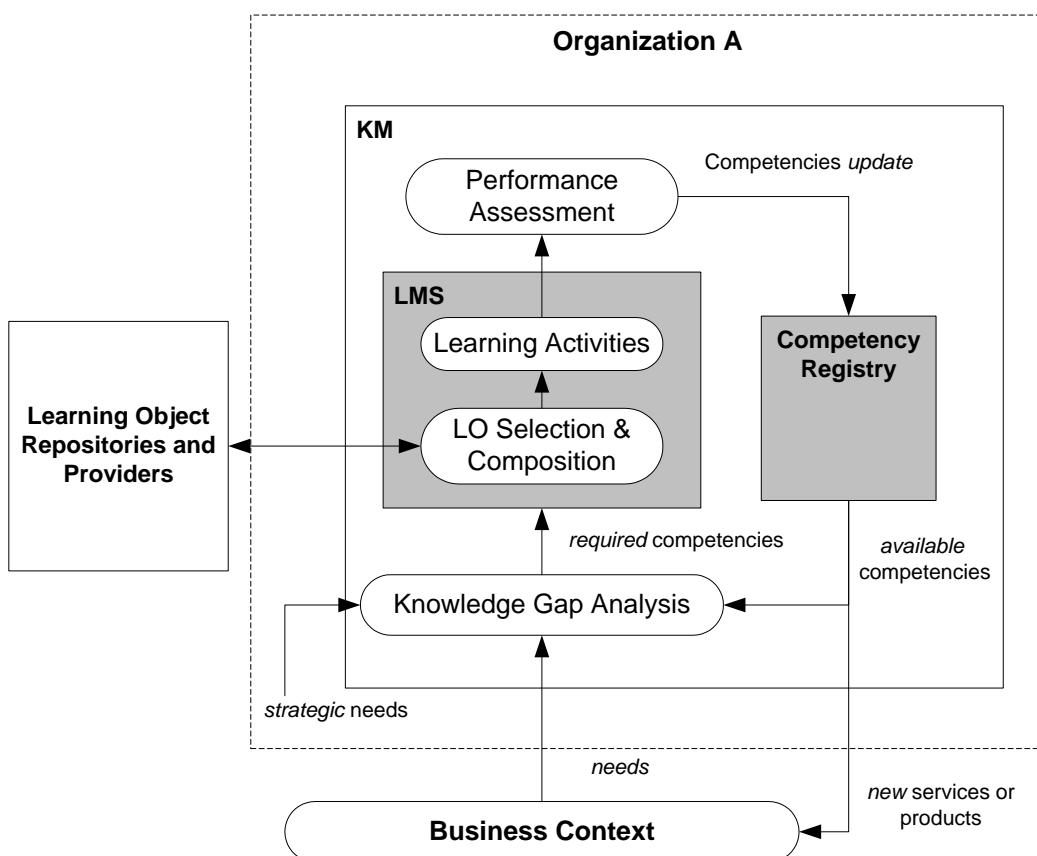

Figure 1. Overall view of e-Learning as a component in KM conduct.

As illustrated in Figure 1, the process of acquisition (usually) starts from a business need emanated from the context of the organization, or eventually from strategic management. Such needs trigger the process of

assessing if the organization is in place to deal with them. Such assessment is commonly referred to as 'Knowledge Gap Analisys' and essentially consists on matching the competencies required for the incoming needs with the available ones. If the result is not satisfactory, the process of searching for available resources should start. This process may entail the selection of learning objects in external or internal repositories and the composition and delivery of the appropriate learning activities. After these activities take place, some kind of assessment would eventually end up with an update of the registry of available competencies. Finally, the newly acquired competencies could change the position of the organization to offer services or products, closing this way the knowledge acquisition loop.

The critical point of the cycle depicted in Figure 1 is the linking of the knowledge goals of the organization with the knowledge adquisition processes enabled through e-Learning activities, which has been referred to as '*Learning Map*' (Wild, Griggs, and Downing, 2002). The partial or total automation of this process requires a rich and detailed knowledge representation for expressing needs and available capacities, and this is the point in which ontologies and Semantic Web technologies provide an appropriate infrastructure. But the provision of a flexible and commonly agreed infrastructure for linking competencies to learning objects requires a considerable amount of further work in several directions. Some of these required milestones are sketched in what follows.

Linking competencies to Learning Object Metadata

The view described so far requires in first place improved learning object metadata annotation that explicitly connects metadata records to ontologies (Sicilia and García, 2004). Current learning object standards and specifications allow in some way this kind of annotations. For example, the *Classifications* element in IEEE LOM can be used to specify concepts in a ontology. Doing that requires that the *Purpose* attribute be set to the value *competency*, and the rest of the attributes in the classification can be used to poitn to the ontology describing the competencies. Nonetheless, the provision of this metadata element is not mandatory in LOM, and there is not standardized way to provide a score or other kind of measure for the expected outcome. In consequence, a special *profile* for learning object metadata for adcquiring competencies should be required. Such profile could simply include idioms or specific practices for producing metadata that is actually usable and useful in KM processes dealing with competencies.

In addition, shared schemas for describing competences are required, so that the competencies provided in learning object metadata actually produce a consistent and explainable effect in KM systems. The HrXML *Competencies* schema (Allen, 2003) represents an important step in that direction. But its orientation as a flexible information exchange model has come at the cost of lacking a strict semantics to differentiate skills, knowledge items and competences, and also the diverse types of relationships between competencies - e.g. *aggregation* versus *'kind-of'* as described in (Sicilia, García & Alcalde, 2004)– are not properly addressed. Competency ontologies as those described in (Sure, Maedche, and Staab, 2000; Vasconcelos, Kimble, and Rocha, 2003) may serve to fill the gap of formal semantics that are lacking in proposed standards oriented to data interchange. With the provision of formal ontologies for defining competences, existing catalogues like O*Net could be enabled for expressing common competencies, taking into account the relationships between competencies where selecting target learners or learning contents.

Summary: Some directions for further work

The Semantic enablement of the competency adcquisition cycle depicted in Figure 1 requires further efforts in several directions. The following list summarizes some of the more important and urgent ones:

- Advancing in a shared ontological representation of competencies, by extending current interchange-oriented specifications like HrXML with improved semantics.
- Designing new approaches to learning object metadata that enable a precise specification of the competencies that are supposedly enabled by each learning object in each given organizational context.

- Integrating the KM and e-Learning views in a shared ontology for the purpose of developing Semantic applications that connect the managerial with the pedagogical perspectives.
- Developing advanced semantic tools for competency definition and for the assessment of knowledge gaps, taking into account relationships between competences, and also the imperfect nature of competency evaluation.

References

Allen, C. (ed.) (2003). *Competencies 1.1 (Measurable Characteristics)*, HrXML Recommendation, 2003 February 26

Lytras, M., Pouloudi, A. and Poulymenakou, A. (2002). Knowledge management convergence: expanding learning frontiers. *Journal of Knowledge Management* 6(1), 40-51.

Metaxiotis, K., Psarras, J. and Papastefanatos, S. (2002). Knowledge and information management in e-learning environments: the user agent architecture. *Information Management and Computer Security* 10/4, 165-170.

Sicilia, M.A., García, E. (2004). On the Convergence of Formal Ontologies and Standardized e-Learning. *Journal of Distance Education Technologies* 2(4) (to appear, Oct 2004).

Sicilia, M. A., García, E., Alcalde, R. (2004). Fuzzy Specializations and Aggregation
Operator Design in Competence-Based Human Resource Selection. *Proceedings of the WSC8 Conference*, Springer Verlag (to appear).

Sure, Y., Maedche, A. and Staab, S. (2000). Leveraging Corporate Skill Knowledge - From ProPer to OntoProPer. In: D. Mahling & U. Reimer. Proceedings of the Third International Conference on Practical Aspects of Knowledge Management (PAKM 2000), 30 - 31.

Van Elst, L. and Abecker, A. (2002). Ontologies for information management: balancing formality, stability, and sharing scope. *Expert Systems with Applications*, 23(4), 357-366.

Vasconcelos, J., Kimble, C. and Rocha, A. (2003). Organisational Memory Information Systems: An Example of a Group Memory System for the Management of Group Competencies. *Journal of Universal Computer Science*, 9(12).

Wild, R.H., Griggs, K.A. and Downing, T. (2002). A framework for e-learning as a tool for knowledge management. Industrial Management & Data Systems 102/7, 371-380.

Reflection on the future of knowledge portals

Martin Dzbor, Enrico Motta, Victoria Uren, Yuangui Lei & John Domingue Hamburg Knowledge Media Institute, The Open University, UK (Contact author: mdzbor@acm.org)

INTRODUCTION

Knowledge management has emerged as a standalone research discipline when academics and practitioners recognized knowledge as a core corporate asset. Knowledge management (KM) has been considered from many perspectives and generally dealt with several phases of the knowledge life cycle [1]: identification, acquisition, development, dissemination, use and preservation of organizational knowledge. Since late 1990-s KM has become almost a pre-condition for what has been labelled as a learning organization [12]. As we illustrated in our earlier work, knowledge gained by the employees during their engagement with a company is often lost in the dynamic business environment. Those who stay with the company are often unaware of critical resources that remain hidden in the vast repositories [8].

A major part of this organizational knowledge does not concern raw facts but rather facts in relation to each other, empirical correlations and ad-hoc dependencies drawing on practitioner's experiences. Knowledge as this is connected in one way or another to various documents published and archived within a company. The task for KM is then to provide a meaningful and straightforward access and interpretation of such documents. One approach to delivering this vision is through *knowledge portals* – particularly those developed within the organizational intranet setting.

A knowledge portal can be seen as a web application providing access to data in a semantically meaningful way, making available a variety of resources for diverse target audiences [5]. Current knowledge portals are dynamic in respect to the content of the repositories and presentation means but fairly prescriptive in terms of processes and individual user interaction. In this paper we contend the traditional view of a knowledge portal as a centralized repository for accessing large-scale corporate resources. Instead we argue there is a need to consider approaches that are more open-ended. Open-endedness does not necessarily mean to surrender the control over the content of knowledge repositories. We use this concept in terms of being end user centred and lightweight.

The core of our argument focuses on data interpretation, and argues that ontology for interpreting data is too restrictive if used for structuring the KM portal. Our approach aims to enable interpretation of data and access to documents from essentially any relevant resource existing in the Intra- or Internet. Rather than *pulling the user* to the knowledge portal, we suggest giving KM a task of *enabling the user to pull* appropriate and semantically relevant knowledge from wherever they currently are, yet using the rich and potentially restricted organizational resources.

CHALLENGING THE CONCEPT OF PORTALS

Several issues related to knowledge portals were identified by [5] as: a great effort to maintain such portals, considerable development curve to integrate the contained information, and difficulties with content presentation for knowledge-intensive dynamic web sites. We want to add another significant shortcoming: each portal reflects the perspective of its designer or in the better case, a group of corporate champions. This traditional approach to KM portals promotes a central repository, whence employees may retrieve important knowledge to make decisions [3].

While the content for KM portals is fairly well-defined, for expressing the contextual links among the chunks of content, unambiguous and shared referential points are needed. And this is the major source of trouble with KM portals. While the well-defined content may be easily captured and encoded, the context remains 'hidden' inside employees' minds. This obstacle is usually tackled by building a portal using an agreed-upon ontology – an explicit, formal, conceptual framework constructed for a particular domain or problem matter [10]. Corcho et al. present a typical scenario with "ontology developers being charged to develop the ontologies […] for describing and indexing the content for browsing" [5].

However, the selection of a particular referential framework for interpreting facts is *consensual* rather than prescriptive. Different stakeholders may want to choose different ontological perspectives to interact with and interpret the same document. The principle of a dynamically chosen perspective is not new. Cognitive scientists have been arguing the existence of *frames* that help people see and understand a problem in a particular light emphasizing some features whilst neglecting others [11]. The same task poses a variety of

problems if investigated from different perspectives. This is also valid for KM portals: the knowledge content may be used and interpreted differently depending on the users' specific perspectives.

Consequently, one of the emerging challenges for both KM and Semantic Web research communities is to enable the end users interact with the content of hand-crafted as well as ad-hoc knowledge repositories in a way that reflects their particular need and task. Admittedly, in an organization, the range of perspectives for enquiry would be limited, but *there is always more than a single perspective* – be it among different organizational units or management levels. This challenge is more visible if we broaden the remit of KM to include not only the hand-crafted, intra-organizational knowledge but also dispersed, almost anarchic sources, such as those existing in the current Web. In the remainder we show an approach towards "personal mini-portals" dynamically constructed and driven by the end users rather than ontology engineers.

FOUR AXIOMS FOR FUTURE RESEARCH

In order to respond to the challenges of multiple perspective, multi-source, and multi-modal interaction with knowledge, we propose the following axioms that define new – the "FAN" (*Frame-Annotate-Navigate*) knowledge portal paradigm. Knowledge content, and consequently 'gateways' to access it, should be:

- Decentralized with respect to *provenance* (i.e. users need to be able to add and re-purpose <u>any content</u>, not simply to access it);
- Decentralized with respect to *entry* (i.e. portals and knowledge inside them should be open to the <u>wider Web</u>, rather than simply cover restricted and relatively rigid internal resources);
- Decentralized with respect to *narrative* (i.e. rather than having a particular ontology structuring a portal, multiple ontological frames should provide <u>multiple perspectives</u> on the same content);
- Decentralized with respect to *functionalities* (i.e. it should be possible not only to add content and semantic annotations, but <u>also services</u>)

Current portals focus largely on the first axiom. Challenge for the Semantic Web community is to abandon the "silver-bullet ontology" approach, and use a single ontological perspective to structuring data directly (incl. the interfaces) rather than merely providing a view over it. Next, we show a *partial implementation* of the axioms presented above using our Magpie prototype. Other implementations of the framework are possible, and indeed a limitation of Magpie is that it is not yet able to use pre-existing annotations. Other limitations include: i) one form of semantic browsing, and ii) although Magpie supports flexibility in "semantic navigation", this is still rather restrictive, batch-like mode. Ideally, one would like to have more seamless mechanism to migrate from one ontological perspective to another during navigation.

FROM MAGPIE TOWARDS 'PERSONAL PORTALS'

Magpie is our framework *partially* responding to the challenge of users interpreting a given content from different conceptual perspectives [9]. The end-user aspect of the Magpie framework comprises plug-in extending standard browsing tools (e.g. Microsoft Internet Explorer). The plug-in enables the user to choose a particular ontology (which codifies a particular perspective on the domain in question) to access knowledge through conceptual "hot spots" relevant to this perspective. The "hot spots" are created automatically as a semantic layer on top of the actual content of the document. The plug-in allows the user to toggle subsets in the selected ontology by dynamically maintaining a "perspective toolbar" representing the categories of semantically relevant knowledge as simple push buttons. An interaction scenario for using Magpie in organizational setting has been described in [9], and Fig. 1 shows a typical end user screen from that scenario.

Unlike traditional KM portal, the information annotated and linked to in the Magpie-enriched document is soft; it changes when the ontological categories (see marker ❶ in Fig. 1) are changed. Moreover, not only does Magpie annotate different concepts for different perspectives but also associates different semantic services with the appropriate categories. Unlike traditional web links, semantic services could be standalone applications providing sophisticated reasoning facilities designed by the individual end users rather than a knowledge engineer in charge. Since users are allowed to create new services (perhaps reflecting their specific objectives and needs), the entire interaction is more dynamic than any KM portal can achieve.
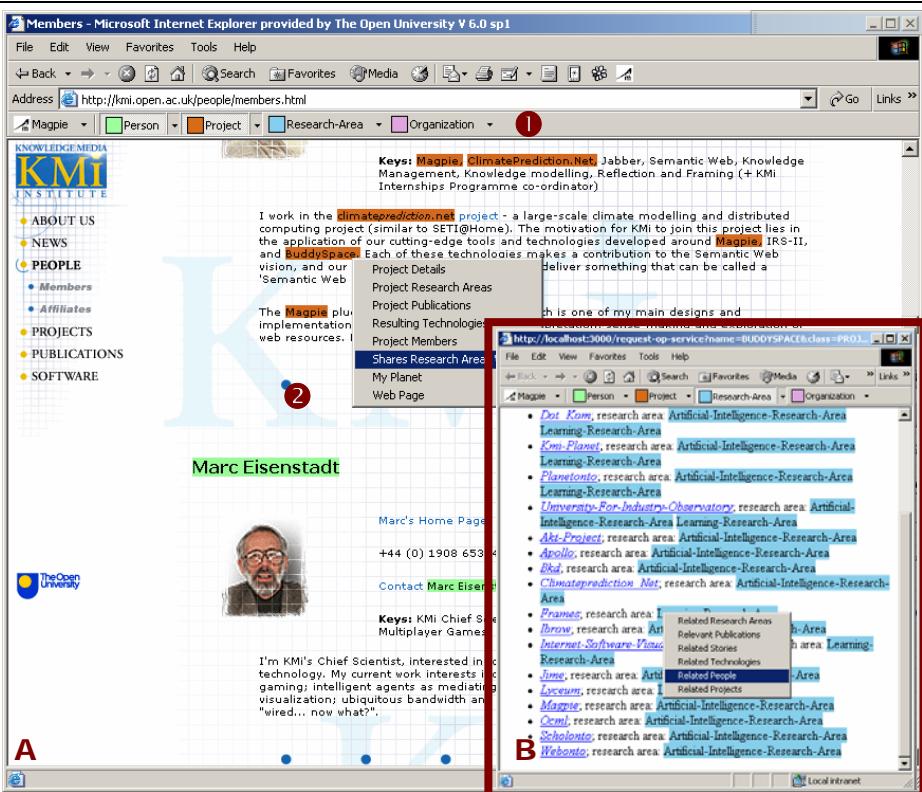
Fig. 1. *Magpie assisting with web page interpretation: A. typical semantically enriched document annotating people and projects on the organizational intranet with a semantic menu for project "BuddySpace" (see ❷), with an annotated response (B.) to the service "Shares Research Area With".*

The dynamics is supported by two additional facets of the Magpie infrastructure. First, any web application and its behavior can annotated in terms of a specific perspective[1] thus turning it into a *semantic web service* [6] – a service that can be comprehended and invoked in a unified manner by a variety of agents [4]. Magpie aggregates semantic services available for a given ontology and given category of the highlighted concept into a streamlined right-click contextual menu that can be displayed when the user interacts with a particular "hot spot". The primary advantage of this soft approach is that the users don't have to use a complex query language nor need they browse through a sequence of documents. Instead, they access appropriate knowledge using the lightweight Magpie toolbar as a gateway enabling rapid exploration of resources.

In addition to the dynamic nature of selected perspectives and relevant services, the users can take advantage of the second foundation the Magpie framework offers. Unlike portal technologies, Magpie distinguishes between on-demand and trigger services; the latter being particularly suitable to facilitate pattern-based knowledge acquisition and allowing the service providers to share initiative with the user. Trigger services also enable asynchronous information exchange unlike traditional portals that are largely based on the synchronous and stateless HTTP protocol. An example of a trigger service collecting on behalf of its user information about technologies related to the user's interests (and not necessarily mentioned explicitly in the visited page) is shown in Fig. 2. More details on the nature of trigger services can be found in [7].

---
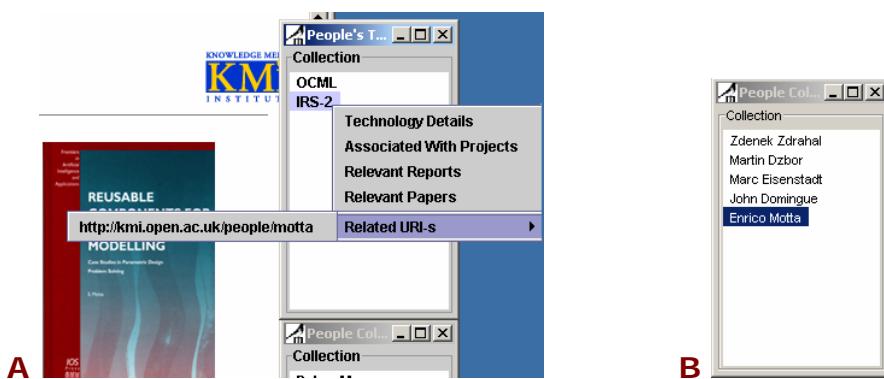
[1] In our example we use automatically populated knowledge base for an academic organization using the AKT reference ontology as a frame (http://www.aktors.org/publications/ontology).

Fig. 2. Two examples of Magpie trigger services: A. service collecting technologies relevant to the interests the user subscribed to (with appropriate semantic services and bookmarks), and B. simple collector of people encountered in the browsing session.

An important side-effect of this user-centered approach to interacting with knowledge through a selectable ontological perspective is that the framework can provide additional functionality basically "for free". For instance, the bookmarking functionality shown in Fig. 2 is done automatically whenever the (trigger service) provider discovers the relevance of a particular entity to the visited document. The relevance is formally captured as annotation and it is automatically associated with the respective URI. This in turn allows the user to tap into this repository too – for instance, by asking the system queries such as "find all pages I visited related to web research and describing some visualization techniques". This is clearly a significant step forward as compared to getting vast amount of replies from a typical search engine or asking the browser to re-visit a page browsed "in the last three weeks".

DISCUSSION

In this paper we propose a shift in the focus of KM portals research from "hunting the white elephants" towards delivering what the end user can apply in many decision-making scenarios. The quest for ever more complex, centralized, single-perspective KM portals is endless, only exacerbating the knowledge maintenance issues. Rather than pouring vast resources into a *centralized access* to knowledge repository, we suggest a more *distributed approach* to accessing what could be centralized knowledge repository, which is fusing the advances in KM and Semantic Web research. Such personal portal prototypes as we described above are less rigid and more open-ended, extensible and customizable by the moderately knowledgeable end users, thus requiring limited support from highly specialized knowledge engineers.

Furthermore, these 'personal portals' are not restricted to bringing the user to a specific web site ("the KM portal") in order to access and re-use knowledge. On the contrary, they bring the knowledge to the user regardless of the document presently visited. Knowledge access is based on its relevance rather than a user's visit of a particular location in the traditional KM portal. As a consequence, one can imagine a scenario with an employee visiting a site of a competitor, discovering an interesting concept, yet still being able to query organizational repositories to associate this knowledge with what is already known in the company – without leaving the original site. Hence, the approach we argue enables interaction with knowledge that is more *situational*, *timely* and uses a wider *variety of interaction modalities* than traditional portals[2]. More comprehensive reviews of related developments in the Semantic Web and HCI research areas can be found in [7, 9].

The regard for situational nature of knowledge makes such 'personal portals' more versatile because the same principles can be used in a wide range of *semantic web applications*. For instance, Magpie has been piloted as a distance-education application for The Open University climatology course, which was extrapolated into a climate science 'portal' for the climateprediction.net project [2]. The scenario presented above originates in another pilot – Magpie as an organizational portal for an academic unit. Another application comprises a sport fan's perspective on UK Premiership football and other sports. Obviously, the details of services differ case by case, but using the same interactive paradigm reduces the need for the user to (re-)learn new applications.

BIBLIOGRAPHY

[1] Abecker, A., Bernardi, A., Hinkelmann, K.*, et al.*, *Toward a Technology for Organizational Memories.* IEEE Intelligent Systems & their applications, 1998. **13**(May/June): p.40-48.

[2] Allen, M., *Do-it-yourself climate prediction.* Nature, 1999. **401**(6754): p.642.

[3] Bank, D., *Know It Alls.* The Wall Street Journal, 1996(Nov.18): p.R28.

[4] Berners-Lee, T., Hendler, J., and Lassila, O., *The Semantic Web.* Scientific American, 2001. **279**(5): p.34-43.

[5] Corcho, O., Gomez-Perez, A., Lopez-Cima, A.*, et al. ODESeW. Automatic Generation of Knowledge Portals for Intranets and Extranets*. In *Proc. of the 2nd Intl. Semantic Web Conf.* 2003. Florida, USA.

[6] DAML-Coalition, *DAML-S: Semantic Markup for Web Services*. 2002, http://www.daml.org/services/daml-s/0.7/.

[7] Domingue, J., Dzbor, M., and Motta, E. *Collaborative Semantic Web Browsing with Magpie*. In *Proc. of 1st European Semantic Web Symposium*. 2004. Greece.

[8] Dzbor, M., Paralic, J., and Paralic, M. *Knowledge Management in a Distributed Organisation*. In *Proc of the 4th IEEE/IFIP Conference on IT for Balanced System (BASYS'2000)*. 2000. Berlin, Germany.

---

[2] Indeed, these three aspects partially reflect the often-quoted KM credo of providing knowledge "in the right form, at the right time, and in the appropriate way".

[9] Dzbor, M., Domingue, J., and Motta, E. *Magpie: Towards a Semantic Web Browser.* In *Proc. of the 2nd Intl. Semantic Web Conf.* 2003. Florida, USA.

[10] Gruber, T.R., *A Translation approach to Portable Ontology Specifications.* Knowledge Acquisition, 1993. **5**(2): p.199-221.

[11] Schön, D.A., *Reflective Practitioner - How professionals think in action.* 1983, USA: Basic Books, Inc.

[12] Senge, P.M., *The Fifth Discipline. The art and practice of the learning organization.* 1990, London, UK: Random House. 424 pages.

Leveraging Knowledge Interoperation Towards The Semantic Era

Weihong Huang
Centre for Internet Computing, Department of Computer Science,
The University of Hull, Scarborough Campus,
Scarborough, YO11 3AZ, United Kingdom
W.Huang@hull.ac.uk

Interoperation Problem Emerges Along with the Semantic Web Develops

Following the incredible success of the WWW in 90's, the inventor Tim Berners-Lee and his colleagues innovatively created the vision of the Semantic Web (SW) [1] as a "web of trust" for the first time in 1998. Six years later, the SW is about to take off based on a set of specifications such as RDF (Resource Description Framework) and OWL (Web Ontology Language)[3]. When the SW wave swashes the knowledge management shore and brings up new beautiful spindrifts, it also brings up some new challenges.

The SW to agents is just as the WWW to our human users. If we expect the SW to be popular to agents as the WWW is to the human users today, we need to consider the sociological issues in deployment. One of the facts behind the success of the WWW is the equivalence between information creators and consumers: human to human using simple HTML encoding. But the situation will change on the SW, the knowledge creator will still be human (at least for a long time), but the consumers will turn into agents. Apparently, writing for agent understanding is a very complicated issue. Actually, this has been reflected by the machine-oriented RDF and OWL specifications, which seem to be not so easy to learn by users. Furthermore, the SW principles require users to build up their SW sites or applications according to the seven-layer cake SW stack [1], which is somewhat unrealistic. As long as the barrier in knowledge authoring is not lowered down enough, and the SW stack has not shown its overwhelming power over other models, the coexistence of the WWW, the SW and other knowledge-based networks is inevitable in the future. Under this situation, it is very important to address the interoperation issue between the exiting unstructured, semi-structured and structured information on the Web, and the fully structured standard-compatible semantic resources on the SW.

The interoperation problem will emerge in knowledge management when the SW is really in place and many existing application still have to use their preferred specifications in businesses. Despite the financial and management issues, to solve the interoperation problem in real business practices surrounding with non-RDF/OWL specifications such as XML Schema[4], XTM (XML TopicMaps)[5], BPML (Business Process Markup Language)[6], the most potential technical solution is to improve the efficiency and effectiveness of intelligent agents in processing heterogeneous knowledge.

Leverage Knowledge With Context-Awareness

To enable agents understand heterogeneous knowledge sounds like a magnificent aim. Ideally, agents are expected to comprehend and process information in a humanlike manner, and consequently reduce our workload in knowledge processing. Recent developments in service-oriented software engineering [2] and Web Services[7] have shown a promising future of service-oriented computing. Among the service-oriented research, one important issue has not received much attention: context. The concept of *context* has been widely used in many computing areas such as pervasive computing [3] and contextual logical reasoning [4]. In this paper, we would like to place the context between traditional existing information service-oriented descriptions and future semantic-aware service-oriented descriptions. Specifically, we define the concept in knowledge management as follows:

Definition: Context of an entity (i.e. an object, an event, or a process) is a collection of semantic situational information that characterizes the *entity's internal features or operations and external relations under a specific situation.*

A simple example about context in practice is like this. Agent *A* and agent *B* would like to find a good deal together on product *P*. Through service discovery, they find out two interesting audition sites: *XBay* and

---

Ymazon. Both the two sites provide the auction service they need, but *XBay*'s service is based on XTM while *Ymazon*'s service is based on XML. To carry out the biddings on the two sites properly, agent A and B have to comprehend the product information in different formats, and monitor the results in real time, and make decision according the situation. Apparently, this is not an easy case for pure SW stack -compatible agents since *XBay* and *Ymazon* are not SW sites. But if some context information is available in a common encoding format such as XML, and following a common description framework including typical contextual information such as service-related, product-related, even seller-related descriptions, the whole process will be easier to carry out comparing with traditional full-text processing. In addition to enabling the interoperability in services for agents, the context model on the same object can also be transformed and ported to other applications. Back to the information provision point mentioned above, providing such structured summary information for special services is a good drive for knowledge creation. Authors will be glad to treat the context descriptions similar to the meta-tags in HTML as value added information as far as they can benefit from the extended accesses of the service. Furthermore, since context is content-focused, there is not restriction on which description language should be used for presentation. The knowledge creator can make the choice based on their familiarity to languages and service requirements.
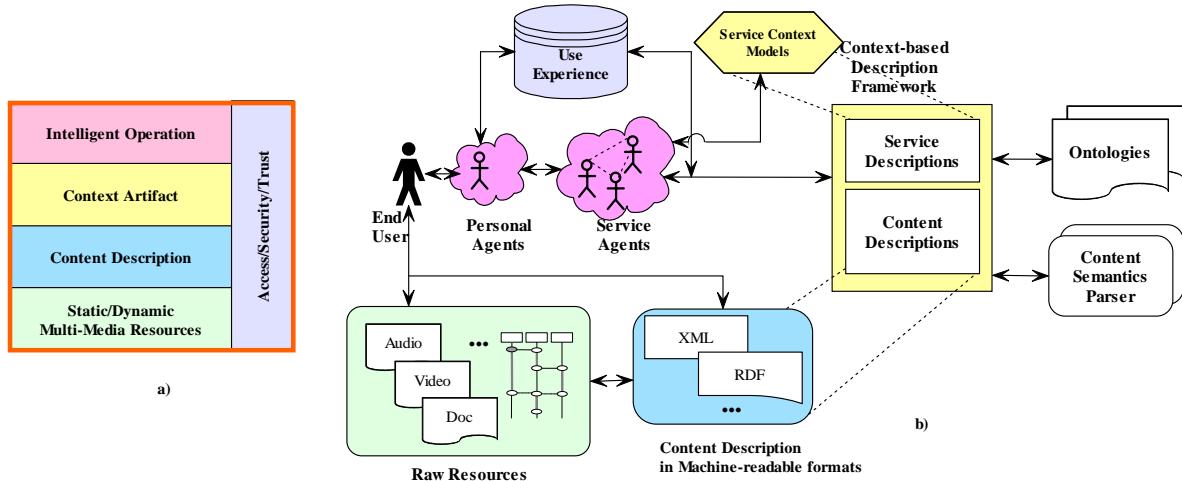

Figure 1. a) Knowledge leveraging model, b) An enabling architecture for context-aware services

Figure 1.a shows the proposed knowledge leveraging model featuring context awareness. The intention to set a content description layer is to enable the interface with existing generic information packaging formats and applications. The intelligent operation layer is set to enable high-level logic reasoning, multi-agent cooperation, and so on. The context artifact layer in the middle is to bridge the current gap between the two layers by enabling more structured and formal information for higher level operation and easy for low-level creation and integration. In practices, context descriptions are supposed to include generic schema information about the concept itself, and the basic context model in services (see Figure 1.b). For instance, in [5] we present a context model for RSS (RDF Site Summary/ Really Simple Syndication)[8] news aggregation. In this process, news providers can still use their preferable RSS formats[9] and do nothing more than usual. On the agent side, after we introduce the basic context service model to agents, they will be able to retrieve news items in various versions and formats. Furthermore, using the content parsers and ontologies, agents can understand the real meaning of the content and consequently filter the information for human users. When the agents go to a different service environment (e.g. BBC News Service) but within a similar modelling framework of service context, the integration and interoperation could be easily carried out at the context level, this is how context-awareness can promote the information and knowledge manipulation techniques in a practical way.

Guiding Knowledge Interconnection with a Reference Model

Adding context awareness to services can contribute to promote knowledge interoperation with intelligent agents, but this approach just touches one point in the global knowledge communication context. On the global network with various levels of heterogeneous descriptions and applications, we many need more generic solutions to address the interoperation issue. Luckily, we find out that the conditions of knowledge interconnection are somewhat similar to that in network interconnection in terms of architecture. As we all know the TCP/IP reference model guides the interconnection of the whole Internet with various protocols at different layers for different communication services. Similarly, we can expect a generic reference model to

---

[8] M. Pilgrim. What is RSS? XML.com, December 2002. http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html
[9] Note that there are three different versions of RSS in use, which are RSS 0.9x and 2.0 in XML and RSS 1.0 in RDF.

guide the knowledge interconnection of the global knowledge network with various description formats for different intelligent services. Here we introduce a KIRM (Knowledge Interoperation Reference Model) model [6] to act as a guideline framework to facilitate the interconnections between different knowledge domains including the SW, structured-data on the WWW, stand-alone knowledge domains, and so on. The KIRM model extends the idea of the knowledge leverage model with context-awareness in a global environment.

The KIRM model consists of five horizontal layers: *Data Link* layer, *Syntactic Description* layer, *Semantic Description* layer, *Context Description* layer, and *Intelligent Application* layer. There is also one vertical layer *Trust Management* layer across the top four layers (see Figure 2).
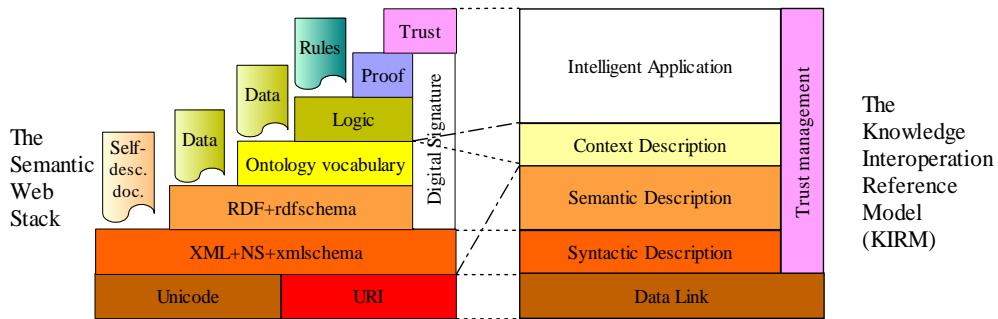

Figure 2. The SW stack versus the KIRM model

Basically, the KIRM model intends to use more generic layers to enable knowledge interconnection at different levels. It also complements the two design flaws of the SW stack. The first one is the functional independence between conjunction layers, which is one of the important requirements in generic layer modelling. The description specifications in the SW are dependent on each other. For example, following the general idea of layer modelling, people can interpret that ontology should only be built based on RDF and RDFS, which is actually not the case in reality. Therefore, we need a generic functional layered model with explicit functional interfaces to include guide the interoperation among instances like RDF and RDFS. This is similar to the TCP/IP model with five function layers, and at the TCP layer we have TCP and UDP protocols as instances for real operation.

In the KIRM model, the Data Link layer is equivalent to the lowest layer in the SW stack. The Syntactic Description layer is equivalent to the XML layer in the SW stack. The Semantic Description layer corresponds to descriptions in two layers in the SW stack: RDF and ontology. The Context Description layer operate based on the context-awareness idea above, and there is no equivalent layer in the SW stack. The Intelligent Application layer compromises the three top layers in the SW stack (i.e., Logic, Proof and Trust) for their tight relations.

The second problem with the SW stack is the position of trust-awareness and its management. Agents cannot always expect a final decision on trustworthiness to be carried out based on full implementation of underlying six layers in the SW. In contrast, the Trust Management layer in the KIRM runs across four functional layers from the bottom up. This mode reflects the nature of trust development in the real world. People can see the trust awareness building up along with enhancement of the knowledge description in a bottom-up manner. Since agents are expected to understand semantic resources at different description levels, the corresponding trust management to agents should also enable the awareness at all operational levels, not just relying on a simple mathematical digital signature in the SW stack. More detailed discussion about the KIRM model could be found in [6].

Conclusions and future work

In this paper, we discussed the problem of knowledge interoperation in the semantic era. To bridge the human-oriented information process mode and agent-based knowledge process mode towards the SW, we present an idea of introducing context awareness to leverage knowledge interoperation for agents. Towards addressing the global knowledge interconnection problem in the semantic era, we present a generic KIRM model referring to the idea of TCP/IP interconnecting the data networks. Ongoing and future work includes context-aware intelligent e-Learning development, KIRM model formalisation, and context-aware trust management for agents.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, May 2001.

[2] Z. Stojanovic and A. Dahanayake (Eds.), Service-Oriented Software System Engineering: Challenges and Practices, IDEA Group, 2004.

[3] T. P. Moran and P. Dourish. Introduction to this special issue on context-aware computing. Human-Computer Interaction, 16, 2001.

[4] V. Akman. Context in Artificial Intelligence: A Fleeting Overview. McGraw-Hill, Milano, 2002.

[5] W. Huang and D. Webster, Enabling Intelligent Agents to Understand Semantic Resources on the WWW and the Semantic Web, to appear in proc. of IEEE/WIC/ACM International Conference on Web Intelligence (WI04), Beijing, China, September 2004.

[6] W. Huang, and T. Tao, Adding Context-Awareness to Knowledge Management in Modern Enterprises, in proc. of IEEE International Conference Intelligent Systems 2004, Varna, Bulgaria, June 2004. (also selected for publication on International Journal of Intelligent Systems)

SWITCH ON!

Dr. Sandra Moffett and Mr. Martin Doherty,
School of Computing and Intelligent Systems
University of Ulster at Magee
Northland Road, Londonderry

Introduction

In today's knowledge-intensive organisations the primary objective of ICT is to lead users to the information they need. This includes creating, gathering, storing, accessing and making available the right information that will result in insight for the organisation's users (Davenport and Prusak, 1998). Thus, the pervasive use of information technology in organisations, qualifies it as a natural medium for information flow (Borghoff and Pareschi, 1999).

The main challenges facing organisational change and development are threefold. Firstly, knowledge discovery, secondly, corporate collaboration and thirdly, rapid decision making (Curley, 1998). From a KM viewpoint improved application of I.T. is a compromise between two polarities. An awareness of the limits of I.T. and realisation that any I.T. deployment will be relatively unsuccessful if not accompanied by a global cultural change towards knowledge value is the essence of KM. Quinn et al, (1996) envisage the development of ICT as, *"allowing many more highly diverse, geographically dispersed, intellectually specialised talents to be brought to bear upon a single project than ever before"*. The Semantic Web is one technology emerging to overcome barriers of limited interactive knowledge sharing and problem solving.

The Semantic Web

Inconsistencies in data/information categorisation have been of concern for many years. For example, in 1995, the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) convened the Metadata Workshop in Dublin, Ohio to address the issue of defining a simple data record to sufficiently describe a wide range of electronic objects. Research (such as Duval et al., 2002) now claims that many of the issues related to poor metadata can be contributed to the use of HTML (Hypertext Mark-Up Language).

HTML is the code mostly commonly used for the storage and transmission of current Internet documents. While HTML is an easy to learn/use simple language, well suited for hypertext, multimedia, and the display of small, uncomplicated documents, it has its limits. The rush to make data available across the Web has demonstrated weakness in its structure, mostly in relation to poor metadata assertions and static data presentation. To overcome such difficulties XML is emerging as a more flexible programming language for the Web, a point noted by Roche, (2000), *"XML (Extensible Mark-up Language) is a more flexible cousin of HTML, that makes it possible to conduct complex business over the Internet"*. Pfaffenberger, (1998) supports this view claiming that by adopting XML, users will be able to express much more finely grained document structures, *"This is the main point of XML, that people, by defining their own mark-up language can encode the information in their documents much more precisely than is possible with HTML. This means that programs processing these documents can understand them much better and therefore process the information in ways that are impossible with HTML"*.

For knowledge to be shared effectively there has to be a degree of computational manipulation of the cybercorp's data, most of which can now be displayed via web pages. The target of the Semantic Web, an initiative currently being developed by WC3, is to enable this manipulation to happen transparently through the Resource Description Framework (RDF). Therefore, while XML markup is the idiom of choice for the encoding and exchange of structured data, making it easier to achieve the principles of modularity and extensibility that HTML lacks, RDF takes this a stage further. RDF is an additional layer on top of XML that is intended to simplify the reuse of vocabulary terms across namespaces. While most RDF deployment to date has been experimental, there are significant applications emerging in the world of commerce (Duval et al., 2002).

According to Berners-Lee et al., (2001) one of the founders of WC3, the Semantic Web will *"bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users"*. By incorporating the meaning of terms or XML codes on a Web page, pointers can be made from the page to an ontology. Ontologies are documents or files that define relationships between objects, thus providing richer integration and interoperability of data. For web-based resources, an ontology typically incorporates a taxonomy and set of inference rules to boost the capabilities of software agents, identifying more significant search results and intelligent information retrieval. This extended functionality of machines, facilitating better comprehension of available data, will have a huge impact on KM development. The aim of this research is to raise awareness of this phenomenon in business organisations, to promote evolving standards and prototypes and to encourage early adoption of the Semantic Web as part of a technological focus for Knowledge Management.

Current Research

Research by Moffett et al., (2003) showed that for successful technological adoption and application within an organisation a number of factors must be present. Firstly, Knowledge Management systems should be well maintained, user-focused systems dedicated to communication and information flow within the organisation. A variety of technological tools should be used for knowledge work; these tools support function classifications as outlined in the literature. Secondly, dedicated roles must be established to promote technological use within the organisation. Employees at all levels should be encouraged to use KM systems for efficient and effective decision making. Reward and recognition must be awarded for their efforts. Thirdly, training must be provided to encourage full organisation of the tools installed. This training should be undertaken at internal, in-house and external levels. Fourthly, emphasis should be placed on web-based systems. This research has shown that use of the Internet is still a relatively new concept in organisations and one that is not yet being used to its full potential. While many organisations are content to use the World Wide Web [WWW] for information gathering, most are apprehensive to employing the Internet as an electronic commerce device. Even though they comprehend that a well designed, organisation-wide, fully implemented technical infrastructure for KM can improve information processing capabilities, knowledge discovery, project collaboration and rapid decision making, they are unsure how to adopt Web technologies to achieve this.

Forthcoming Research

Realisation that many UK organisations are unclear how to fully embrace the Internet for business operations led to ongoing research in this field. Firstly the authors intend to conduct initial scoping with a sample of UK organisations via an on-line survey. This exploratory research will establish the current state of practice of technological presence for KM with particular emphasis placed on Internet technologies, including the Semantic Web. Analysis of the results will highlight organisations suitable for case-study activity. In-depth research will then occur to answer the following research questions:

What are the current technologies in organisations which facilitate KM activity?
To what extent do UK organisations embrace web-based resources for knowledge sharing and application?
What ontologies, taxonomies and protocols are adopted for the structure and transfer of data organisational and/or industry wide?
Is there a need for standardisation to encourage Semantic Web application for UK organisations?

Overall, the aim of this research is to encourage 'Semantic Web Integration Through Converging Hybrid Ontologies'; in other words persuade UK organisations to **SWITCH ON.** By understanding current practice, and identifying gaps in industrial implementation, this research will design a set of guidelines for Semantic Web adoption. In addition, a corporate prototype will be designed to reflect best practice for Semantic Web application. The prototype will be developed based on XML and RDF, incorporating software agent functionality via taxonomies and inference rules, presented through a usable interface. This application will be tested with a number of participant organisations to evaluate intelligent information retrieval and Semantic Web utilisation for sustainable business improvement.

References

- Berners-Lee, T., Hendler, J. and Lassila, O. (2001), *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American.Com, http://www.sciam.com
- Borghoff, U. and Pareschi, R. (1999), *Information Technology for Knowledge Management*, http://www.iicm.edu/jucs_3_8
- Curley, K. (1998), *The Role of Technology*, Knowledge Management: A Real Business Guide, Caspian Publishing Ltd., London, pp. 48-52
- Davenport, T. and Prusak, L. (1998), *Working Knowledge - How Organisations Manage What They Know*, Harvard Business School Press, Boston
- Duval, E., Hodgins, W., Sutton, S., Weibel, S.L. (2002), Metadata Principles and Practicalities, D-Lib Magazine, Vol. 8, No. 4, ISSN 1082-9873, http://www.dlib.org/dlib/april02/weibel/04weibel.html
- Moffett, S. McAdam, R. Parkinson, S. and Patterson, G. (2003), The Utilisation of Technology for Knowledge Management Application, International Conference on Information Systems in Engineering and Construction, Florida, USA
- Pfaffenberger, B. (1998), *Web Publishing With XML in Six Easy Steps*, Academic Press, USA
- Quinn B, Anderson P and Finkelstein, S. (1996), *Managing Professional Intellect: Making the Most of the Best*, Harvard Business Review, March-April, pp. 71-80
- Roche, E. (2000), *Explaining XML*, Harvard Business review, Jul-Aug

Challenges for Knowledge Management in the Biomedical Domain

Christine Golbreich
Laboratoire d'Informatique Médicale, Université Rennes 1,
35033 Rennes, France
Christine.Golbreich@univ-rennes1.fr

Information and services sharing, and multimedia indexing are two crucial issues for knowledge management in the biomedical domain. But they require overcoming several challenges. On one hand, semantic integration is a key bottleneck in the deployment of a wide variety of biomedical information applications including functional genomics, cancerology, anatomy, as well as in many other medical domains, where better patient care, better understanding of diseases and sound decision making in public health need to access large amounts of data distributed in multiple heterogeneous resources. However, traditional data integration, such as ad-hoc systems, datawarehouses, are proved to be not sufficient. On the other hand, due to the numerous health documents now available on the Web, information retrieval with classical indexing tools based on biomedical thesaurus/standards e.g. the UMLS® or GeneOntology™ (GO) becomes problematic.

For scaling up to the Web, newly suggested techniques are more appropriate. Mediator or peer architectures provide more flexible "virtual" integration to access heterogeneous and distributed information. Formal ontologies are a key component of such architectures. They also play a central role in the Semantic Web, since they define which concepts should be used for semantic markup of information resources. Therefore, new formal Web languages, e.g. the standard Ontology Web language OWL or the Semantic Web Rule Language SWRL, that offer powerful inference services are really useful for supporting ontologies construction, validation, and maintenance.

Projects conducted at LIM, in particular the previous terminological server of the French National Agency for Transplantation (EfG) [11], and BioMeKe, a system devoted to gene annotation [**12**], reinforced my opinion that mediators and formal ontologies would benefit to biomedical integration [**4**]. BioMeKe uses GO and the UMLS® to annotate genes, but GO and the UMLS® were not built according to formal ontological principles and exhibit inconsistencies. Indeed, GO top categories Molecular Function, Biological Process, and Cellular Component are structured according different viewpoints, e.g. biochemical viewpoint derived from the EC Enzyme Commission classification, substances they act on. As a consequence, a term may be found both sibling and child of another a term e.g. 'Metal ion transporter activity' is a sibling of 'Cation transporter activity', while in another sub-tree it is a child of 'Cation homeostasis'. It entails that any gene annotation derived from it, may exhibit inconsistencies, redundancies, lacking, or heterogeneity. Same difficulties occurred in a different medical domain concerning end stage organ failure and transplantation, which led us to a new approach. Current work is devoted to the development of a Local as View mediator for querying heterogeneous sources of dialysis and transplantation information. The mediator query answering algorithm is based on two main knowledge components: a global formal ontology of the dialysis and transplantation domain and mappings between that ontology and the local sources. For building that ontology, it was difficult to reuse the previous EfG thesaurus. Indeed, this terminology server has been built to integrate several existing terminologies, e.g. the French Thesaurus of Nephrology and ICD, being driven by a nosological viewpoint. In that server, the diseases are described according a frame-like view, and organized according different dimensions, e.g. "diseases classified by location", "diseases classified by evolution", "diseases classified by finding" etc. But being built pragmatically and manually, and implemented in a language without multiple inheritance management, the different sub-trees exhibit inconsistencies redundancies, or misclassifications, mainly issued from the multiple hierarchies.

These examples illustrate a general issue. Biomedical ontologies play a key part for supporting semantic information integration and indexing. But they inevitably arouse the old problem of multiple viewpoints, which is recurrent in biomedicine and often originate many inconsistencies. Since biomedical ontologies are huge, continually evolving, and involve different viewpoints, they require automatic tools and techniques for their construction, management, and validation.

A main advantage of Description logics (DL) in general and of the OWL-DL language in particular, is that they provide useful reasoning services for consistency checking and classification. The present experience of

building a formal ontology of dialysis and transplantation with the Protégé OWL ontology editor is really convincing [7]. The methodology for building such an OWL ontology is really different from the construction of a frame-based knowledge-base. Classes are defined without boring about which hierarchies they belong to. Next, based on their logical definition or inclusion, the multiple hierarchies are automatically computed from their logical definition by a Description Logics reasoner like RACER [8]. It also enables a progressive improvement of the ontology in checking its consistency. Although it still remains a long and difficult task to build ontology, this approach appears to be really more satisfying.

OWL DL supports powerful automatic reasoning, but also arouses some difficulties. Using OWL for building biomedical ontologies e.g. the brain cortex anatomy [5], or the transplantation ontology, we have been faced to some DL expressiveness intrinsic limitations, which require deeper investigation and developments [6]. For example, it is difficult to express the propagation of properties, like the transfer of properties from parts to wholes, or properties dependencies, as exhibited for instance in the brain-cortex ontology [5 2]. The OWL ontology language is well suited to represent "structured" knowledge by classes, properties and taxonomies. But OWL is not sufficient for representing "deductive" knowledge. Although some knowledge may be represented by any paradigm (i) DLs and rules expressiveness are generally different, (ii) each paradigm better fits some particular type of knowledge and provides specific reasoning services. DL are really useful for the "terminological" part of a biomedical domain and to support some efficient means to reason, such as ontology consistency checking, classification reasoning, multiple hierarchies management, class recognition of instances, which are essential services for biomedical ontologies. But on the other hand, rules are needed to represent the "deductive" part of the biomedical knowledge. "Standard rules" are needed to express the transfer of medical properties along another property, such as the transfer of properties from parts to wholes. "Bridging-rules" are useful for reasoning across several domains e.g. genomics, clinics, anatomy, "mapping rules" for mapping Web ontologies in data integration, "querying-rules" for expressing complex queries upon the Web, "meta-rules" for facilitating ontology engineering (acquisition, validation, maintenance). My claim is that Description Logics and Rules are both required for biomedical ontologies for a future Semantic Web. Therefore, an ontology formal sublanguage extended by a Web rule language that allows interoperating between biomedical ontologies and biomedical rules is required. The recent proposal for a Semantic Web Rule Language, SWRL10 [10] based on a combination of OWL DL and OWL Lite [1] with a RuleML sublanguages is a first step. But a step further is needed for interoperating between SWRL and OWL, not only syntactically and semantically, but also inferentially [6]. We have developed a first prototype of a Protégé Tab widget, SWRLJessTab, to bridge between Protégé OWL11 , Racer [8], and Jess [3] for reasoning with SWRL rules combined with OWL ontologies.

Another important issue is how to take benefit of existing standard terminologies that are largely used for information retrieval, and how to migrate them to formal ontologies. We have recently begun to investigate how to migrate the French CISMEF terminology, a Catalogue and Index of French-speaking Medical Site which is based on MeSH [13] and the Fondational Model of Anatomy (FMA), and are studying how to abstract some general principles for it.

Main challenges to address for semantic integration and indexing in the biomedical domain include:
- Expressiveness of Description Logic with regards to biomedical ontology requirements
- Migrating huge ontologies such as GeneOntology™, UMLS, NCI, FMA to OWL
- Partially automatizing ontology mappings acquisition.
- Hybrid indexing and search, combining formal ontologies and medical standard ontologies
- Alignment of biomedical terminologies supported by a formal ontology
-

---

[10] http://www.daml.org/rules/proposal/
[11] http://protege.stanford.edu/plugins/owl/

References

1. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I. McGuinness, L. D., Patel-Schneider, P. F., Stein, L., A.: OWL Web Ontology Language Reference. W3C Working Draft (2003)

2. Dameron O., Gibaud B., Musen M. Using semantic dependencies for consistency management of an ontology of brain-cortex anatomy, KR-MED 2004.

3. Ernest J. Friedman-Hill: Jess 6.1 manual. Sandia National Laboratories , 2003.

4. Golbreich C, Burgun A, Jacquelinet C. Biomedical Information Integration: two case-studies SIG4, OntoWeb 5 Meeting, ISWC 2003, Florida http://sig4.ago.fr/doc/Technical%20Programme_Ontoweb_SIG4-Florida/LIM_Ontoweb_SIG4-Florida.pdf

5. Golbreich C, Dameron O, Gibaud B, Burgun A, Web ontology language requirements w.r.t expressiveness of taxononomy and axioms in medecine, 2nd International Semantic Web Conference, ISWC 2003, Sanibel Island, Florida, October 2003, Lecture Notes in Computer Science, volume 2870, Springer. 2003.

6. Golbreich C., Imai A., Combining SWRL rules and OWL ontologies with Protégé OWL Plugin, Jess, and Racer. 7th Protégé International Conference, Bethesda, July 2004 Golbreich C., Mercier S., Construction of the dialysis and transplantation ontology, advantages, limits and questions about Protégé OWL, Medical Workshop at 7th Protégé International Conference, Bethesda, July 2004 Haarslev V., Möller R. : Description of the RACER System and its Applications. Description Logics 2001

9. Holger K., The Protégé OWL Plugin, 7th International Protégé Conference, Bethesda, 2004.

10. Horrocks, I., Patel-Schneider, P., Harold, B., Tabet, S., Grosof, B., Dean, M.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Version 0.5 Nov. 2003

11. Jacquelinet C, Burgun A, Delamarre D, Strang N, Djabbour S, Boutin B, Le Beux P. Developing the ontological foundations of a terminological system for end-stage diseases, organ failure, dialysis and transplantation. Int J Med Inf. 2003 Jul;70 (2-3).

12. Marquet G, Golbreich C, Burgun A. From an ontology-based search engine towards a mediator for medical and biological information integration, Semantic Integration Workshop, Second International Semantic Web Conference, 2003, http://SunSITE.Informatik.RWTH-Aachen.de/Publications/CEUR-WS//Vol-82/

13. Soualmia L., Golbreich C. , Darmoni S., Representing the MeSH in OWL: Towards a Semi-Automatic Migration, First International Workshop on Formal Biomedical Knowledge Representation, KR 2004, Whistler, Canada

Emergent Semantics
An organizing principle for Biomedical Informatics and Knowledge Management

Vipul Kashyap

Clinical Informatics R&D, Partners Healthcare System, Inc.

vkashyap1@partners.org

Biomedical and biological research has been transformed from a cottage industry, marked by scarce, expensive data generated manually, to a large-scale data-rich industry, marked by factory scale sequencing. This is in addition to vast amounts of biomedical research literature being generated at an increasing rate and available through various web based sources (e.g., PubMed [3]). Success in the life sciences will hinge critically on the availability of computational and data management tools to retrieve, fuse, interpret, analyze, classify, compare and manage the abundance of data. Biomedicine is fast becoming an information-based science with data/information playing a big role across the "research flow", e.g., Genomics → Transcriptomics → Proteomics → Metabolomics → Final Products/Results. The final products may either be drugs and therapies or positive or negative research results in the field of biomedicine.

Scientific Data Integration has been identified as one of the most daunting challenges at the interface between computer science and biology [1], and has been seen as restraining rapid progress in biomedical research [2]. The standard paradigm in biology today is: Data → Hypotheses → Models → Experimentation → Data, and a solution framework for biomedical data integration should provide support for these artifacts. Biomedical data integration, poses a unique set of challenges [1,4]:

- **Diversity of Information Objects, including Data Types and Queries**:
  o Sequences, complex phenotypic and disease-relevant data, graphs, 3D structures, images
  o Similarity-based queries (e.g., sequence similarity), classification based queries (e.g., Papers about Gene X), what-if hypotheses generating queries (what if Gene X was suppressed, will Protein Z exist?)
- **Diversity of information based computations and operations:**
  o Experimental Plans and Protocols involving complex repetitive computations involving data retrieval, fusion and analytics.
  o Data Curation and Annotation Tasks.
  o Hypothesis validation across multiple scenarios, Federated search/query processing.
- **Semantic Heterogeneity:**
  o Multiple Controlled Vocabularies and Ontologies: Integration, Interoperation and Composition
  o Semi-automatic creation and verification of Mappings
    - Mappings and complex relationships across vocabularies and ontologies
    - Mappings, Annotations of data objects to concepts in controlled vocabularies and ontologies
- **Dynamic and evolving nature of Biomedical Research**
  o Evolution of schemas, ontologies and vocabularies and their impact on underlying mappings or annotations.
  o Evolution of data objects and their impact on associated mapping and annotations
  o Uncertainty and inconsistency of data
  o Support for pro-active data mining and hypothesis generation

Semantics-based approaches are being explored for information integration in the context of the Semantic Web [5]. However, "semantics" or "meaning" is not a fixed entity – it *emerges* from the interaction of people and applications in the context of performing biomedical research. An *emergent semantics-based* information infrastructure would be a pro-active platform where people and applications collaborate for creation of dynamic "semantics" reflecting the current state of knowledge in biomedical research. Some interesting properties of this infrastructure are:

- **Self-description**: The infrastructure shall enable self-description of biological data and content. This is currently
  the focus of various XML-based markup languages (e.g., BioPAX [6], OWL [17]) and ontologies/vocabularies (e.g., GeneOntology [7], UMLS® Semantic Network [8]) developed and to enable data sharing and interoperability.

- **Self-genesis**: The infrastructure shall proactively analyze data and content flowing through it, to create models,

  ontologies and concepts to capture semantics. This enables "bootstrapping" of the meanings prevalent in biological data and content.
- **Auto-emergence**: The infrastructure shall monitor interactions between people and applications to capture and describe new meanings/knowledge that emerge, or existing meanings/knowledge that evolve from these interactions. This capture and description of new meanings that may be required for new applications and services. New biomedical hypotheses and experiments are pro-actively suggested bases on the new emerging knowledge.
- **Self-organization**: The infrastructure shall (re-)organize itself to perform new tasks, in response to new requests for information and services; or due to emergence of new meanings (through self-genesis, auto emergence or otherwise). Two important types of self-organization are:
  - o **Self-interoperation/integration**: The infrastructure shall integrate existing meanings, or interoperate between two different meanings in response to new requests for information and services. This may take the form of annotating data with ontological concepts or mapping concepts across different ontologies.
  - o **Self-provisioning**: The infrastructure shall, monitor the data retrieval and computation operations invoked and pro-actively create experimental plans, protocols and models.

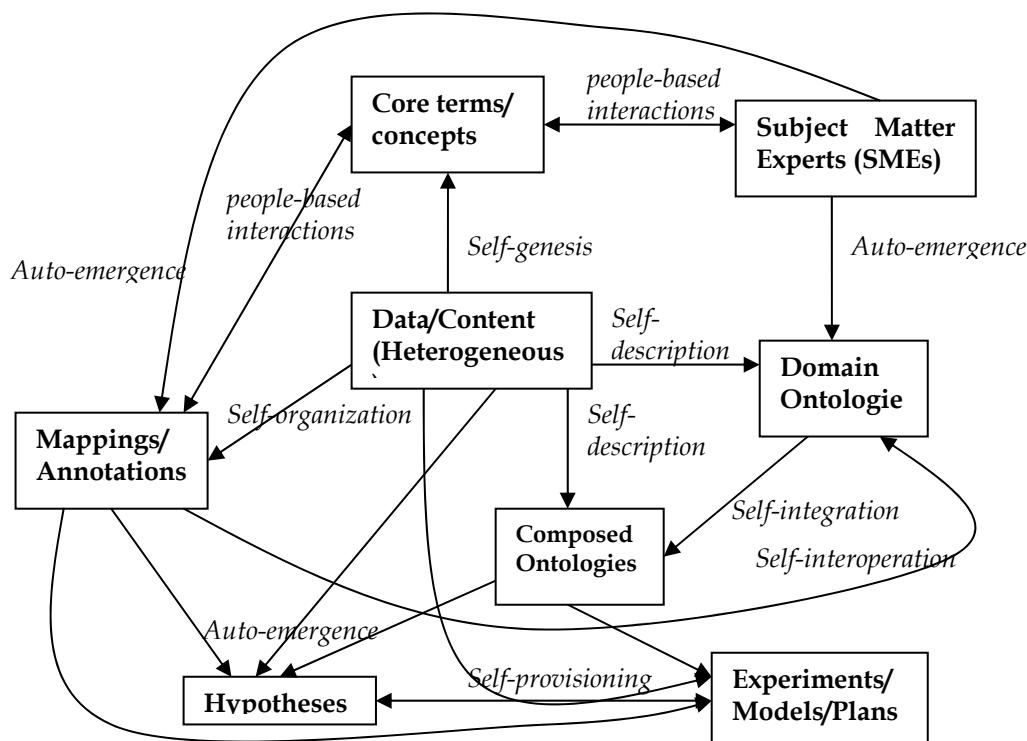*A strawman architecture for the above infrastructure is illustrated in* Figure 1.

Figure 1: A Strawman Architecture

For realizing this vision, the following obstacles need to be addressed.

Ontology Creation: Knowledge acquisition and representation research has struggled with manual and expensive approaches for developing ontologies. A bottom-up process of bootstrapping knowledge using clustering and NLP techniques for initializing a core terminology presented in [9], is a promising approach for enabling self-genesis. Sociological consensus-derivation processes that capture the notion of people-based interactions [10] can be used to enrich rough terminologies, establish agreements on ontologies, annotations and inter-ontology articulations show the potential for supporting auto-emergence. There are a wide variety of communities and user groups on the web, each with their own information, models, databases schemas and ontologies. Top-down processes of interoperation across pre-existing and possibly overlapping ontologies [13,14] are candidates for enabling self-integration. Techniques for enhancing existing knowledge artifacts (e.g., database schemas) to create new ontologies [11] can also enable auto-emergence.

Metadata Extraction, Annotation and Mapping:

The ability to describe web resources using metadata descriptions constructed from domain ontologies is crucial in for curation and annotation of biomedical data and content. Some approaches that can help address this obstacle are:

- Association of a vector space generated from a document collection with ontologies [12] for semi-automatic annotation.
- Use of E-R models to describe and query text information and extraction of metadata from multimedia data [15, 16].
- Machine learning techniques to infer mappings between database schemas, web service descriptions and ontological concepts.

Integration/Interoperation (III):

With a wide variety of communities and user groups on the web, there is a need for interoperation across heterogeneous overlapping ontologies (concepts from which might be used for creating information models and database schemas). **Self-integration/interoperation** is the key property that we seek to enable by addressing the III obstacle. Metadata annotations and mappings discussed above play a crucial role in this context. Approaches that use Description Logics to represent ontologies and their inference capabilities to support inter-ontology integration/interoperation [13,14] can be used to enable **self-integration/interoperation**. Specialized techniques for combining repetitive data retrieval and analysis operations/processes are required for enabling **self-provisioning**.

An emergent semantics-based platform enunciated above, enables a highly adaptable and flexible information infrastructure. The evolution of meaning and biomedical knowledge, supported by the underlying infrastructure, makes it easier to implement new applications for different purposes and requirements, and those that involve continuously evolving knowledge and information. A multi-disciplinary research agenda involving database and information systems research complemented with insights from Knowledge Representation, Machine Learning, Data Mining and NLP is crucial to realizing this vision. There is a need to go beyond traditional information systems and investigate approaches based on social networks and cultural anthropology to achieve a truly flexible e*mergent semantics-based* Knowledge Management platform.

References

1. Digital Biology: The Emerging Paradigm, November 6-7, 2003, NIH Natcher Conference Center, Bethesda, MD
2. NIH Roadmap: Accelerating Medical Discovery to improve Health, Bio-Informatics and Computational Biology, http://nihroadmap.nih.gov/bioinformatics/index.asp
3. PubMed, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed
4. Workshop on Data Management for Molecular and Cell Biology, February 2-3, 2003, Lister Hill Center, NLM, NIH, Bethesda, MD, http://pueblo.lbl.gov/~olken/wdmbio/
5. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities," Scientific American, May 2001.
6. BioPAX: Biological Pathways Exchange, http://www.biopax.org
7. Gene Ontology, http://www.geneontology.org
8. V. Kashyap and A. Borgida, "Representing the UMLS® Semantic Network using OWL (Or "What's in a Semantic Web Link?")", Proceedings of the 2nd International Semantic Web Conference (ISWC), October 2003, Sanibel Island, Florida.
9. V. Kashyap, C. Ramakrishnan, C. Thomas, D. Bassu, T. C. Rindflesch and A. Sheth, "TaxaMiner: An Experimentation Framework for Automated Taxonomy Bootstrapping", Technical Report, Computer Science Dept., University of Georgia, March 5th 2004, http://lsdis.cs.uga.edu/~cthomas/resources/taxaminer.pdf
10. C. Behrens and V. Kashyap, "The "Emergent" Semantic Web: A Consensus approach for Deriving Semantic Knowledge on the Web", Real World Semantic Web Applications, Frontiers in Artificial Intelligence and Applications, Vol 92
11. V. Kashyap, "Design and creation of Ontologies for Environmental Information Retrieval", Proceedings of the 12th International Conference on Knowledge Acquisition, Modeling and Management, October 1999, Banff, Canada.
12. V. Kashyap, C. Behrens and S. Dalal, "Professional Services Automation: A Knowledge Management Approach using LSI and Domain Specific Ontologies", Proceedings of the 14th International FLAIRS Conference (Florida AI Research Symposium), Special track on AI and Knowledge Management, May 2001, Key West, Florida, USA.

13. E. Mena, V. Kashyap, A. Illarramendi, and A. Sheth, Imprecise Answers in Distributed Environments: Estimation of Information Loss for Multi-Ontology Based Query Processing. International Journal of Cooperative Information Systems

14. E. Mena, A. Illarramendi, V. Kashyap and A. Sheth, "OBSERVER: An approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies", Distributed and Parallel Databases – An International Journal, Volume 8(2), April 2000

15. V. Kashyap, K. Shah and A. Sheth, "Metadata for building the MultiMedia Patch Quilt", MultiMedia Database Systems: Issues and Research Directions, Springer Verlag 1995, S. Jajodia and V. Subrahmanian (editors).

16. V. Kashyap and M. Rusinkiewicz, "Modeling and Querying Textual Data using E-R models and SQL", Proceedings of the Workshop on Management of Semi-Structured Data in conjunction with 1997 ACM International Conference on Management of Data (SIGMOD) Tucson, Arizona, May 1997

17. OWL Web Ontology Language Overview, http://www.w3.org/TR/owl-features

Semantic Gossiping:
Coping with Heterogeneous Semantic Knowledge Management Systems in the Large

Karl Aberer and Philippe Cudré-Mauroux
School Of Computer and Communication Sciences
EPFL, Lausanne, Switzerland
{karl.aberer, philippe.cudre-mauroux}@epfl.ch

Coping with heterogeneous systems in the large

With the creation and wide adoption of Semantic Web standards like RDF or OWL, a new breed of semantic networks is about to appear. For the first time, we can expect large numbers of knowledge management systems to interoperate using common languages to express the semantics of the data they share or seek. We however foresee semantic heterogeneity to surface once more as a key problem in information integration, given the scale and variety of the systems we are dealing with: Indeed, we realistically cannot expect common global upper-ontologies to capture with sufficient adequacy the requirements of all the different parties in our heterogeneous environment. Custom ontologies will be developed for various application needs, thus endangering global semantic interoperability by introducing local concepts and properties. Also, the situation is somewhat complicated by the fact that ontologies will not be static in such environments, but will tend to evolve, appear or disappear dynamically as systems join and leave the network.

Cleary, there is a need and potential to develop new semantic integration techniques here, since traditional approaches can neither cope with the scale, nor with the dynamicity of such environments. Observing that fostering semantic interoperability requires some form of agreement or consensus among information sharing parties we focus on mechanisms for establishing such agreements in the large. Unlike traditional approaches, like database schema integration or ontology engineering, we do not require the agreements to be static, global or even accurate. Rather, we consider the existence of mutual local agreements only, and expect global semantic properties to emerge from the continuous interactions of autonomous entities in a self-organizing manner. Establishing semantic interoperability in the large can be understood as studying the dynamics of a complex, self-organizing system. Local communication and decision-making of information sharing agents generate the dynamics of this system. When agents locally reach acceptable agreements that are as consistent as possible with the information they receive, the global system reaches a state that embodies what we call the global semantic agreement. Since this state emerges from the dynamics of a complex system, we call the resulting global semantic agreement also the *emergent semantics* of the semantic interoperability system.

Semantic Gossiping as a new semantic reconciliation technique

Following the principles outlined above, we have developed a concrete approach that we termed Semantic Gossiping [1] where we obtain semantic interoperability in a bottom-up, semi-automatic manner without relying on global semantic models. We assume that some local agreement (e.g., ontology mappings) exists between systems using different ontologies to model their data. Fig. 1 below represents a network where seven distinct semantic systems are related by various semantic translation links. We call such networks *semantic overlay networks*. Requiring some initial consensus between pairs of systems is not that stringent, given the recent developments on automatic and semi-automatic schema and ontology alignment techniques.

Semantic gossiping realizes now sharing of local knowledge on translations by using the principle of gossiping that has been successfully applied for creating useful global behaviors in decentralized environments: When different schemas are involved, local mappings are used to further distribute a search request into other semantic domains. Inferring agreements from transitive closures on local translation links, we can relate systems that would have been semantically disconnected otherwise. By comparing the initial search query with the search queries forwarded through series of translation links (*transformed* query), we characterize the quality of the agreements obtained in this manner in two ways:

- We iteratively determine a *syntactic similarity* measure that accounts for the net loss of information (e.g., attributes which cannot be mapped) between the original and the transformed query
- We determine *semantic similarity* values that reflect the degree of semantic agreement (e.g., precision of attribute mappings) that can be achieved by two systems given a translation link.

Whereas syntactic similarity values can be computed locally by comparing a given query with its transformed version, semantic similarity values have to be derived from the context provided by the semantic overlay network. Thus, we do not depend on any hypothetical local expertise to assess the correctness or precision of the mappings, but rather derive these values by aggregating evidences throughout the network. One possibility to realize this is to analyze cycles in the translation graph that allow us to compare any initial query with a syntactically similar query sent through series of translation links and returning to the originator of the query (e.g., a query issued by *A*, going through *B* and *C* and returning to *A* in the figure below). Another possibility is to analyze the results returned by the systems to which the query was forwarded to determine the end-to-end semantic gap between two semantically heterogeneous systems.
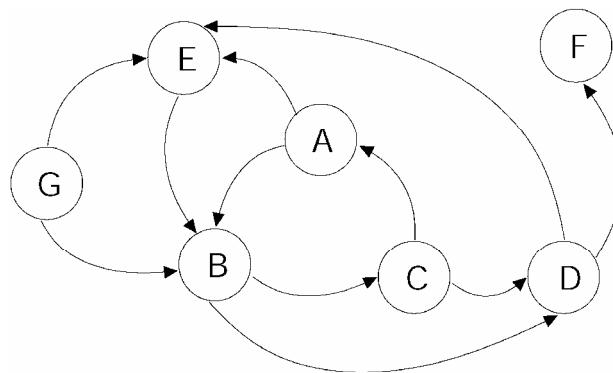

Fig. 1: A network of translation links mapping semantically heterogeneous systems

At a global level, we can view the problem as follows: The translations between domains of semantic homogeneity form a directed graph. Peers in the network compute syntactic and semantic similarity values whenever they want a forward a query to a heterogeneous semantic domain using a translation link and receive from forwarded queries feedback that in turn is used to improve the assessment of semantic similarity. The decision on whether or not it is useful to forward a query to a given system is dependent on the similarity values obtained (*per-hop* forwarding behaviour). In [2], we showed how such heuristics can be applied to maximize recall while limiting the total number of messages generated to answer a given query. Implicitly, this is a state where a global agreement on the semantics of the different schemas or ontologies has been reached. Furthermore, we devised techniques to automatically improve the quality of pre-existing translations based on the results of the similarity computations (*self-healing* semantic networks, see [3]).

A down-to-earth system: GridVine

To demonstrate our approach, we implemented Semantic Gossiping in a concrete system called *GridVine*[4]. In GridVine, we address the problem of building scalable semantic overlay networks by following the principle of data independence and separating the logical from the physical layer (see Fig. 2): At the logical layer, we support various operations necessary for the maintenance and use of a semantic overlay network, including attribute-based search, schema management, schema inheritance and schema mapping. We provide these mechanisms within the standard syntactic framework of RDF/OWL. At the physical layer we provide efficient realizations of the operations exploiting a structured DHT overlay network, namely P-Grid [5] which supports efficient location of resources in a network based on resource identifiers.
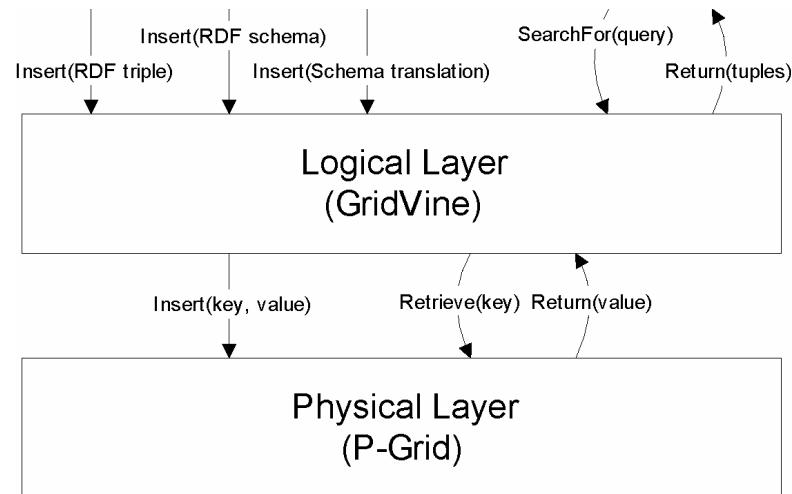
Fig. 2: GridVine: applying the principle of Data Independence

The separation of a physical from a logical layer allows us to process logical operations in the semantic overlay using different physical execution strategies. In particular we identify iterative and recursive strategies for the traversal of semantic overlay networks as two important alternatives. At the logical layer, we support semantic interoperability through schema inheritance and Semantic Gossiping. To the best of our knowledge, GridVine is the first semantic overlay network based on a scalable, efficient and totally decentralized access structure supporting the creation of local schemas while fostering global semantic interoperability.

Conclusion

Studying semantics in the context of large-scale systems will lead to a new breed of approaches and systems to tackle the inherently difficult problem of semantic interoperability. In particular, in these systems the "network" provides a rich source of knowledge that can be processed in a decentralized and automated manner. We can expect that tools for studying complex systems, such as graph theory and dynamical systems theory, will start to play an important role in better understanding the global behavior of such systems and will open exciting avenues for future research.

References:

[1] *A Framework for Semantic Gossiping,*
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth
SIGMOD Record, 31(4), December 2002

[2] *The Chatty Web: Emergent Semantics Through Gossiping,*
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth
Proceedings of the Twelfth International World Wide Web Conference (WWW2003), 20-24 May 2003, Budapest, Hungary.

[3] *Start making sense: The Chatty Web approach for global semantic agreements*,
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth,
Journal of Web Semantics, 1 (1), December 2003

[4] *GridVine: Building Internet-Scale Semantic Overlay Networks*,
Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth, Tim van Pelt,
Third International Semantic Web Conference (ISWC), Hiroshima, Japan, 2004.

[5] *P-Grid: A Self-organizing Structured P2P System*
Karl Aberer, Philippe Cudré-Mauroux, Anwitaman Datta, Zoran Despotovic, Manfred Hauswirth, Magdalena Punceva, Roman Schmidt
SIGMOD Record, 32(2), September 2003

On Knowledge Representation issues

By Alex Abramovich
Gordon College, Israel
webdao@yahoo.com

Abstract.
Knowledge Representations issues take on special significance in the light of development of the novel Web's reality that involves the Semantic Web, GRID, P2P and other today's ITs. In contrast to the previous IT evolution's stages, the recent one utilizes ontology as separated resource. An elaborate knowledge representation approach implies an efficiency of knowledge-based systems and their interoperability. This paper deals with Ontology Engineering approach that allows both build and generate the consistent dynamic autonomous knowledge-based systems.

Keywords: Knowledge Representations (KR), reasoning, human activity, domain world, private world, ontology

1. Introduction

Range of Knowledge Representations' issues, include, but are not limited to:
- measure of KR approach's adequacy to the represented knowledge
- measure of knowledge role with respect to the goal that is trying to be achieved
- measure of overall quality of knowledge within the knowledge representation
- measure of knowledge uncertainty for the knowledge utilization by the autonomous system
- measure of the consistency of knowledge that is provided by the autonomous software agents or by the service providers
- measure of the ontologies' role in autonomous systems
-

Proceeding from the assumption that human behavior is defined by his knowledge, we have a right to expect a successful evolution of autonomous systems only under the stipulation that it exists a reliable KR foundation. Unfortunately, underdetermined system of KR's terminology itself produces numerous problematical KR approaches. In this paper we will attempt to look at aforesaid KR issues as at reasoning's problems and to subordinate knowledge representation to reasoning one.

1.1 What do we mean by knowledge

In order to assess which types of knowledge representation are appropriate for which type of information, including corresponding performance measures as well as to consider other KR issues, it is necessary to define what we mean by knowledge.

Consider some knowledge definitions from Google:

The act or state of knowing; clear perception of fact, truth, or duty; certain apprehension; familiar cognizance; cognition.

 "Knowledge, which is the highest degree of the speculative faculties, consists in the perception of the truth of affirmative or negative propositions." *Locke.*

That which is or may be known; the object of an act of knowing; cognition; -- chiefly used in the plural.

"There is a great difference in the delivery of the mathematics, which are the most abstracted of knowledges." *Bacon.*

"Knowledges is a term in frequent use by Bacon, and, though now obsolete, should be revived, as without it we are compelled to borrow "cognitions" to express its import." *Sir W. Hamilton.*

"To use a word of Bacon's, now unfortunately obsolete, we must determine the relative value of knowledges." *H. Spencer.*

"That familiarity which is gained by actual experience; practical skill; as, a *knowledge* of life. "Shipmen that had knowledge of the sea." *1 Kings ix. 27."*

As we see, knowledge is one of those concepts, concerning which everybody has own opinion. Nevertheless, the last one seems the most operable. Practically, it equates knowledge with an activity representation. In any case, (since the practical skills is used by human in his activity) it means that knowledge is a mental instrument, with is used for the human activity achievement.

Thus it is possible to say that knowledge is an instrument of reasoning.

1.2 Why does a human think?

Before we will define the reasoning model, it is appropriate to put a question – Why does a human think?
"Reasoning is a mediate generalized reflecion of appreciable and regular dependences of reality."[1] As such it is an instrument of human life cycle providing.
"Thinking and acting are the specific human features of man. They are peculiar to all human beings. They are, beyond membership in the zoological species homo sapiens, the characteristic mark of man as man."[2]
Since a human life cycle is constituted by set of profession/living activities, reasoning serves these activities' achievement.
At that, knowledge is used as the human activities awareness.
To Ludwig Edler von Mises [2] "human activity is a goal-seeking behavior" and "human action is necessarily always rational".
And so by human activity we mean:
Definition 1. Human activity is time-, place-, state-, and event- ordered set of multidisciplinary actions aimed to achievement of socio-specified goal.

1.3 Activities' types

In spite of the obvious differences of social institutions and persons, their life cycle as set of activities, on closer examination, seems in the following way:
1. An activity (activities) that provides the means of subsistence (both profession and other socially specified activities),
2. Properly living activities, namely, learning, execution, repair, protection, *an advancement of results*, *supply*, *an analysis and control.*
The first activity (activities) belongs to certain area (areas) of expertise (domain). As domain activities we differentiate *domain generic activities* and *private activities.*
The properly living activities we designated by the common name of *generic living activities.*
Domain generic activity is a basis framework of actions, operations and/or activities aimed to achieve one or more *domain specific goals*, where domain goal is a socio-claimed product or service.
Private activity is an adapted domain/living generic activity provided by a social unit, where by *a social unit* we mean a government, an enterprise, a community and a person.
Thus we differentiate the following activity's types: generic living activities, domain generic activities, and private activities.

1.4 A human mental activity

Now, there is time for *correlate with each* other a social unit's life cycle, activities, knowledge and reasoning:
Definition 2. Reasoning is a human mental activity that operates with human activities knowledge for the purpose of the social unit's life cycle providing.

At that, on the *level of the social unit's life cycle organization* reasoning operates with activities as with data type, and on the *level of the activities' implementation* it operates with an activity's components (see below) as with data types.

It is important to note that reasoning trace is a certain algorithm, and its data types' names constitute *a reasoning ontology.*

We differentiate life cycle of domain (*domain world*) and life cycle of a social unit (*private world*).

Reasoning's algorithm of domain world we denominate as *domain world activity,* and, analogously, reasoning's algorithm of private world we denominate as *private world activity*.

We emphasize the domain world activity and the private world activity, since, as a matter of fact, they define *behavioral/management models* of domain or of a social unit.

Definition 3.
1. Domain world activity (Adw) is a resultant activity of the domain community, composed of domain generic activities (owned by domain experts) and private activities (owned by the other domain community's members), aimed to the domain life cycle providing;
2. Private world activity (Apw) is a resultant activity of the private profession/living activities, owned by social unit and aimed to the private life cycle providing.

2. THE Reasoning

The suggested Ontology Engineering approach forms a core of THE (Total Human Experience) Web project. In the network of THE Web project it is proposed to build an integrated Web knowledge resource (THE KB) with the purpose of the exhaustive Web service providing of the profession/living activities. THE Web service will be realized by an integrated multi-agent system (THE MAS) under multilevel dispatching.

THE KB is constituted by human activities' representation and derived ontological as well as causal environment.

Human Activity is represented in form of Activity Proposition (AP) on the Reasoning Language (RL). RL is THE Web's internal language that data types are represented by Core Ontology (CO), *Domain Ontology (DO) (as CO* extension), Private activity's ontology (PO) as a certain DO extension, Domain World Activity's Ontology (DWO) as DO extension, Private World Activity's Ontology (PWO) as CO and a certain DOs extension that are derived from corresponding activities' propositions (see below).

AP represents an algorithm of the activity performance' steady states transformation. So called Steady Reasoning (SR) serves (validates and directs) this algorithm performance. SR operates the following knowledge types:

- A private activity's *initial state* (AIS)
- A state *transforming* private activity (STA)
- Set of *possible STA effect* states as result of STA,
- where a state knowledge includes
- a state ontology,
- a state determinant,
- determinants of state's components;
- and a private activity knowledge includes
- an activity ontology,
- an activity's states,
- an instrumental private activities toolkit and
- an activity's determinant.

At that, an activity's and activity state's determinant is a semantic framework of its ontology's components that is a mandatory for inheritance at all generations.

RL provides also transient reasoning's means for the purpose of Transient Reasoning (TR) achievement. In addition to above mentioned, TR operates the following knowledge types, derived from THE KB:

- Network of generalized causalities,
- Generalized cause (that is, set of causes that derive from the same state the same effect),
- Causality determinant.
- RL is interpreted by THE MAS reasoning framework (THE Reasoning).
-

THE Reasoning process is provided by the following agents:

- *Recognizer* that recognizes determinants of activities and activities' components,
- *Executor* that executes the AP's sequence of operations,
- *Predictor* that predicts an eventual course of events,
- *Reason _detector* that detects a reason of deviation from the specified steady state and generates a target setting,
- *Activity_generator* that derives from the KB a new activity proposition as a discovered (or received) problem solving.

3 Activity Proposition

This paper is not RL presentation. Therefore we will consider RL features that concern Knowledge Representations issues only. RL is a procedural, a markup, an ontology language as well as an action language, destined for the description of reasoning that required for the activities' performance.

As a procedural language it allows to describe an activity's algorithm.

As an action language it represents a causality in the form of a triplet {I,C,E}, where I is an initial condition, C is a cause, E is an effect.

As an ontology language it allows to input both concepts and concepts' relations.

As a markup language it provides a semantic marking of AP text that allows the ontology mapping.

Activity Proposition plays a part of a canned program and at the same time it is considered as knowledge module. At once on completion of AP design, it occupies THE KB position in compliance with its causal interpretation.

At that, it is necessary to note that we extend concept of an activity actor beyond the social units. We mean by Activity Proposition (AP) a semantically marked description of *purposeful system of operations that producible by human(s) and/or by service provider(s) and/or by apparatus(es) and /or by software applications.*

At that,
- *Activity's ontology* is AP text's remainder of deletion both RL's terms and lexical forms as well as semantic tags (that is, a semantic ordered set of words (*ontology units*) used for AP representation).
- *Ontology unit's semantics* is fixed by the nearest semantic tags (opening and closing) and
- *Ontology unit's meaning* is Web, THE Web or private resource.

3.1. Personal world

Private world (PW) is constituted by set of actual private profession/living (p/l) activities derived from Basis and Domain generic activities. At that, ever it remains the PW composition, namely, *learning*, *practice (*that is, an execution of a socially specified activity(ies) that provide(s) the livelihood), *repair*, *protection*, *an advancement of results*, *supply*, *an analysis and management.*

Every p/l activity is correlated with others by time, by place, by preferences and by cost. Space of correlated p/l activities is rank-ordered by APpw that represents a scenario of parallel/sequential executable private p/l activities, which are marked by a special set of tags. RL keeps *AP special sets of semantic tags* that define an activity's position in the personal world. APpw provides a semantic sharing of private p/l activities as well as of private p/l ontology. Private world's activity represented by APpw is aimed to the achievement of it's owner p/l goals with a cost minimization.

A priority of APpw's performance produces *a particular causal stipulation* of private activities as well as *particular reasons of response* to external occurrences (*a private logic*). A corresponding APpw ontology has, therefore, *private semantic features.*

A private logic induces interoperability issues both on the profession and on the living level that must be considered as an operation problems both of PW management and of PWs interaction. In case that a response

to an external occurrence is not contradict APpw performance's logic it will be executed. If not, a response's execution will hurt the PW.

The response's motivation takes on special significance for reasoning, particularly, for Reason _detector and Activity_generator.

3.2 Domain world

Domain world (DW) is constituted both by domain generative activities and by private activities of professional communities, of enterprises and of specialists. THE Web engine keeps *AP special sets semantic tags* that define a profession position of all domain world participants. A corresponding domain world AP (*APdw*) provides a semantic sharing of domain activities as well as of domain ontology.

Domain world activity, represented by APdw, is aimed to the achievement of domain socioeconomic, sociopolitical and socio-productive goals with cost minimization. APdw performance is achieved via domain Web portal.

3.3. THE self-organization

Ontology constitutes the external level of human experience's knowledge representation. Every Onto-unit has THE KB's multi-semantic position represented by set of DW related triplets (APdwName, APName, SemanticTag) as well as by PW related triplets (APpw, APName, SemanticTag). At that every Onto-unitName is accomponied by links to DOName or CO (that is, to Onto-unit parent's name).

This Ontology organization grounds an opportunity of the interoperability issues' solving.

Recognition of an activity's determinant in the current input activates THE Reasoning process.

3.3.1 Target setting's processing

A target setting as an output of Reason_detector or due to a customer's initiative is sent to Activity_generator in form of an initial and a finite state.

Using knowledge of activities states' determinants, Activity_generator search the corresponding THE KB nodes and AP paths between them.

The next problem is a correction of one of this paths with the purpose of utilization it by PW owner.

This correction is a type of a semantic translation that represents a sequential revision of the inter-tags' spaces. An impossibility of the inter-tags' spaces filling is fixed as a problem that involves a target setting for Activity_generator.

As result of this recursive procedure is a new AP.

4. Measure of KR approach's adequacy to the represented knowledge

Suggested Ontology Engineering approach deals with unified model of above mentioned knowledge types (see paragraph "THE Reasoning") representation. AP representations of existing software tools/agents/applications utilization's procedures will extend THE KB. Representations ones will be used as procedures of access to these resources. In the same way it represented an implement's, an apparatus', equipment's, a sensor's (and so on) utilization as an activity states' components representations. At that, principles of operation of above-named devices are represented by means APs too.

Thus THE Web operates with active knowledge forms, for which AP representation is adequately.

5. Measure of knowledge role with respect to the goal that is trying to be achieved

According to M. Polanyi [3], the components of an optimally organized system must not be further partible in the certain, defined for this system, ratio. M. Polanyi made out of a system's components at a ratio of their contributions to the goal achievement. A component's position in the system's organization defines its semantics. Its contribution defines the component's significance.

Due to RL notation, semantic tags define an ontology unit's contribution to the AP, and an ontology unit is utilized as a pointer of a related resource that details an access procedure (or this knowledge principle of

operation). Thus THE KB represents knowledge system, ordered in M.Polanyi sense, and THE KR approach provides a contribution of every knowledge unit to the goal's achievement.

6. Measure of overall quality of knowledge within the knowledge representation

Since an ontological design is provided by domain expert or by APpw owner, the overall quality of knowledge within the knowledge representation is depended on its author's skill level or on the APpw owner's preferences that always may be submit for consideration of new customer. THE engine provides the AP designers' rating and chooses (in the presence of choice) the best AP version.

7. The ontologies' role in autonomous systems

Among a manifold of an ontology definitions the Protégé' one is the most close to RL notation:
"Ontologies are explicit specifications of the types of resources that exist and possible relationships between them, and specific instances of concepts in the ontologies" (http://protege.stanford.edu/).
THE Reasoning utilizes an ontology as a semantically ordered set of Web resources' pointers. Similarly, a human operates on concepts. At that, as concepts it is used both scientific/technical/common terms and arbitrary identifiers of arbitrary objects sets as well as of various process' parts, of states, of situations and so on.
Therefore in THE notation the problem of primary importance is *a reconstruction of the individual conceptual system* (that is, the private ontology mining). A discovery of corresponding DO/CO terms grounds a semantic translation of private situation to the DO/CO specification. Only after that it is possible generating for customer a personalized Web service.
Remind that in the previous chapters we considered an ontology as data type names' space.
Thus, since reasoning process is grounded by conceptual schemes,
an ontology plays a part of primary importance for all knowledge based systems include autonomous ones.

8. Conclusion

We considered a particular Knowledge Representations' approach. We simplified a problem by consideration unified KR form called Activity Proposition. We consider that it optimally satisfies both human and machinable reasoning and that in such a way we are able to build of a personalized Web service.

Reference:
1."Reasoning is a mediate generalized reflecion of appreciable and regular dependences of reality." (http://azps.ru/articles/proc/proc9.html)
2.Ludwig Edler von Mises, "*Human Action: A Treatise on Economics*",The Foundation for Economic Education, Inc. Fourth revised ed., 1996, printed 1998
3. M. Polanyi, Personal Knowledge, Harper & Row, New York, 1958

Alex Abramovich received his M.Sc. in mathematics from State University of Rostov-on-the-Don, Russia in 1975. He is an author of a programming language YASEN for mathematical physics equations solving by difference schemes (1976) that until now is used by military-industrial establishment of Russia. As research fellow of Rostov's University (1975 -1996) he was involved (as an executor, an analyst, knowledge engineer, Master Solution Architect) in the development of unmanned system of federal level for aircraft building, for ferrous metallurgy, for hydro-acoustics, for mechanics of continua and other areas. From 1996 he is citizen of Israel. As Gordon College's fellow he is engaged to work in the Distance Learning group as group leader. During the last 7 years he independently investigates problems of Ontological Engineering. He is an author of user-centric oriented ACTIVITY Language (ACTIVITYL, 2003) and of Reasoning Language (RL, 2004).

Ontologies for the Semantic Web: Can Social Network Analysis Be Used to Develop Them?

Henry M. Kim
Schulich School of Business, York University
hkim@schulich.yorku.ca

Abstract

   According to Tim-Berners Lee, the inventor of the WWW, a semantic Web in which software agents find meanings of terms that describe tasks it performs is the next progression of the Web. Ontologies as repositories of these machine-interpretable meanings are key to his vision. However, ontologies are distributed and not, and will likely never, be centrally organized. Enabling agents to find the right meanings then is an important challenge for realizing the semantic Web. As ontologies evolve, they will likely form in clusters exhibiting small-world effects, just like web pages. In this paper, questions about bring to bear findings from social network analysis to the design of these ontologies are raised. One questions stems from the argument that ontology use over a competing technology (XML) may occur when mitigating uncertainty is important. A research direction to study social networks for dealing with uncertainty then is posited.

Contact:
Prof. Henry M. Kim
Schulich School of Business
York University
4700 Keele St., Toronto, Ontario, Canada M3J 1P3

Tel: 1-416-736-2100 x77952
Fax: 1-416-736-5687
Email: hkim@schulich.yorku.ca

Key Words: Ontologies, Semantic Web, Data Modeling, Social Network Analysis

Introduction

The WWW is inherently a set of standards such as HTTP (HyperText Transfer Protocol) and HTML (HyperText Markup Language) for transmitting and rendering hypertext, developed to be consistent with Internet standards. Just with these standards and a web browser, performing meaningful tasks with hypertext documents is mainly a human endeavor. According to Tim Berners-Lee, the inventor of the Web, this is a limited use of the Web's possibilities [Berners-Lee *et al.* 2001]: "computers will find the meaning of semantic data by following hyperlinks to definitions of key terms and rules for reasoning about them logically. The resulting infrastructure will spur the development of automated Web services such as highly functional agents." Efficient computers, not error-prone humans, will then be able to perform rote tasks. In this vision, meanings that computers can find and reason about are represented using ontologies.

An ontology is a data model that "consists of a representational vocabulary with precise definitions of the meanings of the terms of this vocabulary plus a set of formal axioms that constrain interpretation and well-formed use of these terms" [Campbell and Shapiro 1995]. A sufficiently expressive ontology represents enough knowledge as computer encoded instructions and data to enable automated instruction execution. Because precise definitions and axioms exist, proper interpretations by a computer or a decision maker that did not develop the definitions and axioms are possible. There are numerous ontologies available on the

Web—everything from "light-weight" ones [Uschold 1998] that represent taxonomies of terms with little or no meanings formally represented, such as those for Yahoo!™ [Labrou and Finin 1999], to those that represent meanings for specific community of users, such as VerticalNet™ ontologies [Das *et al.* 2001], to an ambitious endeavor to represent common-sense meanings of the world [Lenat 1998]. For software agents, these ontologies serve as dictionaries with which the agents can determine meanings and proper interpretations of terms that describe tasks that need to be performed. Imagine the agent's dilemma though: There are numerous dictionaries within which there are numerous definitions, and the term that is given to the agent to define may not be readily found or its proper meaning may differ from a symbolically-identical term with a different meaning. This is analogous to how Yahoo!™'s primary web page seems overwhelming to an uninitiated user, or a searcher's frustration at typing 'jaguar' on Google™ to investigate felines and being referred to sites about cars and computers.

At least human users are intelligent enough to choose appropriate branches on Yahoo!™'s classification tree to "drill down" or add other keywords to their Google™ search. There are efforts from AI (semantic inter-operability), database (schema integration), and information retrieval (IR) communities to enable computers to mimic such intelligence. In fact Google™'s search algorithms are a result of IR work. These efforts can be informally characterized as locally-aware but globally-unaware, and locally-unaware but globally-aware. The first describe those, such as mapping techniques, that take advantage of relationships between terms within some cluster of repositories (ontologies, databases, or documents), but not between clusters; the second describe those, such as co-occurrence techniques, that take advantage of tendencies in which all terms are generally represented. Is it possible to develop an approach to design and use repositories, specifically ontologies for software agent use, based on a hybrid, locally-aware and somewhat-globally-aware approach? In this paper, research towards this approach is initiated by exploring, then relating to ontology development, a phenomenon that can be characterized as locally-aware and somewhat-globally-aware: How people find other people.

Relating to Social Network Analysis

The small-world effect has been attributed as the explanation for the "six-degrees of separation" result of Milgram's experiment [Watts and Strogatz 1998]. Adamic [2001] has discovered that web pages exhibit this effect. If web pages whose hyperlinks are used by humans are organized in small-worlds, then one can suppose that ontologies whose hyperlinks are used by software agents will be organized in small-worlds, if not already so organized. Though there are efforts at standardizing upon use of common terms for all ontologies, such as UpperCyc [Cycorp 2002], most researchers believe that these terms and definitions will not be globally used, but rather that numerous domain, industry-, value-chain, and company- specific ontologies will evolve that are incomplete and inconsistent with respect to each other. Emergence of competing and non-complementary XML (extensible Markup Language) tag sets is a testament to this; XML is the standardized syntax and format for representing structured data over the Web, and is the *de facto* language for transporting instances (facts) represented in ontologies.

In fact, it can be argued that prevalent use of XML can preclude widespread adoption of ontologies. Kim [2002] makes the following analogies to argue this point. HTML use is likened to paper and pen use for letters. Data can be encoded on paper by an author, but its interpretation is left solely to the recipient, who brings to bear an understanding of natural language and knowledge about the author for interpretation. XML use is likened to use of business forms. The recipient brings to bear knowledge of the format and conventions of the business form to bear in interpreting that form. As long as the author follows the format and conventions, s/he does not need to know the recipient, and vice versa. Accurate interpretation is possible even if the author and recipient do not share a common natural language. Business forms can be more efficiently processed because the recipient can, for instance, sort by looking at one field where the indexed data should always be rather than perusing the whole form. Taking advantage of this, lower-paid data entry clerks could be hired to perform such tasks as well as more expensive domain experts. Similarly, data structured in XML versus unstructured HTML can be efficiently processed using computers, not humans. The proviso is that there must be very consistent, informal understanding of meanings and proper uses of terms between authors and recipients of a given XML document. Ontology use is likened to use of business forms with standard operating procedures (SOP's) and requisite training to properly apply SOP's. Where interpretation and processing of

business forms is not as rote and more uncertain, SOP's and training assist clerks' decision-making. Ontologies are still meant for computer use, but meanings and terms are represented more formally for computer interpretation. Inasmuch as SOP's and training are more expensive than use of business forms only, so is ontology use more expensive than XML use; just as increase in uncertainty of processing necessitates SOP's and training, so may increase in uncertainty of machine interpretation compel ontology use.

Relating this back to social networks, those networks that exhibit small-world effects but inter-relate to mitigate uncertainty should be more carefully studied. This may yield clues to circumstances in which ontologies will be adopted—XML use is much more prevalent, so a clear advantage of ontology use must be demonstrated. Characterization of these circumstances can then be factored into design and use of ontologies that will be practically used.

Concluding Remarks

This paper discusses an initiation of a program of study rather than the results of one. It does provoke relating two somewhat disparate fields of study, and endeavors to make a contribution to the advancement of a ubiquitous technology, the WWW. The research questions raised are the following. How can de-centralized ontologies with no central organization be designed for a semantic Web that enables a society of software agents to automatically perform tasks that are currently done manually? What research, methods, and tools from the social network analysis fields can be brought to bear for this design? Can a characterization of circumstances in which ontology use will prevail over XML use be discerned? And can social networks that exist to mitigate uncertainty be studied for this characterization? The semantic Web is a lofty vision, not likely to be realized in full for a long time. However, it may be partially realized, and it is believed that this research will contribute to that.

References

[Adamic 2001] Adamic, Lada A., 2001, "Network Dynamics: The World Wide Web", *Ph. D. Thesis*, Department of Applied Physics, Stanford University, Stanford, CA.

[Berners-Lee *et al.* 2001] Berners-Lee, Tim, Hendler, James, and Lassila, Ora, 2001, "The Semantic Web", *Scientific American*, May.

[Campbell and Shapiro 1995] Campbell, A. E., and Shapiro, S. C., 1995, "Ontological Mediation: An Overview", In: *Proceedings of the IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing*, AAAI Press, Menlo Park, CA.

[Cycorp 2002] Cycorp, 2002, "Welcome to the Cyc Public Ontology", Available – http://www.cyc.com/cyc-2-1/index.html, Accessed: May 30, 2002.

[Das *et al.* 2001] Das, Aseem, Wu, Wei, and McGuinness, Deborah L., 2001, "Industrial Strength Ontology Management", In *Proceedings of the International Semantic Web Working Symposium*. Stanford, CA, July.

[Kim 2002] Kim, Henry M., 2002, "Predicting how the Semantic Web Will Evolve", *Communications of the ACM*, February.

[Labrou and Finin 1999] Labrou, Y. and Finin, T., 1999, "Yahoo! as an Ontology - Using Yahoo! Categories to Describe Documents", In: *Proceedings of the 8th International Conference on Information and Knowledge Management*, Kansas City, MO, November, pp. 180-7.

[Lenat 1998] Lenat, Doug, 1998, "From *2001* to 2001: Common Sense and the Mind of HAL", In:, Ed: Stork, David G., MIT Press: Boston, MA, pp. 193-210.

[Watts and Strogatz 1998] Watts, D.J. and Strogatz, S.H., 1998, "Collective Dynamics of "Small World" Networks", *Nature*, 393, pp. 440-2.

[Uschold 1998] Uschold, Mike, (1998), "Where Are the Killer Apps?", In: *Proceedings of ECAI-98 Workshop on Applications of Ontologies and Problem-Solving Methods*, Brighton, England, August.

An ontology model for the exploitation of knowledge in group decision making settings

Christina Evangelou and Nikos Karacapilidis
Industrial Management and Information Systems Lab, MEAD, University of Patras
26500 Rio Patras, Greece
{chriseva, nikos}@mech.upatras.gr

Introduction

Group decision making, as a collaborative process of shaping a position, opinion or judgment in order to resolve a problem, attain a goal or seize an opportunity, is indisputably a core organizational activity (Karacapilidis *et al.*, 2003). At the same time, it is a highly complicated task that depends both on the organization's knowledge capital, as well as on the underlying knowledge gathering and sharing activities. Especially in cases where decisions should be made through argumentative discourse, communication (both in social and technical level) imposes a series of quandaries. However, such discourses are valuable for the acquisition and exploitation of the decision makers' knowledge and should be thoroughly exploited. It is widely argued that the amount of knowledge exchange and the level of the decision makers' communication are related to the existence of a common language and terms of reference (Shen, 2003). In this direction, ontologies are a means to provide exhaustive and rigorous conceptual schemata for the specification of semantics within a certain knowledge domain; they establish a shared understanding of different knowledge domains and facilitate the sharing and reuse of pieces of knowledge across diverse groups and applications (Duineveld *et al.*, 2000; Chandrasekaran *et al.*, 1999). Taking into account the above issues, we present in this article an ontology model for the management of knowledge that is embedded in a collaborative decision making setting.

The proposed ontology model

Our approach is based on the assumption that the required knowledge to make a decision is expressed through arguments, whereas a decision that is based on rigid arguments constitutes new knowledge. According to this, knowledge acquisition, knowledge sharing, argumentation and decision making are interrelated processes. We have developed an ontology model that is based on the principal concepts (entities) of knowledge, argument and decision, and the related processes of knowledge sharing, argumentation and decision making. Figure 1 illustrates an instance of the XML schema of the proposed model; as shown, it consists of three main classes, namely *decisionMaker*, *entity* and *process*. The first class contains all the information related to the decision makers' personal and professional background, as well as information related to their behaviour in decision making and knowledge sharing processes. The *entity* class comprises the *knowledge*, *argument* and *decision* subclasses, which are explicitly defined by their *instances* and *identifiers*. Similarly, the *knowledgeSharing*, *knowledgeGathering*, *argumentation* and *decisionMaking* processes are all defined as *processType* subclasses of the *process* class. In this way, all their relevant entities, relationships and rules are assigned to the *initiative*, *mechanism* and *result* classes.
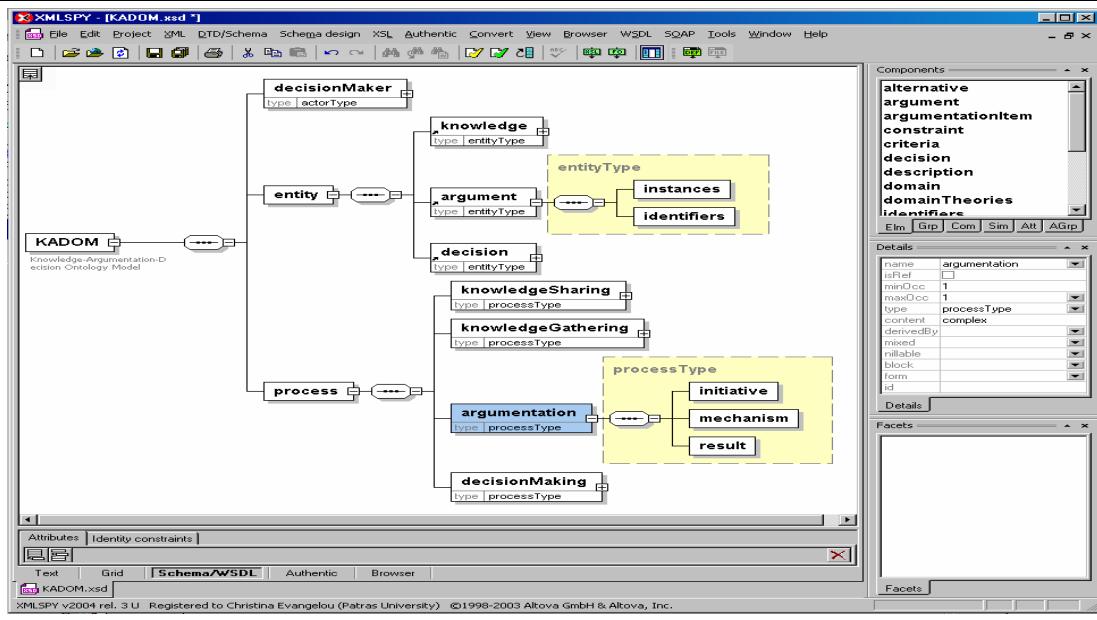
Figure 1: An instance of the XML schema of the proposed model.

The rationality conveyed in the proposed ontology model is that the processes of (argumentative discourse-based) decision making and knowledge sharing should be simultaneously and seamlessly considered in a collaborative setting. Therefore, the *entity* and *process* classes, as described above, are multiply interrelated. For instance, as shown in Figure 2, the *instances* subclass of the *knowledge* class comprises the *argument* and *decision* classes. In such a way, the model declares explicitly that the knowledge embedded in such a setting can be found in the form of arguments and decisions. Furthermore, members of the *argument* class can be found as subclasses of the *knowledgeSharing*, *knowledgeGathering*, *argumentation* and *decisionMaking* classes.
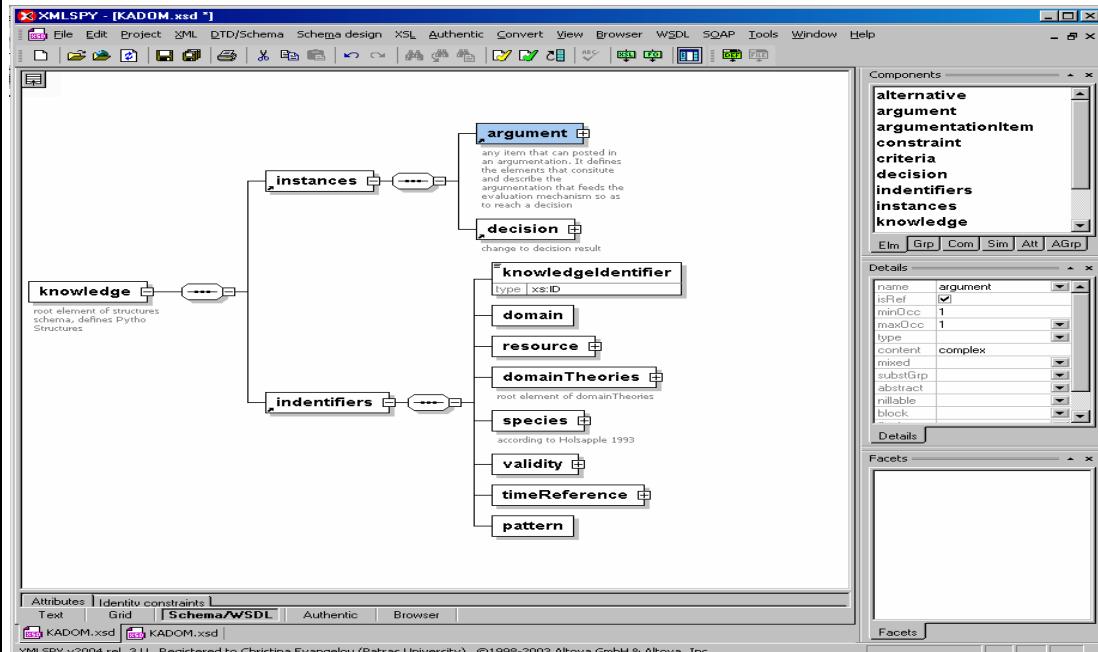

Figure 2: The hierarchy structure of the "knowledge" element.

Focusing on the *argument* class, any kind of statement a decision maker may assert during a discourse is considered as an argument. For instance, the *argument* class consists of the *domain*, *subject*, *alternative*, *criteria*, *support* and *constraint* classes. In such a way, our model interrelates the structure of an argumentative discourse with the knowledge domain of the discourse topic. Furthermore, the *subject*, *alternative*, *criteria*, *support* and *constraint* classes comprise subclasses that refer to specific knowledge domains. In addition, these classes are constructed with respect to semantics used in methods, models and techniques coming from the *Multicriteria Decision Aid (MCDA)* discipline. This is based on the requirement that the argumentation's final goal is to reach a decision; therefore, the discourse structure should comply with a decision making model.

| Criteria concerning the equipment | ```
<owl:Class rdf:ID="maintenanceCriteria">
        <rdfs:subClassof rdf:resource="criteria"/>
</owl:Class>
<owl:Class rdf:ID="maintenanceCriteriaName">
        <rdfs:subClassof rdf:resource="maintenanceCriteria" />
        <location rdf:ID="maintenanceCriteriaName" />
        <request rdf:ID="maintenanceCriteriaName" />
        <workOrder rdf:ID="maintenanceCriteriaName" />
        <response rdf:ID="maintenanceCriteriaName" />
</owl:Class>
``` |
| Criteria concerning the production | ```
<owl:Class rdf:ID="productionCapabilityCriteria">
        <rdfs:subClassof rdf:resource="criteria"/>
</owl:Class>
<owl:Class rdf:ID="productionCapabilityCriteriaName">
        <rdfs:subClassof rdf:resource="productionCapabilityCriteria"/>
        <personnelCapability rdf:ID="productionCapabilityCriteriaName" />
        <equipmentCapability rdf:ID="productionCapabilityCriteriaName" />
        <materialCapability rdf:ID="productionCapabilityCriteriaName" />
        <processSegmentCapability rdf:ID="productionCapabilityCriteriaName" />
</owl:Class>
``` |

Figure 3: Definition of the criteria class for two cases.

To present the proposed model's additional functionalities, we use in the sequel an instance (extension) of it that applies to the manufacturing management knowledge domain. In brief, manufacturing management involves decision making about the product design and development, process selection, plant location and design, capacity management, manufacturing planning and control, quality control, workforce organization, equipment maintenance, product distribution, and inter-plant coordination. All the above are issues to which the proposed model could be applied, provided that the appropriate extensions will be made. For instance, a decision about the equipment maintenance can be considered as a collaborative issue. In this case, the proposed ontology model defines as related criteria the *location of equipment*, the *kind of request*, the *work order* and the *response time*. Similarly, in the case that the discourse is about the production capability issue, the decision makers should use as criteria the capability concerning *personnel*, *equipment*, *material* and *process segment*. For these two cases, Figure 3 presents the proposed ontology model definition of two *criteria* subclasses, namely the *maintenanceCriteria* and *productionCapabilityCriteria* classes, as defined with the use of the OWL syntax (W3C, 2004).

Discussion

The proposed ontology model has been constructed through the exploration and exploitation of concepts and theories originally coming from the disciplines of Knowledge Management, Argumentation Theory, Decision Making and Multicriteria Decision Aid. It can appropriately address the decision makers' requirements to interpret and reason about knowledge during an argumentative discourse. Furthermore, it provides decision makers with the necessary means to communicate individual knowledge, while it facilitates their creative thinking. In addition, it is developed with the use of XML technologies, thus attaining an interoperable, generic, extensible and neutral information modeling. Its structure exploits the advantages of the object oriented paradigm, the relational paradigm and the hierarchical structure furnished by the XML schema. With respect to the above, we argue that the proposed model can be considered as both flexible and expendable, as far as both people communication and machine interpretability are concerned, and may serve the manipulation of diverse knowledge domains (such as manufacturing management, product design, etc.).

References

Chandrasekaran, B. Josepheson, J. and Benjamins, V.R. (1999), Ontologies: What Are They? Why Do We Need Them?, *IEEE Intelligent Systems*, Vol. 14, No 1, pp.20-26.

Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B. and Benjamins, V.R. (2000), WonderTools? A Comparative Study of Ontological Engineering Tools, *International Journal of Human-Computer Studies*, Vol. 52, pp.1111-1133.

Karacapilidis, N., Adamides, E. and Evangelou, C. (2003), Leveraging Organizational Knowledge to Formulate Manufacturing Strategy, in *Proceedings of the 11th European Conference on Information Systems (ECIS2003)*, Naples, Italy, June 16-21.

Shen, W. (2003), Editorial of the Special Issue on Knowledge Sharing in Collaborative Design Environments, *Computers in Industry*, Vol. 52, No 1, pp. 1-3.

W3C (2004), *The World Wide Web Consortium Organization*, http://www.w3.org.

Knowledge Services as Web Services: Representation for retrieval

Panos Georgolios, Kostas Kafentzis, Gregoris Mentzas, Panos Alexopoulos,
National Technical University of Athens, 9 Iroon Politechneiou, Greece
{pgeorgol, palexop, kkafe,
gmentzas}@softlab.ntua.gr
http://imu.iccs.ntua.gr

Introduction

The Semantic Web is an effort to express the information in the web in a machine processable way. This is achieved by annotating documents with meta-information that defines what the documents are about. The Semantic Web is today a major research trend.

In parallel, another trend has emerged as a succession of major knowledge management investments in the previous decade. An increasing number of enterprises are getting interested in commercially exploiting their knowledge assets outside the organizational borders. To harness this opportunity enterprises need to blend their knowledge management systems, that have contributed to the creation of knowledge bases and a plethora of knowledge assets within enterprises, with electronic marketplaces, that provide adequate transaction mechanisms and viable business communities; see e.g. [5], [9], [13]

A contemporary approach to the subject of knowledge provision through the web calls for the utilization of semantic web services in order to create mechanisms that fully exploit the advantages these services offer. To the best of our knowledge, there exist no electronic marketplaces in the knowledge provision domain that use the technologies of the semantic web services.

Services, e-services and knowledge services

In economics and marketing, a service is considered as the non-material equivalent of a good. In the context of these sciences, a service can be defined as an "activity whose output is not a physical product", and its purpose is to add higher-order (i.e. intangible) value to its recipient [12]. Services exhibit several key attributes like intangibility, perishability, non-transportability, non-standardization or heterogeneity, labour-intensity, demand fluctuation and buyer involvement. Services can be delivered by electronic means (e-services) [13].

A specific class of electronic services is knowledge services. A knowledge service utilizes sources of knowledge objects upon which it performs a specific set of functions in order to create value and enable a human agent to act in a knowledgeable and goal-oriented manner. We consider knowledge web services as semantically enriched web services qualified enough to represent knowledge services.

The sources of knowledge objects are various knowledge repositories available online where knowledge objects are stored and organized. These knowledge objects may represent explicit or tacit knowledge.

An example may illustrate the case more eloquently. A Human Resources manager faces a problem with a specific operation in her department. Being unable to solve the problem on her own, she searches for a solution in the web. The solution will come in the form of a knowledge service that will fulfill her specific need. Using an appropriate mechanism she comes up with a methodology– of course in the form of a knowledge object- that applies in the situation. Yet, she may also need to hire an expert to realize the methodology on her behalf. The knowledge service generates a list of experts with relevant experience based on their profiles - which are again represented as knowledge objects. The representation schema of a knowledge service should also aim to support the whole lifecycle of a knowledge service [6].

Representation of knowledge objects

In order to support knowledge retrieval and application, a highly sophisticated knowledge-representation mechanism for the discovery of knowledge-based services is required. For this purpose we make use of an ontological model which main idea is the Information Ontology [1].

Adopting this work, we present the Knowledge Object Ontology for representing knowledge objects as the codification of knowledge assets. Such an ontology comprises (see also [2]) among others:

- A specification of all attributes a Knowledge Object for trading knowledge may possess.

- The value ranges, and – if necessary – supplementing related ontologies – for defining the ranges of attributes used.
- A specification of all links and relationships that may exist between Knowledge Objects (indicating, e.g., that some knowledge object could provide prior knowledge useful for understanding and applying some other knowledge object).

The Knowledge Object Ontology comprises of eleven metadata facets: content, context, community, security, business, transition, application, evaluation, history and domain. The idea behind the faceted description is that, if all the facets are sufficiently described, it should be possible: to assess the content and potential usage and value of a knowledge object comprehensively; to support all processes, transactions and modifications during the lifetime of a knowledge asset describing a knowledge service.

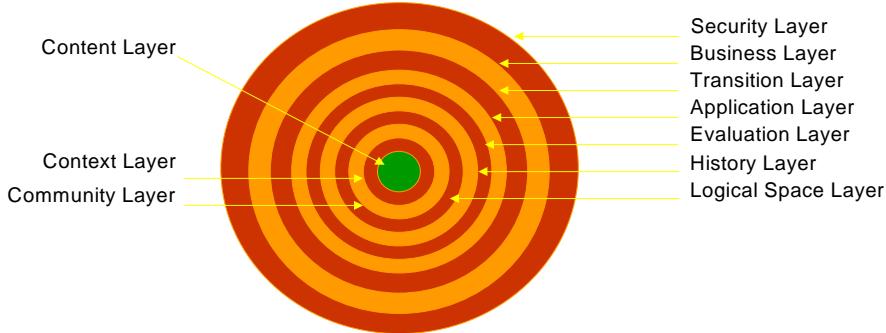Here is a short overview of the content and context facets:

Figure 1: Faceted representation of a knowledge object

The content facet shall describe the core content of an Knowledge Object, i.e. both what it is about (e.g., "this is a textbook about operating systems") and how it is physically manifested (e.g., "the book has 342 pages").

The context facet shall describe under which circumstances a knowledge product may be used and applied in a customer organization. For instance, we could know that some lesson learned should be useful in all marketing processes of car manufacturing companies.

It may be the case that only one of these two central IO description dimensions will be used in a concrete example (e.g., Digital Libraries typically talk only about content, not about context, whereas lessons learned (LL) systems may talk only about the context where some LL could add value), but we discuss both dimensions and feel that it opens promising chances to consider both.

Another interesting aspect of our ontology is the way we represent tacit knowledge. Taking into account the fact that tacit knowledge can not be codified we create knowledge objects following the above mentioned representation schema, where attributes are the expertise, the availability, the price of consulting time and the contact information of an expert. This way we create a knowledge object which constitutes a pointer to an expert.

We need a reference model upon which we build the representation of a knowledge service. In the following section knowledge services are represented as a modification of the daml-s service representation model.

DAML-S representation for k-services

The application of services on the Web is manifested by Web Services [11]. There have been a number of efforts to add semantics to the discovery process of web services. An upper ontology for services has already been presented to the DAML-s Project [3].

As previously mentioned knowledge services leverage knowledge that is codified in knowledge objects. The representation of a knowledge object, and subsequently a knowledge service, is a complicated task and requires the generation of ontologies for its efficient description. Communication protocols or message formats employed by knowledge services are no different than typical web services.

Despite the fact that we could add specific attributes to the current description, in our implementation we prefer to elaborate the Knowledge Object Ontology and then embed it to the upper service ontology for reusability purposes. Figure 2 shows the relationship between SERVICEPROFILEs [3], knowledge services and knowledge objects. The extended SERVICEPROFILE is used to represent a knowledge service that retrieves and composites knowledge objects from a single or a variety of sources. According to this approach a knowledge service is described as a Semantic Web Service with a usual DAML-s SERVICEMODEL and SERVICEGROUNDING and a SERVICEPROFILE based on the elaborated ontology we have created.
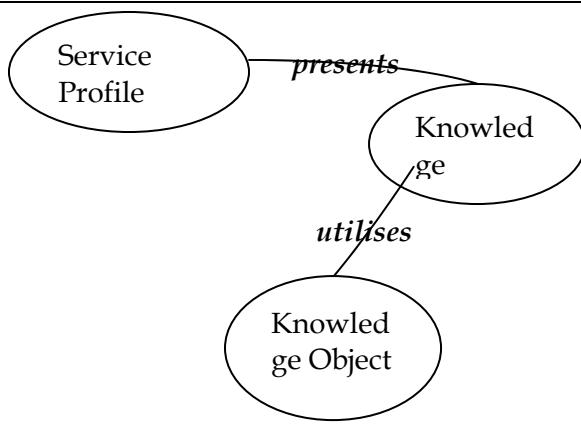
Figure 2: Extention of the daml ontology for services

An architecture for delivering knowledge services through web service interface

Figure 3 illustrates an architecture for delivering knowledge services through web services interface.
The main idea of this figure is to enlighten the utilization of daml-s+wsdl[4] in conjuction with the Knowledge Object Ontology. What we do is create a wsdl file where the messages to be exchanged contain the metadata of a knowledge object. In order to achieve this we use the daml-s ontology unparsed as java classes.

The knowledge objects are provided by different knowledge service providers, mentioned as KPi , i=1,2,… in figure 1. Reference and metadata of the knowledge provided by the KP are stored in the Metadata Database whose schema complies to the Knowledge Object ontology, presented in the previous section. The retrieval service is described by wsdl files that provide the end point to the SOAP interface.

The eleven facets identified in the Knowledge Object Ontology are expressed as relation classes in the daml knowledge representation model in any knowledge service discovery and delivery situation**.**

Figure 3:
Knowledge Web Services: Overall Approach

A user can perform queries in the database based on these relations. For example she can query for all knowledge objects which are related to a specific industry or business process. Taxonomic relations are used only in Industry types according to NACE and business processes based on experience from existing projects [7]. Validation of the input queries is performed through the use of Jena DAML Parser according to the Knowledge Object Ontology,. The whole application is hosted on an Oracle 9i Application Server. Main

components of the architecture, which arethe Knowledge Object ontology, the metadata database and the Web Service interfcace.

Through such queries we build a retrieval service for any content type with selected business processes or industries. Based on similarity measures, results ranking is available. Six knowledge Web Services are currently provided extending the retrieval mechanism described in the precedent paragraph.

Conclusions

Two are the future research milestones for Knowledge Web Services. The first is to implement our architecture when knowledge is exchanged in distributed environment. This can be succeeded following the peer-to-peer computing model. For this a P2P ontology based infrastructure is required [8]. The second is to address the issue of automatic composition of knowledge services [6]. Through the web service interface of knowledge services the already existing composition techniques for web services may be applied. Composition of knowledge services consists a promising research field for the area of knowledge management.

References

1. Abecker, A., Apostolou, D., Franz, J., Maass, W., Mentzas, G., Reuschling, C., Tabor, S.; *Towards an Information Ontology for Knowledge Asset Trading*, in Weber, Pawar, Thoben (Eds) Proceedings of the 9th International Conference on Concurrent Enterprising, June 2003, Espoo, Finland.

2. Abecker, A., et al.: Towards a Technology for Organizational Memories. IEEE Intelligent Systems, (1998), 13 (3), May/June.

3. DAML-S, Coalition: *Web Service Description for the Semantic Web*. ISWC2002. www.daml.org

4. DAML-S, Coalition: *Describing Web Services Using DAML-S and WSDL.* working document, May 2003, www.daml.org

5. Fahey, K. et al (2001) Linking e-business and Operating Processes: The Role of Knowledge Management, *IBM Systems Journal*, Vol. 40, No 4, 2001, pp. 889-907.

6. Georgolios etal, Towards knowledge oriented service composition. Implementing web services, submitted to ICSOC 2004: 2nd International Conference on Service Oriented Computing, Co-sponsored by ACM SIGWEB and ACM SIGSOFT, New York City, USA, November 14-18, 2004

7. INKASS (2000) The INKASS Project Web Site, www.incass.com (1 August 2002)

8. INKASS (2004) Agora deliverable, www.incass.com

9. Kocharekar, R. (2001) "K-Commerce: Knowledge-based Commerce Architecture with Convergence of e-commerce and Knowledge Management," *Information Systems Management*, Spring 2001.

10. Mentzas, G.; Apostolou, D.; Abecker, A., Young, R.; *Knowledge Asset Management.* Springer-Verlang London (2003)

11. Papazoglou, M, *Service -Oriented Computing: Concepts, Characteristics and Directions, Fourth International Conference on Web Information Systems Engineering (WISE'03)*, **IEEE**

12. Quinn, J.M., Baruch, J.J., Paquette, P.C.: Technology in Services. Scientific American, 257(6): 24–32, 1987.

13. Satyadas, A. and U. Harigopal (2001) "Cognizant E-Business Solutions: Linking the New E-Business Wave with Knowledge Management," IBM paper (2001).

Ongoing Research Column: Real World SW Cases

The Research in Progress or Ongoing Research Column is dedicated to the presentation of interesting research works with important achievements and critical milestones towards the realization of Information systems that prove the value potential of semantic web. In this issue we present one short article.

Measuring the Semantic Web

Rosa Gil, Roberto García, Jaime Delgado
Universitat Pompeu Fabra (UPF), Departament de Tecnologia,
Pg. Circumval·lació 8, E-08003 Barcelona, Spain
{rosa.gil,roberto.garcia,jaime.delgado}@upf.edu

Introduction

The book *Weaving the Web* by Tim Berners-Lee [i] presents a plan to build what is called The Web. Basically, it can be described as a decentralised knowledge system that self-organises and evolves scaling to unforeseen conditions. The Semantic Web effort is introduced as the last step towards completing this idea, from the result of the World Wide Web effort.

This knowledge system is not like previous ones. It is open and growths freely, without central control, and this can produce many undesired outcomes that can be also seen as opportunities. The World Wide Web is based on the same principles, e.g. there are link consistency problems, but has largely succeeded.

However, these are only vague words. What have we really built with the World Wide Web? And what are we building with the Semantic Web? How near we are from the original plans and what is the metric? These are difficult questions. The WWW and the Semantic Web are acquiring a size and a complexity that puts them out of our control and even from our direct conception.

What can we do? Just looking around we realise that the WWW and the Semantic Web are just as complex as many other systems. Other research communities have faced similar problems and found a common approach, which has properly been called the study of complex systems. Are they complex systems?

Complex systems (CSs) are made up of the combination of a great amount of elements. However, their behaviour is not the sum of the behaviour of their parts. Examples of CSs are metabolic networks [i] acquaintance networks [ii], food webs [ii] or neural networks [ii]. Actually, it has also been shown that the WWW is a CS [ii].

What do all this systems have in common? How to identify a CS? Scientists have looked for a way to achieve this using some mathematical tools, concretely graphs and statistical mechanics. In the next section this methods are presented.

Modelling and Analysing Complex Systems

Graphs are used to model CSs in order to analyse them. Nodes represent the CS parts (chemical components, people, species, neurons, web pages…). Edges model the relationships among the parts (chemical reactions, acquaintanceship, species dependences, neuron axons, web links…).

The resulting graphs show statistical properties that characterise CSs. Some of them are highlighted here. They are considered sufficient conditions for identifying a CS:

- **Degree distribution**: the resulting graphs, although they model systems that are shaped without a central control, are not random graphs as it was first believed. The probability $P(k)$ that a vertex has a degree k does not follow a Poisson distribution as in random graphs. Instead, it shows a power-law distribution, $P(k) \approx k^{-r}$. This kind of distributions are characterised by the γ exponent and are called scale-free networks [ii]. In other words, they show the same properties independently of the scale at which they are observed.

- **Small world**: a graph is a small world if the average minimum path length d between vertices is short [ii,iii], usually scaling logarithmically with the total number of vertices. Graphs showing an average path length similar to random graphs of the same size and average degree are very likely small worlds [ii], $d \approx d_{random}$.
- **Clustering coefficient**: It measures the probability that two neighbours of a given node are also neighbours of one another. For random graphs it is a small quantity. However, CSs show a high clustering compared to random graphs, $C >> C_{random}$. A high clustering confirms small-worldness.

Is the Semantic Web a Complex System?

We are now going to study the Semantic Web as a CS. It is modelled as a graph and then analysed using the statistical methods already presented. The results are analysed in order to check if it is a CS and to compare it with other ones. All the tools that have been used and the complete results are available at the project web page [ii].

The Semantic Web Graph

The first step towards analysing the Semantic Web is to build an appropriate graph model. Due to self-similarity and scale invariance of CSs, we can perform this analysis selecting a significant portion of the Semantic Web and the results can be inferred to other scales.
We have focused on the ontological part of the Semantic Web, i.e. we model the graph from a set of semantic web ontologies. We could also use instance metadata but we consider that at this first stages focusing on ontologies makes the conclusions more relevant.
Instance metadata usually models "real networks" that should be analysed on their own or have already been shown to be CSs. For instance, FOAF metadata models social networks that have been extensively studied as CSs [viii].
   Therefore, in order to collect the semantic web ontologies that are analysed, an RDF crawler is launched over the DAML Ontology Library [ii]. The processed URIs are combined in a RDF graph built from 160,000 triples.

Graph analysis

In order to analyse the obtained Semantic Web graph we use Pajek [ii], a large networks analysis tool. The RDF triples are translated to Pajek network format. The triples subjects and objects became network nodes connected by directed edges from subject to object.
For this first analysis we will focus on the explicit nature of the Semantic Web. Only triples explicitly stated in the processed Ontologies are considered. Therefore, for the moment, the potential triples that could be inferred applying RDF, DAML+OIL or OWL semantics are ignored.
The Pajek network has 56,592 nodes and 131,130 arcs. Once loaded in Pajek, the available tools are used to obtain the required information about the graph: average degree, average path length, clustering factor and degree distribution.

Results

The results of the graph analysis are shown in Table 1. The first line, DAMLOntos, shows the results for the graph built from the ontologies at DAML library. It can be compared with the same parameters for other CSs networks: the results from some WWW studies [ii,vi], WordNet [ii] and human language words networks [ii].

Table 1. Some CS statistical properties. Networks, number of nodes, average degree <k>, clustering factor C, average path length <d> and power-law exponents γ

Network	Nodes	<k>	C	<d>	γ
DAMLOntos	56592	4.63	0.152	4.37	-1.48
WWW	~200 M		0.108	3.10	-2.24

WordNet	66025		0.060	7.40	-2.35
WordsNetwork	500000		0.687	2.63	-1.50

First of all, from the previous data, we can deduce that the Semantic Web is a small world comparing its average path legth <d>=4.37 to the corresponding value for a random graph with the same size and average degree, <d>$_{rand}$=7.23. Moreover, the clustering factor C=0.152 is much greater than C$_{rand}$=0.0000895 for the corresponding random graph.

The final evidence is the degree distribution; it is clearly a power-law. The degree Cumulative Distribution Function (CDF) for DAMLOntos is shown in Fig. 3. The linear regression of this function gives an exponent γ = -1.485 with a regression error $\varepsilon_\%$ = 1.455.
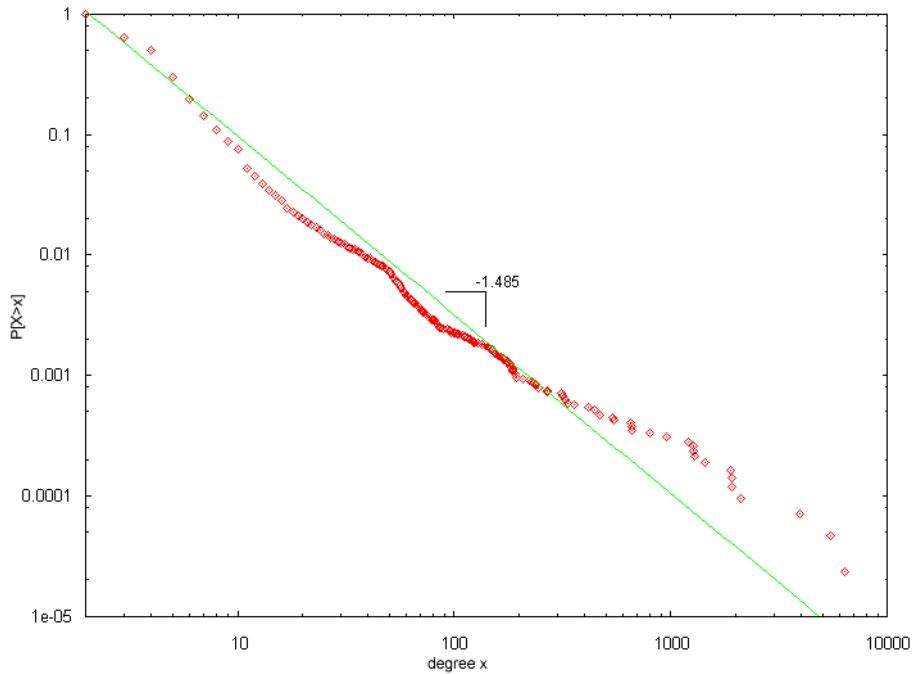

Fig. 3. Degree CDF (Cumulative Distribution Function) for the set of studied DAML library ontologies (DAMLOntos) plus linear regression and computed exponent

Therefore, the graph for the portion of the Semantic Web that has been analysed shows clear evidences that the Semantic Web behaves like a CS. It is a small world, with a high clustering factor and a power-law degree distribution. It has also a scale-free nature, so the same properties can be observed at a different scale.

Indeed, the analysis has been repeated for smaller graphs yielding the same conclusion. For instance, for a 971 nodes graph corresponding to the IPROnto [i] ontology: C = 0.071 while C$_{rand}$ = 0.0034272, <d> = 3.99 while <d>$_{rand}$ = 5.38 and γ = -1.06 with $\varepsilon_\%$ = 4.45.

Conclusions and future work

It has been shown that the Semantic Web behaves like a CS. When it is viewed as a graph, it reproduces all their characteristic patterns that. Once the Semantic Web is studied from this perspective, these patterns can be used as a kind of Semantic Web metric. With them, we can figure out its current situation and compare it to other CSs.

We have just started this work and a lot of questions have emerged. We plan to apply inferences to the retrieved triples in order to check the resulting graph. What do the implicit semantics do from the perspective of the whole RDF graph? Instance metadata is also going to be studied. Do the resulting graphs show the same statistical properties than the "real networks" that they model? And, what can we learn if we compare the Semantic Web with other "semantic" CSs like WordNet? It is sure that more questions are to come.

References

(i). Berners-Lee, T.: Weaving the Web. HarperBusiness (2000)

(ii). Wolf, Y., Karev, G. & Koonin, E.: Scale-free networks in biology: new insights into the fundamentals of evolution? Bioessays, 24, 105-109 (2002)

(iii). Amaral, L.A.N., Scala, A., Barthélémy, M., Stanley, H.E.: Classes of small-world networks. Proc. Natl. Acad. Sci, USA 97 (2000) 11149-11152

(iv). Montoya J.M., Solé, R.V.: Small World Patterns in Food Webs. Journal of Theoretical Biology, 405-412 (2002)

(v). Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Reviews of Modern Physics, 74 (2002) 47-97

(vi). Adamic, L.A.: The Small World Web. Proceedings of ECDL'99, LNCS 1696, Springer-Verlag (1999) 443-452

(vii). Barabási, A.L., Dezso, Z., Ravasz, E., Yook, S.H., Oltvai, Z.: Scale-free and hierarchical structures in complex networks. To appear in Sitges Proc. on Complex Networks (2002)

(viii). Pool, I., Kochen, M.: Contacts and influence. Social Networks 1 (1978) 1-48

(ix). Milgram, S.: The small world problem. Psychology Today 2 (1967) 60-67

(x). Solé, R.V., Ferrer, R., Montoya, J.M., Valverde, S.: Selection, tinkering and emergence in Complex Systems. Complexity, 8(1) (2002) 20-33

(xi). Living Semantic Web project web page, http://dmag.upf.es/livingsw

(xii). DAML Ontology Library web page, http://www.daml.org/ontologies

(xiii). Pajek, http://vlado.fmf.uni-lj.si/pub/networks/pajek

(xiv). Kleinberg, J., Lawrence, S.: The Structure of the Web. Science, Vol. 294 (2001) 1849–1850

(xv). Sigman, M., Cecchi, G.A.: Global organization of the Wordnet lexicon. Proc. Natl. Acad. Sci, vol. 99, no. 3 (2002) 1742-1747

(xvi). Ferrer, R., Solé, R. V.: The small world of human language. Proc The Royal Society 268 (2001) 2261-2265

(xvii). Delgado, J., Gallego, I., Garcia, R., Gil, R.: An ontology for Intellectual Property Rights: IPROnto. Extended poster abstract, International Semantic Web Conference (2002) http://dmag.upf.es/ontologies/ipronto

Special Section:
Semantic Web Leading Research Centers

In this Issue

DERI: Making Semantic Web Real

- Interview with Prof. Christopoh Bussler
- Presentation of DERI
- Projects' Description

Note from the editor:

I am really gratefull from the kindness of Prof. Bussler to prepare for us this special section. During his holidays we had an excellent collaboration and he spent a lot of time. Dear Chris many thanks to you.

Hope you will find this section very interesting. In the next issues of Bulletin we plan similar special sections for LSDIS Lab, MindSwap, AIFB, Protégé, ILRT, China Knowledge Grid, etc.

An Interview with Christoph Bussler
The DERI approach – Making the Semantic Web and Semantic Web Services Real

"The times are exciting and right to bring a vision and powerful idea to its full potential..."

Christoph Bussler

Science Foundation Ireland Professor
Executive Director
Digital Enterprise Research Institute (DERI)
National University of Ireland, Galway
Galway
Ireland
Web http://www.deri.ie

Short Bio: Christoph Bussler (http://hometown.aol.com/chbussler) is Science Foundation Ireland Professor at the National University of Ireland, Galway in Ireland and Executive Director of the Digital Enterprise Research Institute (DERI, http://www.deri.ie). In addition to his role as Executive Director of DERI, Chris leads the Semantic Web Services research group at DERI.

Miltiadis: Dear Professor Bussler we are really honored for your kindness to provide us this interview. I would like to start by asking you to share with us your opinion for the so-called digital world. Are things going well?

Christoph: First, thank you very much for your invitation to this interview. Just recently my family went on a short holiday and in preparing that we were able to organize every aspect of the preparation electronically. Buying equipment, travel arrangements, etc. When you think back (not that long actually) and compare how this was done before then it is quite amazing that this is all possible. Especially when all upfront arrangements actually happen as planned during the trip. So from this customer-centric perspective things are going well. Similar observations could be made from a business-centric perspective. When you look behind the scenes, however, and study what information systems infrastructure has to be put in place and maintained in order to provide that level of services, be it for customers or businesses, the state of affairs can be improved quite a bit from a technological side, especially the semantics side of it. Too many glitches happen due to missing semantic underpinning. Once semantics-based technologies are available, the situation for customers and businesses can be advanced a lot beyond the current state, too, in addition to overcoming today's problems. We in DERI address many of the technological issues that need to be overcome through research in the Semantic Web and Semantic Web Services.

> **Once semantics-based technologies are available, the situation for customers and businesses can be advanced a lot beyond the current state, too, in addition to overcoming today's problems.**

Miltiadis: Dear Chris, could you give us upfront a brief introduction to DERI?

Chris: DERI stands for Digital Enterprise Research Institute. DERI has two locations, one in Galway, Ireland, at the National University of Ireland and one in Innsbruck, Austria, at the Leopold-Franzens-University. You'll find the two sites when going to www.deri.org. DERI's mission is to make the Semantic Web and Semantic Web Services real through academic research and bringing research results successfully into industrial exploitable settings. In a nutshell, accomplished basic and applied research results from DERI researchers are transferred into domain specific application

areas for solving real problems. Examples are Knowledge Management, Enterprise Application Integration, eCommerce, the travel and financial industries. We have major national sponsors in Hewlett-Packard Galway, the Science Foundation Ireland, Cooperate, FFF and The Tiroler Zukunftsstiftung, as well as European funding from the IST program of the European Commission. Professor Dieter Fensel and I are leading the institute that has currently over 80 members with the goal to grow to about 150.

> **"What makes organizations successful are people, people, and people. In research environments it is precisely the same: researchers, researchers, and researchers. From a management perspective it is the highest priority to establish a world-class, high-powered and stimulating research environment where researchers can excel, grow and achieve outstanding research results".**

Miltiadis: Dear Christoph, you are the Executive Director of DERI, one of the top Research Institutes worldwide. How difficult is to manage such a research institute pursuing high research outcomes?

Christoph: What makes organizations successful are people, people, and people. In research environments it is precisely the same: researchers, researchers, and researchers. From a management perspective it is the highest priority to establish a world-class, high-powered and stimulating research environment where researchers can excel, grow and achieve outstanding research results. Researchers are involved in all aspects of DERI, which gives them a fantastic opportunity of personal growth in learning all aspects of a research institute, and it motivates them being part of it. This includes the whole range from accomplishing high-quality research results, to project acquisition and execution, and to contributing to applied industrial projects. A group of well-established and respected senior researchers leads the institute. With Professor Dieter Fensel, the Scientific Director of the Institute, DERI has an invaluable asset for establishing a world-leading research environment. All other aspects of managing an institute derive from the research environment requirements. We have an outstanding administrative backbone with project managers, outreach officers, business development and administrative support, all supporting the research environment in order to maximize research time and effectiveness.

Miltiadis: I have several times visited your web site at DERI (www.deri.ie) and I am really impressed from the huge amount of research outcomes related to projects deliverables, publications as well as events. What do we have to expect from your research in DERI in the near future?

Christoph: If you go to www.deri.at you will find many more already accomplished outcomes from DERI Innsbruck in addition to those from DERI Galway. You will see several different future developments. One is that research results will transfer into industry through joint industrial projects as well as spin-offs. DERI continues to participate in standardization activities in the Semantic Web and Semantic Web Services. In addition, DERI is still in a growth phase. We will increase the number of researchers as well as the number of research and industrial projects. This means that additional research areas as well as application domains are opening up for DERI. Just recently we acquired two additional projects, M3PE (Multi Meta Model Process Execution) and ASG (Adaptive Services Grid). M3PE is a basic research project in the area of process execution whereby ASG is an EU-funded project bringing together Grid, Semantic Web and Semantic Web Services technology.

> **" People need to be made aware of the development in general, they need to be educated about the benefits and they need to be enabled to use the technological results. EU projects provide this opportunity. And to increase the impact, several projects joined in these so-called dissemination activities and established a SDK cluster".**

Miltiadis: It is more than interesting that in your on-going projects in DERI, you are participating as leaders in the majority of EU-funded projects concerning semantic web (DIP, Knowledge Web, SWWS, SEKT). Can you outline their contribution to the realization of the SW for the vast majority of people? Are they parts of a critical puzzle towards SW services?

Christoph: EU projects are multi-faceted in their impact. First, industrial partners and academic partners work together in accomplishing the research results. This means that there is a continuous mutually beneficial exchange of ideas and real requirements between the two groups leading to significant research results. Furthermore, real impact is to be shown in terms of software as well as general community outreach. This means that the research results become accessible to industry as well as to society. All these aspects are

essential to make the Semantic Web as well as Semantic Web Services a reality. People need to be made aware of the development in general, they need to be educated about the benefits and they need to be enabled to use the technological results. EU projects provide this opportunity. And to increase the impact, several projects joined in these so-called dissemination activities and established a SDK cluster (see http://www.sdk-cluster.org/). This is a focus point as well as multiplier for the outreach aspect. DERI is involved in all these aspects further increasing the impact throughout society and industry.

Miltiadis: Dear Christoph if I asked you to select one critical milestone for the evolution of the Semantic or Knowledge Web what would you choose? Several critiques on the semantic web state that SW is just a fashion or a bubble... What you think?

Christoph: In terms of milestones we can look in three directions, into the past, the future and the current situation. Along this time axis we can observe many important milestones. Some are the establishment of research groups and funded research projects in the space quite a while ago, the successful cooperation of several research groups creating the OWL standard of W3C, the successful establishment of start-up companies in the space and, last but not least, the establishment of DERI as a major nationally, industrially and European funded research institute. All these events are important ones for the academic community, society, and industry. The Semantic Web is a long-term effort working towards a clear goal, not at all changing every year. Solid progress based on real impact creates a successful area, solid as well as healthy growth, and never a bubble. All involved parties, DERI, research groups, industry, standards organizations and customers are interested in making the Semantic Web a reality, not a fashion or a bubble at all. Milestones going forward will be industry-wide pick-up of Semantic Web and Semantic Web Services technology, broad application in all industrial domains and

> *"The Semantic Web is a long-term effort working towards a clear goal, not at all changing every year. Solid progress based on real impact creates a successful area, solid as well as healthy growth, and never a bubble".*

an ongoing establishment of Semantic Web and Semantic Web Services research groups in universities and research institutes world-wide.

Miltiadis: Dear Chris in this issue of SIGSEMIS Bulletin we have a special section for DERI. Could you outline for our audience your mission and your personal vision for DERI??

Christoph: The mission of DERI is to make the Semantic Web and Semantic Web Services real. This is based on solid research results as well as their transfer into industry. Personally I'd like to make a real impact, scientifically and economically by making DERI and its researchers successful in accomplishing their mission.

> *DERI is part of this vision with the Semantic Web and Semantic Web Services as an important foundation and element of a knowledge based society. Through the ongoing transformation the industrial sector is very open for advancing and collaboration.*

Miltiadis: Based on your previous answer I would ask you to comment on your personal experience for the so-called Irish Miracle in Information Technology. Ireland in Europe is considered as an excellent paradigm on how IT can provide enormous benefits for the society. What is your experience in DERI and why DERI in Ireland?

Christoph: Ireland, an important manufacturing region with access to the European market, realized quite a while ago that manufacturing will become mobile with the global change going on in the world. This insight created the vision for a knowledge based society moving up the value chain of Ireland as a country. IT is part of a knowledge based society with excellent opportunities for research, education and industry. Ireland took the opportunity for a transformation. Science Foundation Ireland was created with the goal to establish world-class research in Ireland to contribute to the transformation to a knowledge based society. Ireland will become

the place for successful research directly contributing to the knowledge based society. DERI is part of this vision with the Semantic Web and Semantic Web Services as an important foundation and element of a knowledge based society. Through the ongoing transformation the industrial sector is very open for advancing and collaboration. DERI cooperates with many companies in the space giving them a significant competitive edge.

Miltiadis: Dear Chris, your career path is just amazing. I would like you to comment on your European and USA experience. Does Europe fall behind USA in research? What you gained from this mix?

Christoph: I hear the question about Europe versus US many times and it is fascinating that it continues to come up as a general question. I will only comment on the Semantic Web and Semantic Web Services research. If you look at the history of the accomplishment of the Semantic Web up to date then European research groups have been a key player and are becoming more and more the center of the Semantic Web research. This goes hand in hand with the level of funding national as well as European funding agencies provide for this space. The same applies to the Semantic Web Service research, no doubt. There is no sign whatsoever that European research falls behind US research. Working for significant amounts of times in different cultures has many benefits. One clearly starts to experience and understand cultural differences in society as well as professional life. It is interesting to observe how different real world problems are approached and solved. Having worked in large as well as small organizations, it is important to understand the gap between customer and provider perception of problems and solutions. Also, the awareness of timelines in terms of time to market of research results is an important experience. For DERI, an impact strategy from first day was an important topic based on all these experiences.

> *"If you look at the history of the accomplishment of the Semantic Web up to date then European research groups have been a key player and are becoming more and more the center of the Semantic Web research. This goes hand in hand with the level of funding national as well as European funding agencies provide for this space".*

Miltiadis: A lot of people know you from your presence in ORACLE. What is your opinion for the perception of these big Corporations towards the SW? In my opinion several times technology acceptance is happening through push or pull strategies. Do you believe that these organizations will force push strategies?

Christoph: Every one of the big software vendors has a different strategy and approach. Some push, some wait for the pull, some work in the background, and some launch huge marketing efforts. The strategy and approach also depends on the technology itself, if it is a disruptive technology, a replacement technology, an incremental technology, etc. In the relation to the Semantic Web and Semantic Web Services, all understand the value and the benefit. At the same time big software vendors have a so-called Ecosystem around them consisting of customers, suppliers, standards organizations, research groups, user groups, etc. Any new development has to allow the Ecosystem to be part of it. However, one of the biggest forces is customers. If customers start consistently asking for a specific functionality, companies will react and satisfy the customer's needs. In the Semantic Web and Semantic Web Services domain this is currently happening more then ever. I am convinced that Semantic Web and Semantic Web Service technologies will be the foundation of software products across all application domains. That's why it is essential that DERI not only engages in basic research but also in applying the research results in various industry sectors.

> *I am convinced that Semantic Web and Semantic Web Service technologies will be the foundation of software products across all application domains. That's why it is essential that DERI not only engages in basic research but also in applying the research results in various industry sectors.*

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

Christoph: You are correct; the term Knowledge Society has many interpretations. For me it means that the society is economically based mainly on knowledge. The productivity of a society is achieved through knowledge. Knowledge allows creating value that is economically exploitable. This requires that Knowledge can be created. This in turn requires an excellent educational as well as research environment to train people appropriately and to provide them a creative research environment. DERI is part of this environment. It also means that knowledge is turned into economically valuable products, e.g., software and information. This requires an extremely well functioning knowledge and technology transfer into the knowledge industry. DERI contributes to this though the industrial activities. As all this is currently ongoing, it is clear that this is not utopia at all. The contrary is the case, the transformation is happening.

Miltiadis: Dear Chris, it is obvious that SW research is in the middle of supporters and people who criticize it. I would like to copy to you a small portion of an opinion of an interesting thinker John Sowa: "In 6 years (1998 to 2004) with ENORMOUS hype and funding, the semantic web has evolved from Tim BL's book to a

few prototype applications, which are less advanced than technologies of the 1970s such as SQL, Prolog, and expert systems -- and they're doing it with XML, which is far less advanced than LISP, which was developed in the 1950s. This contrast does not give me a warm, hopeful feeling about the semantic web…" What would you answer to such a critique for the importance and performance of SW?

Christoph: The Semantic Web is not just a matter of research by itself and neither a matter of technology only. The need for the Semantic Web is established by the current rather cumbersome state of the technology and customers as well as businesses simply asking for its capabilities. The uptake of Semantic Web and Semantic Web Services technology by industry is significant, with many companies, small and big, established and start-ups, customers and software providers applying Semantic Web technology and building products with Semantic Web technology. Funding agencies fund Semantic Web research as well as the technology transfer into industrial settings. Standards have been established with many more to come in this space. All this constitutes already a significant successful effort and is a success that will continue given the clear and enormous potential and benefit. DERI contributes and takes part in all these aspects and will continue to be part of the success of the Semantic Web and Semantic Web Services.

> *"The uptake of Semantic Web and Semantic Web Services technology by industry is significant, with many companies, small and big, established and start-ups, customers and software providers applying Semantic Web technology and building products with Semantic Web technology".*

Miltiadis: You have organized several workshops, tracks and special issues on SW related themes. Do you plan anything for the near future?

Christoph: Yes, in the near future the VLDB workshop on the Semantic Web and Databases will take place in Canada in conjunction with the Conference on Very Large Databases, the AIMSA conference will take place in Bulgaria, a workshop on the implementation aspects of the Web Service Modeling Framework in Germany is being organized, as well as other events like an ECAI, FOAF and ISCW workshops. Our web pages contain always a current list of activities. Also, next year the 2nd European Semantic Web Conference (Greece) as well as the International Semantic Web Conference 2005 (Ireland) will take place. These events are important to continue to build the academic and industrial community around the Semantic Web and Semantic Web Services, which is an important goal of DERI.

> **Semantic Web Services address this situation by providing a mechanism to formally describe the meaning of Web Services. This then allows not only to integrate the services but also to ensure that their integration is semantically correct.**

Miltiadis: A major research stream in DERI is the Semantic Web Services area. Would you like to share your vision for Semantic Web Services towards more effective and efficient services?

Christoph: Without addressing the semantic aspect Web Services will never be effective and efficient. Semantics-free web services have already proven to not solve the interoperability problem any better then known technologies that are available for quite some time. The majority of the effort in making businesses (B2B integration) or applications (EAI) interoperable lies in ensuring the semantically correct cooperation. In a semantics-free environment this means that defining the integration has to be achieved manually and based on human user experience. That is very time consuming, costly, and there is no guarantee that the result of this human design activity is correct. Semantic Web Services address this situation by providing a mechanism to formally describe the meaning of Web Services. This then allows not only to integrate the services but also to ensure that their integration is semantically correct. The Web Service Modeling Ontology (WSMO) at www.wsmo.org is such formalism.

Miltiadis: A few days ago I had a conversation with a colleague in the university. He wanted to learn more about 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?

Christoph: The key to success lies in both, technologies and theories. Furthermore, their application to real world problems. This is not an either – or question at all since both are essential to make the Semantic Web and Semantic Web Services real. My advice would be for a newcomer to be engaged in all three areas: technologies, theories and real

applications of both. Only when you are concerned about all these three areas, the development of theories and technologies will lead to useful ones due to their application to real world problems.

Miltiadis: Chris undoubtedly your role in DERI incorporates vision, inspired management, effective HR management, Leading Edge IT management, etc. How do you manage such a complicated role?

Christoph: There are several aspects when leading a research institute like DERI. First and foremost, we are recruiting top people, be it researchers, managers or administrative staff. Experience is very important as is personality and team spirit. Team spirit is another big factor that enables a larger research group to be successful. Everybody contributes and everybody can rely on everybody else. And a shared vision and a shared goal provide the common direction necessary to focus on a successful way forward

"Team spirit is another big factor that enables a larger research group to be successful. Everybody contributes and everybody can rely on everybody else"

Miltiadis: Several times I am contacting people from research institutes worldwide. A general conclusion is that they all share a very optimistic vision for the role of new technologies and also a list of problems. What are the major problems that you see in the promotion of the Digital World?

Christoph: Adoption of new technology always takes time. It is slow initially and only the curious will pick it up. But then the adoption growth increases dramatically until it is in common use. As all technologies before, there are problems to overcome along the way, but I don't see any roadblocks at all. We are well under way, have a great start under our belt and I would agree: the future looks bright.

Miltiadis: How do you find the formation of the New Special Interest Group on Semantic Web and Information Systems on AIS?

Christoph: I think that this is a very important step for the Semantic Web and Semantic Web Services community. First, it is an important event and especially when the leaders of the field are immediately involved. Second, the bulletin is an important channel of research, news as well as discussions, all necessary elements for the continued growth of the community. I definitely expect this interest group to grow and to play a major role.

The times are exciting and right to bring a vision and powerful idea to its full potential. I encourage everybody in the community to be part of this fantastic opportunity

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

Christoph: Thank you very much for the opportunity of talking to you and the readers of the bulletin. I am looking forward to future professional interactions to make the Semantic Web real. The times are exciting and right to bring a vision and powerful idea to its full potential. I encourage everybody in the community to be part of this fantastic opportunity

Christoph Bussler (http://hometown.aol.com/chbussler) is Science Foundation Ireland Professor at the National University of Ireland, Galway in Ireland and Executive Director of the Digital Enterprise Research Institute (DERI, http://www.deri.ie). In addition to his role as Executive Director of DERI, Chris leads the Semantic Web Services research group at DERI.

Before taking this position he was Member of Oracle's Integration Platform Architecture Group based in Redwood Shores, CA, USA. He was responsible for the architecture of Oracle's next generation integration product providing EAI, B2B and ASP integration. Prior to joining Oracle he was at Jamcracker, Cupertino, CA, USA, responsible for defining Jamcracker's ASP aggregation architecture, Netfish Technologies (acquired by IONA), Santa Clara, CA, USA, responsible for Netfish's B2B integration server, The Boeing Company, Seattle, WA, USA, leading Boeing's workflow research and Digital Equipment Corporation (acquired by Compaq, acquired by Hewlett-Packard), Mountain View, CA, USA, defining the policy resolution component of Digital's workflow product.

He has a Ph.D. in computer science from the University of Erlangen, Germany and a Master in computer science from the Technical University of Munich, Germany. Chris published a new book titled 'B2B Integration', two books in workflow management, over 60 research papers in journals and academic conferences, gave tutorials on several topics including B2B integration and workflow management and was keynote speaker at many conferences and workshops.

Digital Enterprise Research Institute (DERI)

The Digital Enterprise Research Institute (DERI, http://www.deri.org) is the world-leading institute in Semantic Web and Semantic Web Service research.

The major objective of the Digital Enterprise Research Institute (DERI) is to bring current Web technology to its full potential by combining and improving recent trends around the Web. DERI is a cooperation of Hewlett Packard Galway, the Leopold-Franzens Universität Innsbruck, and the National University of Ireland, Galway. In short, the mission is to make the Semantic Web and Semantic Web Services a reality.

DERI has two locations, DERI Galway, Ireland (http://www.deri.ie) and DERI Innsbruck, Austria (http://www.deri.at). DERI is led by Professor Dieter Fensel (http://www.fensel.com) and Professor Christoph Bussler (http://hometown.aol.com/chbussler). Currently DERI has over 80 members and will grow to over 150 members when reaching a fully staffed research environment.

The academic and industrial research work is carried out in context of international, European, national as well as industrial projects. The projects range from basic research projects with exploratory nature to very applied projects that immediately result in industrial applications. The projects are alphabetically listed and outlined below in more detail with the corresponding URLs that point to more detailed information.

In addition to establishing research contributions DERI is involved in educational outreach as well as standardization activities. The Semantic Web Services Initiative (SWSI, http://swsi.semanticweb.org/) is co-chaired by Professor Dieter Fensel and SWSI's architecture sub-committee is co-chaired by Professor Christoph Bussler. DERI is member of W3C (http://www.w3c.org) as well as OASIS (http://www.oasis-open.org).

DERI is very involved in the research community. DERI's researcher organize conferences like the ISWC 2005 as well as the ESWS 2004, workshops, lead major research initiatives and are members of many conference program committees. A complete overview can be found at the institute's web sites as well as the individual researcher home pages. In addition, several DERI researchers are part of the Semantic Web Science Association (SWSA, http://www.iswsa.org/) management board and members.

DERI is supported by many industrial partners and funding agencies. The major partners are Hewlett-Packard, Galway (http://h40055.www4.hp.com/galway/), Science Foundation Ireland (http://www.sfi.ie), Information Society Technologies (IST, http://www.cordis.lu/ist/), Tiroler Zukunftsstiftung (http://www.zukunftsstiftung.at/), Forschungsförderungsfonds für die gewerbliche Wirtschaft (FFF, http://www.fff.co.at/) and Cooperate (http://www.cooperate.at/).

For contact information of DERI researchers and open research positions please go to http://www.deri.org, http://www.deri.ie and http://www.deri.at.

Selected Projects

Adaptive Services Grid (ASG)

The goal of the newly started Adaptive Services Grid (ASG) project funded by the European Commission is to develop a proof-of-concept prototype of an open development platform for adaptive services discovery, creation, composition, and enactment. To achieve its goal, ASG addresses scientific and technological issues making use of the knowledge and expertise of major European research institutions with significant contributions from the software, telecommunications, and telematics industry. ASG provides the integration of its sub-projects in the context of an open platform, including tool development by small and medium sized enterprises.

Based on semantic specifications of requested services by service customers, ASG discovers appropriate services, composes complex processes and – if required – generates software to create new application services on demand. Subsequently, application services will be provided through the underlying computational grid infrastructure based on adaptive process enactment technology. In the proposed project, methods and concepts from software architectures, software development methodologies, Web Services composition and workflow process planning and coordination will be complemented by recent results in domain engineering, software generation and Semantic Web and agent negotiation research.

Impact on a European level is supported by strong industry involvement – both with respect to platform development, deployment, and exploitation – in the areas of telecommunications and telematics. To this end, ASG makes a difference for the people in the society since it helps bridging the gap between member states and important candidate states in the areas of telecommunications, telematics, and enterprise IT. DERI's contribution to ASG is in the area of Semantic Web Services enrichment of the Grid infrastructure and matchmaking [1] using Semantic Web infrastructure.

Corporate Ontology Grid (COG)

The objectives of the COG Project are to:

- Demonstrate the applicability of Grid technologies to industry
- Realize the concept of an Information Grid incorporating real corporate data
- Give semantics to corporate data formats including legacy, relational and XML through a detailed mapping to a central ontology
- Demonstrate the technological innovation of automatically translating data between data formats on the Grid by way of a semantic mapping to a central ontology
- Publish a reusable ontology with concepts from discrete manufacturing and specifically automotive
- Capture best practices in achieving above objectives and document the methodology
- Disseminate the methodology and results to Europe's corporations
- Reuse of methodology in industry

The major work involved in the project includes

- Creating a library of ontologies for the automotive industry (which by extension will be reusable in other discrete manufacturing industries)

- Studying innovative ways in which ontological software can work on top of underlying Grid infrastructure
- Mapping real underlying structured data sources, as supplied by the end use, to the ontologies
- Rationalizing and providing a unified 'virtual view' of data source by way of the ontologies, for knowledge navigation purposes
- Generating transformation scripts in a fully automatic manner from the ontology and mappings thus turning the ontology into an active thesaurus
- Attaining proof of success by utilising the test data in practice
- Capturing the methodology for repeating this success on other corporate grids in industrial settings
- Disseminating the methodology and best practices of the project to a wide audience by way of seminars, workshops, papers and a book

More information can be found at http://www.cogproject.org/

Data, Information and Process Integration (DIP)

With current Web technology the computer is used as a device for rendering information for the human reader only. Providing actual support in information processing and information exchange requires machine-processable semantics of data and information. This is precisely the goal of the Semantic Web. Based on Ontologies, the computer will be enabled as a device for querying and managing semi-structured information. Recent complementary efforts try to lift the Web to a new level of service based on integrating it with computational aspects. Software programs can be accessed and executed via the Web based on the idea of Web Services, which can significantly increase the Web architecture's potential by providing a way of automated program communication, discovery of services, etc.

The major mission of the European Commission funded DIP project is to further develop Semantic Web and Web Services and especially to enable their combination. Web Services are the proper means to access semantically enriched data and semantic enrichment of Web Services is essential for their scalability and maturity. This new area is called Semantic Web Services. Semantic Web Service technology will allow structural and semantic definitions of documents providing completely new possibilities in knowledge management, Enterprise Application Integration, and eCommerce. Semantic Web Services will provide a new infrastructure for eWork and eCommerce, just as the telephone did a century ago, based on its ability to provide semantic processing of data, information, and processes. DIP will develop this technology and will focus on applications in eWork and eCommerce including sub topics such as Knowledge Management, Enterprise Application Integration and eGoverment.

DIP's mission is to make Semantic Web Services a reality, providing an infrastructure (i.e. an architecture and tools) that will revolutionize data and process integration in eWork, and eCommerce as the Web did it for human information access.

The main contribution of DERI in DIP revolves around ontology reasoning and querying, ontology management, service description and service mediation, technology watch and standardization. Furthermore DERI is involved in DIP management and in the dissemination process. Ontology reasoning and querying contribution is aiming to add relevant extensions to the current state-of-the-art in the field (ontology querying, representation and reasoning). The reasoning contribution is focusing on hybrid reasoning and reasoning over dynamics.

In regard to ontology management we are active trying to suggest practical approaches to deal with large-scale ontologies, making the handling of ontologies with several thousands of concepts possible; manage heterogeneous ontologies networks, even if they have conflicting or complementary definitions; and organize the process of change of ontologies, so dynamic ontologies can be used.

Concerning Service description DERI is active in developing the ontology and Semantic Web infrastructure for adding semantics to Web Service descriptions. This will allow

- The mapping of elementary and complex Web Service descriptions with goal descriptions
- The integration of heterogeneous business processes and interaction protocols by means of formalizing them in ontologies, which will significantly help to scale up ad-hoc formation of business coalitions
- The ability to smoothly align heterogeneous data structures of combined Web Services
- Goal ontology. Instead of describing goals, Web Service functionalities, and quality of services from scratch
- The reuse of terminologies in order to allow a simple plug-and-play process for establishing definitions of goals Web Service functionalities, and quality of services

Regarding Service mediation DERI is working towards providing a general mediation function layer that will enable the mediation of heterogeneous processes. Such a layer will require the interpretation of goals as well as workflow and flexible Web Service invocation.

Technology watch and standardisation is researching into the state-of-the-art main technology and application areas relevant for DIP, is monitoring them and recommending the most suitable for DIP usage. Also identifies the most relevant standardisation bodies for effective development and acceptance of DIP standardisation activities.

DIP is part of the SDK cluster (http://www.sdk-cluster.org/). The SDK's cluster goal is to coordinate the projects SEKT (see below), DIP and KnowledgeWeb (see below).

More information about DIP can be found at: http://dip.semanticweb.org/

DERI Lion
DERI Lion (Lion is the Irish term for 'Web') is a substantial national long-term project funded by Science Foundation Ireland (www.sfi.ie). It is the core project of DERI Galway. The mission of DERI Lion is to combine and harness the power of Web Services and the Semantic Web. The end result will be a scalable semantics-based Web Service environment where business partners can seamlessly share services across dynamic business communities.

DERI Lion has five major areas of research. Two basic areas are Semantic Web and Semantic Web Services. Both are introduced in more detail below. These basic research areas provide the foundation for application oriented work in the areas of Knowledge Management (http://lion.deri.ie/clusters/km/), Enterprise Application Integration (EAI) and eCommerce (http://lion.deri.ie/clusters/eai/). These application oriented research clusters are applying the results of the Semantic Web and Semantic Web Services in their respective application domains with the goal of providing superior functionality.

More information about DERI Lion can be found at: http://lion.deri.ie/

DERI-Lion: Semantic Web Cluster (SW)
The current World Wide Web (WWW) is, by its function, a syntactic Web where the structure of the content has been presented while the content itself is inaccessible to computers. Although the WWW has resulted in a revolution in information exchange among computer applications, it still cannot provide interoperation among various applications without some pre-existing, human-created agreements outside the Web. The Semantic Web is aiming at more functionality, described by the following quote from [4]:

"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."

This quote from the Scientific American guides the work of the Semantic Web cluster within DERI-Lion. The quote mentions three different kind of collaboration – all of them are relevant for the Semantic Web cluster at DERI

- Computer and Computer. This is typically associated with the Semantic Web, since it is often quoted as making information machine readable and understandable. And indeed – simplifying the interoperability problem among computers is one of the goals within DERI. Developing technologies like RDF Inference Engines, able to deploy and use RDF data is one of the major goals within DERI. An example is the TRIPLE RDF Rule engine, able to reason and transform RDF metadata (see http://triple.semanticweb.org). However, computer-to-computer collaboration is just one aspect of the Semantic Web.

- People and Computer. Personal Data Management and Annotation is one challenges within the Semantic Web effort. DERI projects like SECO [2] and GECO, aiming at the aggregation and annotation of Web resources, simplify the deployment of Semantic Web data in collaborative environments.

- People and People. Social Networking is a recent trend within Web community. The Friend-of-a-Friend people representation is a standard which originated within the Semantic Web community (see http://www.foaf-project.org/). DERI is involved in helping to shape the future of Social Networking on the Semantic Web by organizing and coordinating the first Workshop on Social Networking on the Semantic Web [3].

More information about the Semantic Web cluster can be found at: http://lion.deri.ie/clusters/sw/

DERI-Lion: Semantic Web Services Cluster (SWS)

Today, Web Services and Semantic Web technology are separate from each other, each addressing a different problem space. Semantic Web technology is used in order to describe data semantically meaningful. Consequently, it is meaningfully accessible far beyond the current conventional access methods. Web Service technology is used for remote invocations in order to link programs and allow them to reuse each others functionality. It is possible to invoke remote functionality over the Internet with ease. However, Web Services must eventually be semantically defined and Semantic Web concepts must contain behavioral aspects. Therefore, instead of looking at Semantic Web and Web Services technology as different technology addressing different problem spaces, they can be looked at as complementary technology. Their combination called 'Semantic Web Services' results in a technology more powerful then the sum of both.

The goal of the Semantic Web Services cluster is to combine Semantic Web with Web Service technology, called 'Semantic Web Services'. A complete software stack is going to be implemented that consists of a conceptual model of Semantic Web Services, a language to define Semantic Web Services as well as a modeling and execution environment for the definition and execution of Semantic Web Services. The end result will be a scalable semantics-based Web Service environment where business partners can seamlessly share services across dynamic business communities. This automated Web Service Invocation Framework should enable flexible on-the-fly configuration of software, eWork, and eCommerce relationships.

The Semantic Web Services Execution Environment (Lion-WSMX) developed through DERI-Lion's Semantic Web Services cluster aims to provide a platform for matchmaking, selection, mediation, composition and invocation of the Semantic Web Services. The work carried out by both DERI-Lion SWS cluster and by WSMX working group hosted by DERI provides a reference implementation for Web Services Modeling Ontology (WSMO). The group aims to build an environment to provide both a test-bed for WSMO and to demonstrate the viability of using WSMO as means to achieve dynamic inter-operation of Web Services.

While the first version of Lion-WSMX provides a complete architecture for dynamic matchmaking, mediation, selection and invocation, the implementation of these components is going to be minimal. The initial functionality of the system focuses on achieving some user-specified goal (e.g. purchasing a novel) with some preference (e.g. from the most reliable Web Service) by invoking a Web Services and providing it with a mediated ontology fragment. The first version of Lion-WSMX is going to be released mid year 2004 as an open source effort. Subsequent versions of Lion-WSMX will incorporate the ongoing research of the Semantic Web

Services community, in particular the WSMO, WSML and WSMX working groups, augmenting and empowering Lion-WSMX to provide the full Semantic Web Services platform.

More information about the Semantic Web Services cluster can be found at: http://lion.deri.ie/clusters/sws/

Esperonto

The goal of the Esperonto project is to provide a bridge between the current Web and the Semantic Web. The current Web is based on HTML and meant for human users: HTML is about the layout of Web pages. The Semantic Web is based on new languages such as XML, RDF(S), OIL, DAML+OIL, etc., and is meant to be understandable by software programs.

Esperonto has two main objectives. First, the development of an annotation service provider that will help content providers bridge the gap between the current Web and the Semantic Web. For this objective, the necessary ontological infrastructure is being developed for proper coding and structuring of digital content. It includes support for multimedia and multilingualism. The second objective consists of providing added value knowledge-based services on top of the constructed Semantic Web by content aggregation using semantic indexes and routers; exploring innovative ways of visualizing Semantic Web content and ontologies. This objective is being illustrated by several pilot applications in different domains.

More information can be found at: http://esperonto.semanticweb.org/

h-TechSight

The main objective of the h-TechSight project is to improve the capabilities of technology intensive organisations to monitor, assess, predict and respond to technological trends and changes. The project aims at enabling SME's to become intelligent players in the global market. Technology from the project will provide SME's with cost-effective venues to achieve state-of-the-art awareness in research and development and extend their activities around the world. To achieve that, the project will develop technologies for systematically and automatically updating, dynamic knowledge maps of specific technology domains. These will act as advanced business intelligence support for market, products and technology watch and will represent a novel knowledge management practice in an area (technology evolution assessment), which is of tremendous importance to the high technology sector.

The major outcomes of the h-TechSight project are:
- Development of the Knowledge Management Platform KMP, which integrates
  - Evolving ontology management suite
  - Semantic Web search engine (agent technology)
  - Knowledge discovery tools (natural language processing, knowledge processing facilities)
- Domain ontologies for sample application areas
- Methodology for knowledge management in technology-intensive industries

More information about h-TechSight can be found at: http://prise-serv.cpe.surrey.ac.uk/techsight/

Infrawebs

INFRAWEBS will develop an ICT (Information and Computer Technology) framework, to enable software and service providers to generate and establish open, extensible and reconfigurable development platforms for Web Service applications. Such open platforms consist of coupled and linked INFRAWEBS units, whereby each unit provides tools and adaptable system components to analyze, design, conjointly compose, and maintain Semantic Web Services within their whole life cycle. Due to the versatile and modular structure, the proposed framework enables the handling of a broad class of services

- Providing a comfortable and "easy to use" knowledge brokering unit
- Allowing the design and composition of Semantic Web Services
- Preserving run-time and maintenance tools supplemented with secure and privacy

These objectives will be achieved by developing and realizing software and system components consisting of

- An ontology and tools for the extracting and retrieval of SW service specific knowledge and data
- A P2P, Multi-Agent and SW service based interoperability environment (Distributed SW service repositories) for the CBR based design and composition of SW services (experience based CASE- tool), and
- A run-time supervision (Executor, QoS-Broker) and maintenance module enriched with distributed decision making features considering security aspects

The role of DERI in the INFRAWEBS project revolves around

- The Analysis and Design of SWS registry, to enable basic storage and retrieval functionalities
- The specification of top level architecture, a service modeling framework, and local registry, to support the design, composition and reuse of Web Services
- The specification execution and QoS specific metrics and behavior patterns in designing / composing process, for QoS monitoring and service execution

More information about INFRAWEBS can be found at: http://infrawebs.aspasia-systems.de/

Knowledge Web (KW)

Knowledge Web is a FP6 Network of Excellence that aims to support the transition of Ontology technology from Academia to Industry. It started on January 1st 2004, with a budget of around 7 million euro and 18 participants including leading partners in Semantic Web, Multimedia, Human Language Technology, Workflow and Agents.

In a nutshell, it is the mission of Knowledge Web to strengthen the European industry and service providers in one of the most important areas of current computer technology: Semantic Web enabled e-work and e-commerce. The project efforts will be concentrated around the outreach of this technology to industry. This includes education and research efforts to ensure the durability of impact and support of industry.

Knowledge Web establishes:

- An Ontology Outreach Authority, being "the" meeting place for interacting with interested industrial parties to take advantage of the latest research results, including tools. In the end, Knowledge Web will strive to set up an alliance with several industry bodies in order to set up an Ontology Outreach Authority, certifying, and serving validated ontologies
- A Virtual Institute for Semantic Web Education (VISWE) where a specialized and adapted curriculum, which no single university can offer, is created for students coming from all over Europe
- The Virtual Research Centre will coordinate the research carried out within Knowledge Web and take care that its results are shared and disseminated

The current consortium will ensure that Knowledge Web is open for further academic members and research institutions that provide substantial contributions on any of the main goals of the network: outreach to industry, outreach to education, and research. Based on the importance of Knowledge Web's main research topics, the high number of participants in the OntoWeb thematic network (http://www.ontoweb.org, 143 participants) and the exponential growth rates in the area of ontologies and the Semantic Web, Knowledge Web expects to count with many additional academic and research members.

Knowledge Web will be also open for further industrial members that provide substantial contributions to the network goals. Best-practice cases, interesting applications, training courses, deploying ontology-based applications, learned lessons in the application of ontologies and Semantic Web technology, etc., are good examples of the contributions expected from these industrial members. Industrial partners play also the role of a window to the standardization efforts. DERI is especially aiming at versioning of Semantic Web data and Ontologies and Semantic Web Services.

KnowledgeWeb is part of the SDK cluster (http://www.sdk-cluster.org/)

Further and up to date information about the project can be found at: http://knowledgeweb.semanticweb.org/

Multi Meta-Model Process Execution (M3PE)
The Multi Meta-Model Process Execution project is the latest in the set of nationally founded projects of DERI. This project is a basic research project with the goal to provide an execution model as well as an implementation for process execution. However, unlike all existing approaches, the vision is to allow not only to define processes, but also their meta-model underneath that defines the process execution semantics.

M3PE's approach to support any number of process execution models is based on the observation that every existing process or workflow model has a fixed semantics [5] [6]. This is considered the normal case since every process or workflow model has to define what it means to instantiate a model, schedule processing steps and define the data flow.

However, in the world of interoperability many process models can be found. This includes a variety of workflow systems as well as process standards in the world of Business-to-business integration [5]. In order to make processes interoperable, interoperating entities have to adjust their process model to the one they want to interoperate with. This is easiest done by having available a process execution system that allows to define the process execution model dynamically for a given process. This enables the communicating entity to adjust to the process model of the partner that requires interoperation.

The project is a basic research project with a duration of three years.

OntoWeb
The goal of OntoWeb Network is to bring researcher and industrials together enabling the full power ontologies may have to improve information exchange in areas such as information retrieval, knowledge management, electronic commerce, and bioinformatics. It will also strengthen the European influence on standardization efforts in areas such as Web languages (RDF, XML), upper-layer ontologies, content standards such as catalogues in electronic commerce, and the use of language technology in ontology development and knowledge markup.

The main workpackages of OntoWeb are
- Technical Roadmap: A document will outline the current state of the art of the field in Europe and world wide, providing the most promising lines of research, development and application
- Guides to industrial and commercial Applications
- Ontology-based content standardization recommendations to promote the development of ontology-based metadata standards and content harmonization/interoperability across different standards for the creation, communication and sharing of capability, process and activity models on the Web

- Ontology language standardization recommendations because a joint agreed standard for specifying and exchanging ontologies is essential if ontologies are to fulfill their pivotal role in realizing the vision of a Semantic Web
- Establishing an information portal on language technology in ontology development and use as part of the OntoWeb portal, which will have a bridging function between the language technology and knowledgement management communities
- Co-operation with non-European initiatives like DARPA DAML and the IEEE Standard Upper Ontology Study Group
- The network will establish a Web portal covering all aspects of the knowledge Web
- Organisation of workshops and meetings
- Setting up a scientific journal
- Providing a bridge to the multi-media community and their work on content description
- Education support to manage, co-ordinate and initiate educational initiatives related to Web standardization activities and to topics related to the Semantic Web in general
- An annual report will be produced that is readable and accessible for a wide audience

The main results of OntoWeb are

- A technical roadmap on the state of the art of the WWW of the next generation plus guides to Industrial and Commercial Applications
- A series of European and international workshops that bring together leading researchers and industrials
- A Web portal on advanced knowledge management and electronic commerce. Contributions in content and language standardization, and in language engineering for knowledge management
- A scientific journal and educational material

More information about OntoWeb can be found at: http://ontoweb.aifb.uni-karlsruhe.de/

Semantically-Enabled Knowledge Technologies (SEKT)

The vision of SEKT is to develop and exploit the knowledge technologies which underlie Next Generation Knowledge Management. We envision knowledge workplaces where the boundaries between document management, content management, and knowledge management are broken down, and where knowledge management is an effortless part of day to day activities. Appropriate knowledge is automatically delivered to the right people at the right time at the right granularity via a range of user devices. Knowledge workers will be empowered to focus on their core roles and creativity; this is key to European competitiveness.

The SEKT strategy is built around the synergy of the complementary know-how of the key European centers of excellence in Ontology and Metadata Technology, Knowledge Discovery and Human Language Technology, besides major European ICT organizations. Specifically, SEKT will deliver software to: semi-automatically learn ontologies and extract metadata, and to maintain and evolve the ontologies and metadata over time; to provide knowledge access; besides middleware to effect integration of all the SEKT components. SEKT will also develop a methodology for using semantically-based knowledge management. The software components and the methodology will be evaluated and refined through three case studies, in the legal, media and telecoms industries.

SEKT will also undertake a program of dissemination. This will be aimed not just at specialist technical communities, but also at wider information technology management and general management communities. The object of this dissemination program will be not just the specific results of SEKT, but also the general applicability of the semantically-based approach to knowledge management. Some of this program will be in

concertation with other projects in the semantically-enabled knowledge systems strategic objective. There will also be standards activities jointly with these projects.

SEKT is part of the SDK cluster (http://www.sdk-cluster.org/)

More information about SEKT can be found at: http://sekt.semanticweb.org/

Semantic Web Fred

The mission of the Semantic Web Fred project is to develop an integrated platform for task delegation to cooperative agents along with an up to date technology for automated and dynamic Task-Service-Resolution. The objective of the project is to extend the FRED system, developed by Net Dynamics, with technologies emerging in the area of the Semantic Web and Semantic Web Services in order to provide advanced support for Semantic Web applications.

The objective of the Semantic Web Fred project is unique because it combines agent technology with Semantic Web Services technologies. As there is no comparable research or development effort existing that combines these technologies in a coherent manner, the project will have significant impact on the development of runtime environments for the Semantic Web.

The FRED system, developed by Net Dynamics, is an environment for agent-based applications that allows import of ontologies and integration of external Web Services. Moreover, the FRED system integrates technologies for mediation of possibly heterogeneous resources with agent-based techniques for automated execution of tasks, thereby combining the core technologies of the Semantic Web. Thus, the FRED system can serve as an integrated platform for Semantic Web applications. The current status of development of the FRED technology provides a promising architecture, but some technical components are very basic at this point of time.

The Semantic Web Fred project takes the FRED system and WSMO (see below) as a starting point for development. The outcome of the project will be

- A Framework for the SWF which identifies the building blocks and the functionality of the system, comprised of
    o Specification of the Task-Service-Resolution Technology
    o Specification of the Agent Cooperation Technology
    o Combination of the 2 former aspects into the SWF Framework
- Implementation of enhanced Task-Service-Resolution Technologies
    o Description Language for Goals and Services; Services cane be Plans, Processes, or Web Services
    o Enhanced Resolution Mechanisms for Meeting Creation and Service Discovery, as well as basic mediation facilities needed
- Implementation of enhanced Agent Cooperation Environment
- Showcasing SWF in Use Case Implementations
- Presentation of SWF and the project results in scientific and industrial boards and events

More information about Semantic Web Fred can be found at: http://www.deri.at/research/projects/swf/

Semantic Web Annotator (SWAN)

SWAN is about Large Scale Annotation of human language for the Semantic Web (SW) using Human Language Technology (HLT). The SW and Semantic Web Services rely on formal semantics in the shape of ontologies and related instance sets, or knowledge bases. HLT provides the missing link between language and formal data, and a glue to fix Web Services to their user constituency and facilitate enterprise integration.

SWAN is an experiment in scaling up automated metadata extraction for industrial strength Semantic Web applications development. The SWAN project leverages language technology to advance a number of existing DERI cluster elements covering both basic research and applications demonstrators. The Web revolution has been based largely on human language materials, and in making the shift to the next generation knowledge-based Web, human language will remain key.

More information about SWAN can be found at: http://www.deri.ie/projects/swan

Working Groups

DERI established several public working groups in the area of Semantic Web Portals and Semantic Web Services in conjunction with the SDK cluster (http://www.sdk-cluster.org/). These working groups are open to everybody to become member and are leading efforts with the goal to establish coherent leading results in this space that are becoming part of standardization efforts as well as joint projects. The working groups are

- **Semantic Web Portal Project**. The mission is to create a Semantic Web Portal, demonstrating the maturity of Semantic Web technology in a real application. The technology, which will be developed together with industrial partners, during the course of the project, will be applicable to community portals for different communities.

  The two pillars for the portal will be the usability of the portal, both for inexperienced and expert users, and the support of communities, enabling cooperation within and between networks of people. The community support in the portal will not only be support in the passive consumption of information, but also support in the active publishing and collaboration by community members, through consensual vocabularies. Within the portal, the aim is to bring together networks of people and to facilitate collaboration between them.

  The use case for showcasing the Semantic Web Portal technology will be a community portal for the Semantic Web community, at http://www.semanticweb.org. Aim is to bring together research groups, research projects, software developers and user communities in the Semantic Web area.

  The technological mission is to create a satisfying ontology management environment needed for Semantic Web enabled community portals and to make extensive usage of Semantic Web technologies for enhanced information processing facilities and to create means for the Semantic inter-operation between different communities and even different Semantic Web portals.

  More informationa about this working group can be found at:
  http://www.deri.at/research/projects/sw-portal/

- **Web Service Modeling Ontology (WSMO)**. The mission is to create a Web Service Modeling Ontology, for describing services and its automation process. A world-wide standard will be provided, which will be developed together with industrial partners, research groups, and aligned with many different research projects.

  The pillars of the project will be the Web Service Modeling Framework (WSMF), which will provide some basic concepts that will be further developed in the course of the project, and the current available initiatives that try to address similar problems, which drawbacks will be overcome.

  The use case for showcasing our Web Service Modeling Ontology will be based on an eTourism

application. We aim to develop an ontology that can be easily used by research groups, research projects, software developers and user communities in the Semantic Web area.

More information about this working group can be found at: http://www.wsmo.org/

- Web Service Modeling Language (WSML**).** The SDK WSML working group, part of the SDK WSMO working group aligns the research and development efforts in the areas of Semantic Web Services between the SEKT , DIP and Knowledge Web research projects.

  Members of this working group include key participants with expertise in Semantic Web-related research areas.

  It is the mission of the SDK WSML working group to, through alignment between key European research projects in the Semantic Web Service area, further the development of Semantic Web Services and works toward further standardization in the area of Semantic Web Service languages and to work toward a common architecture and platform for Semantic Web Services.

-

  Specifically, the working group aims developing a language call Web Service Modeling Language (WSML) that formalizes the Web Service Modeling Ontology (WSMO). Hereby, we have a twicefold mission: Developing a proper formalisation language for Semantic Web Services and providing a rule-based language for the Semantic Web.

  More information about this working group can be found at: http://www.wsmo.org/wsml

- **Web Service Execution Environment (WSMX)**. The SDK WSMX working group, part of the SDK Cluster aligns the research and development efforts in the areas of Semantic Web Services between the SEKT, DIP and Knowledge Web research projects.

  Members of this working group include key participants with expertise in Semantic Web-related research areas.

  It is the mission of the SDK WSMX working group to built up a reference implementation of an execution environment for WSMO. The goal is to provide both a testbed for WSMO and to demonstrate the viability of using WSMO as a means to achieve dynamic interoperability of Web Services. The development process for WSMX includes defining its conceptual model, defining the execution semantics for the environment, describing an architecture and software design and building a working implementation.

  More information about this working group can be found at: http://www.wsmo.org/wsmx

References
[1] H. Tangmunarunkit, S. Decker, C .Kesselman: Ontology-Based Resource Matching in the Grid - The Grid Meets the Semantic Web. International Semantic Web Conference 2003: 706-721
[2] A. Harth: SECO: Mediation Services for Semantic Web Data. IEEE Intelligent Systems. May/June (Vol. 19, No. 3). See: http://seco.semanticweb.org
[3] http://www.w3.org/2001/sw/Europe/events/foaf-galway/
[4] T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", Scientific American, May 2001
[5] C. Bussler: B2B Integration. Springer-Verlag, 2003
[6] S. Jablonski, C. Bussler: Workflow Management. Model, Architecture and Implementation. Thomson Publisher, 1995

REGULAR COLUMNS

Semantic Search Technology Technologies by Dr. Peter Alesso

H. Peter Alesso,
h.alesso@comcast.net
Computer Science Department,
Ohlone College, CA

BOOKS:

- "Building Semantic Web Services," A.K. Peters Ltd., 2004.
- "The Intelligent Wireless Web," Addison-Wesley, Dec. 2001.
- "e-Video: Producing Internet Video as Broadband Technologies Converge," Addison-Wesley, July 2000.

SOFTWARE PUBLICATIONS:

- "Wealth Insurance," Compton's NewMedia, Inc., 1989.
- "Engineering Design," VSL, 1994.
- "Semantic Web Author," A. K. Peters, Ltd., 2004.

Column Description

SCOPE
Articles and news covering explanations, examples, and advances in emerging semantic search applications including: semantic search technology, latent semantic indexing, ontology matching, semantic search agents and semantic data clustering. In addition, we will include current development, algorithms, inference applications and development software tools.

DESCRIPTION

Search engine's, such as, Google with its 300 million hits per day and over 4 billion indexed Web pages are a vital part of today's World Wide Web. The prevaling attitude of surfers on the Web is: When you have a question - fire up Google.

Current commercial search technologies has been based upon two approaches: human directed search and automated search. In general, human directed search engine technology utilizes a database of keyword concepts and references. A great deal of existing search engine technology uses keyword searches to rank pages, but this often leads to irrelevant and spurious results. Some specific types of human-directed search engines, such as Yahoo!, use topic hierarchies to help to narrow the search and make search results more relevant. These topic hierarchies are human created. Because of this, they are costly to produce and maintain in terms of time, and are subsequently not updated as often as the fully automated systems.

The automated form of Web search technology is based on the Web crawler, spider, robot (bot), or agent which follows HTTP links from site to site and accumulates information about Web pages. This agent-based search technology accumulated data automatically and is continuously updating information.

As Semantic technologies become more powerful, it is reasonable to ask for better search capabilities which can truly respond to detailed requests reducing the amount of irrelevant results. A semantic search engine seeks to find documents that have similar 'concepts' not just similar 'words'. However, most semantic-based search engines suffer performance problems from the scale of a very large semantic network. In order for the semantic search to be effective in finding responsive results, the network must contain a great deal of relevant information. At the same time, large network must process many paths to a solution.

In this column, we will explore semantic search applications including: semantic search technology, latent semantic indexing, ontology matching, semantic search agents and semantic data clustering. In addition, we will include current development, algorithms, inference applications and development software tools.

AUDIENCE
Web Service developers, Web site developers, Semantic Web specialists, and search technology researchers will all benefit from this exposition of semantic search technology supporting automatic Web services.

Column: Search Engines Explore Semantics (6/04)

Semantic search engines seek to find documents that have similar concepts not just similar words. As Semantic Web technologies become more powerful, it is reasonable to ask for better search capabilities which can truly respond to detailed requests. Using RDF and OWL tags will offer semantic opportunities for new search methods.

Background

In early 1994, Jerry Yang and David Filo of Stanford University started the heirarchical search engine, Yahoo!, in order to bring some order to the otherwise chaotic collection of documents on the Web. Some months later, Brian Pinkerton of the University of Washington, developed the search crawler WebCrawler. Also in 1994, Dr. Michael Maldin of Carnegie Melon University created Lycos.

In late 1995, Metacceawler, Excite, AltaVista, and later Intomi/Hotbot, AskJeeves and GoTo appeared. Yahoo, though actually a directory, was the leading search engine at that time, but then AltaVista was launched and began to gain popularity.

By late 1998, Stanford's Larry Page and Sergey Brin reinvented search ranking technology with their paper "The Anatomy of a Large-Scale Hypertextual Web Search Engine" and started what became the most successful search engine in the world, Google**.** The uncluttered interface, speed and relevancy of the search results were cornerstones in winning the tech-literate public.

Search engine optimization became more important as experts tried to boost the rankings of their commercial websites in order to attract more customers. In 2000, Yahoo and Google become partners, with Google handling over 100 million daily search requests. In 2001, AskJeeves aquired Teoma, and GoTo was renamed Overture.

Google with its 400 million hits per day, and over 4 billion indexed WWW pages, is undeniably the most capable search engine today. The prevaling attitude is: When you have a question - fire up Google.

As the use of the World Wide Web has become increasingly widespread, the business of commercial search has become a vital and lucrative part of the Web. Search engines have become commonplace tools for virtually every user of the internet, and company names such as Yahoo! and Google, have become commonly recognized. These commercial search technologies have been based upon two approaches: human directed search and automated search.

In general, human directed search engine technology utilizes a database of keyword concepts and references. A great deal of existing search engine technology uses keyword searches to rank pages, but this often leads to irrelevant and spurious results. In its simplest form, a content-based search engine will count the number of the query words (keywords) that occur in each of the pages that are contained in its index. The search engine will then rank the pages. More sophisticated approaches take into account the location of the keywords. For example keywords occurring in the title tags of the Web page are more important than those in the body.

Some specific types of human-directed search engines, such as Yahoo!, use topic hierarchies to help to narrow the search and make search results more relevant. These topic hierarchies are human created. Because of this, they are costly to produce and maintain in terms of time, and are subsequently not updated as often as the fully automated systems.

The automated form of Web search technology is based on the Web crawler, spider, robot (bot), or agent which follows HTTP links from site to site and accumulates information about Web pages. This agent-based search technology accumulated data automatically and is continuously updating information.

Semantic search engines seek to find documents that have similar concepts not just similar words. As Semantic Web technologies become more powerful, it is reasonable to ask for better search capabilities which can truly respond to detailed requests. Using RDF and OWL tags will offer semantic opportunities for new search methods. However, most semantic-based search engines suffer performance problems from the scale of a very large semantic network. In order for the semantic search to be effective in finding responsive results, the network must contain a great deal of relevant information. At the same time, a large network creates difficulties in processing the many possible paths to a solution.

Several major companies are seriously addressing the issue of semantic search. Microsoft's growth on the Web may depend on its ability to compete with search leader Google. As a result, Microsoft has launched a new search program called MSNBot, which scours the Web to build an index of HTML links and documents. The homegrown system, which performs agent/robot functions previously done by Inktomi may pose a significant threat to Google.

Google has increased their commitment to content-targeted advertising with products that are based on semantic technology, which understands, organizes, and extracts knowledge from websites and information repositories in a way that mimics human thought and enables more effective information retrieval. A key application of the technology is the representation of the key themes on Web pages to deliver highly relevant and targeted advertisements.

The business of commercial search has become very profitable, not only for Google and rival Overture Services, but also for their partners MSN, America Online and Yahoo. Google and Overture share revenues with their distribution partners every time someone clicks on a sponsored link. With an estimated 500 million online searches taking place daily, the targeted ad business is predicted to generate more than $7 billion annually within four years, according to analysts.

Search Engine Categories
Search can be catorized into several types, lexical, linguistic, mathematical, metasearch, semantic, SQL query, and XMLquery, as follows:

- Lexical: searches for a word or a set of words, with Boolean operators (AND, OR, EXCEPT).
- Linguistic: analysis allows words to be found in whatever form they take, and enables the search to be extended to synonyms.
- Mathematical: semantic search operates in parallel with a statistical model adapted to it.
- Metasearch engines do not crawl the Web compiling their own searchable databases. Instead, they search the databases of multiple sets of individual search engines simultaneously, from a single site and using the same interface. Metasearchers provide a quick way of finding out which engines are retrieving the best results for you in your search. http://www.ixquick.com/ - http://www.profusion.com/ - http://vivisimo.com/
- Semantic: the search can be carried out on the basis of the conceptual meaning of the query.
- SQL structure query: a search through a sub-set of the documents of the database defined by SQL.
- XML structured query: the initial structuring of a document is preserved and the request is formulated in XPath.

There are two basic forms of Web search engine technology: (1) Small-scale human-based search engines that use a category hierarchy for each category described by a set of keywords, and (2) Large-scale agent/robot-based search engines which rely on bots to retrieve Web pages and store them in a centralized database. Both forms of Web search engines are based on keywords, and are subject to the two well-known linguistic phenomena that strongly degrade a query's precision and recall:

- Polysemy (one word might have several meanings) and
- Synonymy (several words or phrases, might designate the same concept).

There are several basic characteristics, we require for a search engine and how to improve them. It is important to consider useful searches as distinct from fruitless ones. To be truly useful, there are generally three necessary criteria:

- Maximum relevant information.
- Minimum irrelevant information
- Meaningful ranking, with the most relevant results first.

The first of these criteria - getting all of the relevant information available - is called recall. Without good recall, we have no guarantee that valid, interesting results won't be left out of our result set. We want the rate of false negatives - relevant results that we never see - to be as low as possible.

The second criterion - minimizing irrelevant information so that the proportion of relevant documents in our result set is very high - is called precision. With too little precision, our useful results get diluted by irrelevancies, and we are left with the task of sifting through a large set of documents to find what we want. High precision means the lowest possible rate of false positives.

There is an inevitable tradeoff between precision and recall. Search results generally lie on a continuum of relevancy, so there is no distinct place where relevant results stop and extraneous ones begin. The wider we cast our net, the less precise our result set becomes.

This is why the third criterion, ranking, is so important. Ranking has to do with whether the result set is ordered in a way that matches our intuitive understanding of what is more and what is less relevant. Of course the concept of 'relevance' depends heavily on our own immediate needs, our interests, and the context of our search. In an ideal world, search engines would learn our individual preferences so well that they could fine-tune any search we made based on our past expressed interests and peccadilloes. In the real world, a useful ranking is anything that does a reasonable job distinguishing between strong and weak results.

Google Search

The heart of Google Search software is PageRank, a system for ranking Web pages developed by the founders Larry Page and Sergey Brin at Stanford University

PageRank relies on the uniquely democratic nature of the Web by using its vast link structure as an indicator of an individual page's value. Essentially, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

Important sites receive a higher PageRank, which Google remembers. Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to the search. Google goes far beyond the number of times a term appears on a page and examines all aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for the query. The following table presents the PageRank calculation.

Table 1 - Google's PageRank Algorithm

== The PageRank algorithm is calculated as follows:

$$PR(A) = (1-d) + d\ (PR(T_1)/C(T_1) + ... + PR(T_n)/C(T_n))$$

where:

PR(A) is the PageRank of a page A
PR(T1) is the PageRank of a page T1
C(T1) is the number of outgoing links from the page T1
d is a damping factor in the range $0 < d < 1$, usually set to 0.85

The PageRank of a Web page is therefore calculated as a sum of the PageRanks of all pages linking to it (its incoming links), divided by the number of links on each of those pages (its outgoing links).
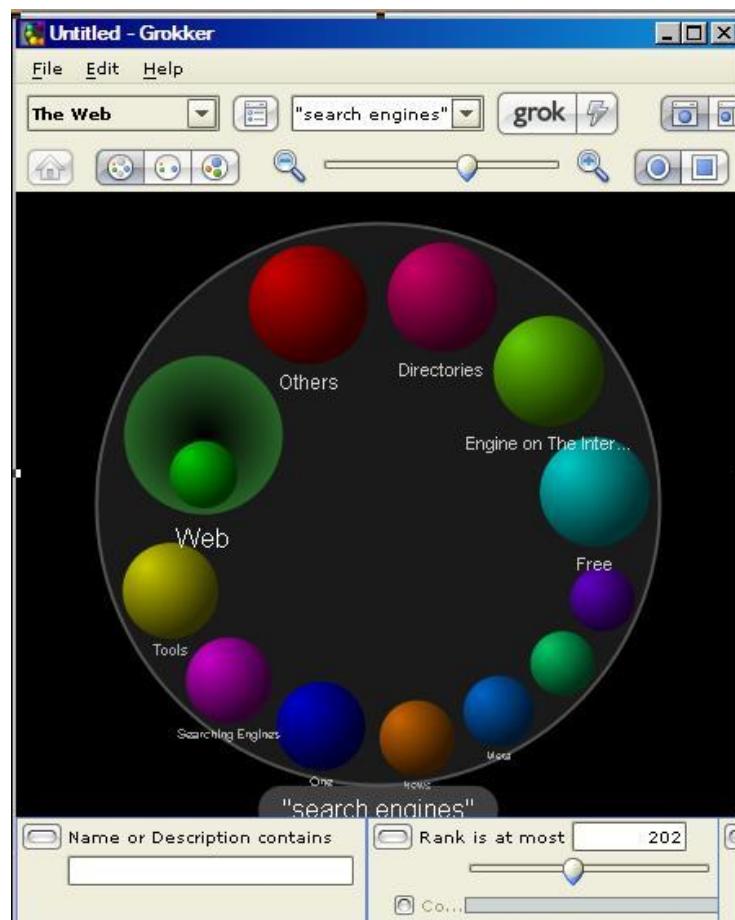
A Stanford University start-up, Kaltix, which use a Semantics-based technology, was recently acquired by Google after it had taken Google's model one step further, so that different search results are produced for every user based on their preferences and history. Without discussing Kaltix's plans publicly, the company's founders have published research that claims to offer a way to compute search results nearly 1,000 times faster than what's currently possible. Kaltix's method is similar to looking for a tree in a forest by examining only a clump of trees rather than the whole forest.

Grokker

In 2004, release of a new Web search product called Grokker has changed the search environment by using data clustering and improved graphics. Grokker is not a web service but an application that sits on your PC. Grokker takes the raw output of a search and organizes it into categories and subcategories. This means that a wide variety of types of databases can be Grokked. Currently, Grokker can search with six different engines simultaneously—Yahoo, MSN, Alta Vista, Fast, Teoma, and WiseNet.

Grokker creates a visual representation of the search. When you type in, say, "nanotechnology," Grokker starts organizing data from the multiple search engines. You see a big circle, within which are smaller circles with labels including "conference," "technology," "science," "research," "reports," "news," "molecular," "material," and so on. Each represents a subset of data on nanotechnology. Click on, say, "molecular," and that circle will enlarge so you can see several further subcircles, one of which is "molecular assemblies." Click on that, and another category becomes visible entitled "molecular assembly sequencing software." Now you could, in theory, have typed that exact phrase into Google and gotten to the same websites. However, in many cases you can't be sure what you're looking for because you simply don't know what's out there. Grokker gives you an easy way to delve into a data set, and it often leads to info-revelations (see Figure 1).

Figure 1: Example of Grokker Graphical Results.

Grokker is available at www.groxis.com and the company is making the Grokker SDK (software development kit) available as open source soon, to allow developers to expand this capability.

Conclusion

While Google and Yahoo! expand the current search capabilities by reaching for one small semantic step at a time other organizations are experimenting with bold semantic jumps, such as ontology matching algorithms. In this column, we will explore many of the innovation in semantic search technology coming from explorers around the world, including Latent Semantic Indexing, Stanford's TAP and more.

References:
1. Alesso, H. Peter, and Smith, Craig F., Developing Semantic Web Services, A. K. Peters, Ltd., ISBN: 1-56881-212-4, 2004.

Semantic Web Technologies by Dr. Jessica Chen Burger

Yun-Heh (Jessica) Chen-Burger
Room 4.08, Appleton Tower
AIAI, CISA, Informatics
The University of Edinburgh, UK
+44-131- 650-2756 (Office)
jessicac@inf.ed.ac.uk

An Over-Arching Description for the
Semantic Web Technologies Column
For SIGSEMIS: Semantic Web and Information Systems

http://www.sigsemis.org/columns/technologiesColumn/

For this bi-monthly Semantic Web Technologies Column, I plan to cover various advanced technologies that is relevant to the field of semantic web technologies.

Research topics cover but not limited to:

- Knowledge Management techniques;
- Advanced Knowledge technologies;
- Grid Computing technologies, esp. Semantic Grid technologies;
- Enterprise Modelling and its applications in assisting the development of semantic web and knowledge management;
- Verification and validation techniques that is applicable to semantic web/rich technologies;
- Collaborative systems and their cooperative operations based on semantic web/rich technologies;
- Workflow systems that understand, manipulate and execute semantics rich information;
- Web services as well as over-arching architecture that holds different web services together;
- Advancements in process modelling and workflow technologies, esp. their relations to the semantic web;
- Applications based on advanced semantic web/rich technologies, e.g. advancements in the bio-informatics;
- Development and applications of ontology technologies; e.g. mapping, evolution, negotiation and the use of ontologies;
- Advanced information technologies, e.g. information extraction, knowledge capture, natural language generation/presentation based on information captured using IE, etc.
- Knowledge portal applications;
- Evaluation and critique of current semantic web/rich technologies and their applications;
- A combination of some of the above technologies.

While some/most columns will be entirely contributed on my own, guest authors may be sometimes invited to contribute to the content of the column, when appropriate. Guest authors may also be different each time. This is an attempt to provide in-depth knowledge to the column as well as broaden its views. In order to acknowledge their efforts, their name may appear as a co-author, when appropriate. The responsibilities for the make-up of the column, however, entire rest on myself.

Advanced Knowledge Technologies Meet the Challenges of the Semantic Web

Yun-Heh Chen-Burger[1], Kieron O'Hara[2]

[1]Informatics, The University of Edinburgh,
[2]Department of Electronics and Computer Science,
The University of Southampton

As information is becoming more abundantly available through the World Wide Web (WWW), the need for more effective storage, retrieval and presentation methods of such information becomes more prominent. Equally important are the capabilities for extraction and thus utilisation of the underlying *knowledge* that is often deeply embedded in the information mass that has been made available. Technologies related to such capabilities are often referred to as Knowledge Technologies. One interesting application area for such technologies is the Semantic Web, that emphasises the intelligent manipulation of the knowledge that is made available via the web.

Richard Benjamins, et al [5] described in this journal six challenges of the Semantic Web (SW) as being the availability of semantic web content; the availability, development and evolution of ontologies; scalability of managing knowledge content; multilingual information access; visualisation; and semantic web language stability. They make the challenges as well as opportunities of the SW admirably clear.

The Advanced Knowledge Technologies (AKT) project [1] has looked at issues of the creation, extraction, derivation and dissemination of knowledge at all points of the knowledge management life cycle, including knowledge acquisition and modelling, reuse, retrieval and publishing, right through to the tricky issues of maintaining knowledge repositories. These knowledge technologies interact with the Semantic Web and therefore had to address all the challenges described by Benjamins, et al. For instance, many AKT technologies exploit state-of-the-art natural language processing techniques, both to extract content from large-scale unstructured texts, and to publish human-readable as well as machine-processable content from structured repositories; such techniques will support multilingual data [3][9] [10][11].

On the other hand, decentralised web services powered by underlying semantically rich processing methods that may be flexibly coupled with specialised problem solvers can address some of the scalability issues [6][8][12][15]. Creation, evolution and utilisation of knowledge models [7] and ontologies play a central role for most of the AKT work and remain a main focus for its vision [4][13][14]. The AKT Reference Ontology provides a set of coherent ontological modules that may be used complimentary to suit individual applications [2]. In addition, many AKTors play prominent roles on various SW standards committees, and so are helping shape the future of the web and its languages.

Throwing technologies at the SW alone will not by itself realise SW's full potential; social aspects of information/knowledge handling need to be understood by technology designers and adopted by its users. Part of AKT's work is thus to look at the behaviour of communities that may (or may not) operate with the assistance of WWW/SW. Social theories and practices are being borrowed and adapted to strengthen technologies that are being offered to such communities [16][17].

Ultimately, the opportunity that the SW affords is to provide easier access to information and knowledge stored in many heterogeneous sources. This power of obtaining knowledge is combined with shared and collaborative distributed computational capabilities to assist both humans and software, that may scatter in different geographical areas, in achieving their goals. AKT contributes towards this goal by providing a wide spectrum of Semantic Web related advanced research as well as practical applications. One of its results, the CS AKTive Space [11] won the 2003 Semantic Web Challenge awarded at the International Semantic Web Conference in Florida – show that the effort will be well worth it.

This paper includes some of AKT's work; for more information, its official website at http://www.aktors.org lists approximately 50 selected technologies, together with reader-friendly descriptions of the real-world knowledge problems they are intended to address.

References

[1] Advanced Knowledge Technologies: http://www.aktors.org.

[2] AKT Reference Ontology: http://www.aktors.org/publications/ontology/.

[3] Alani, Harith and Kim, Sanghee and Millard, David and Weal, Mark and Hall, Wendy and Lewis, Paul and Shadbolt, Nigel. Automatic Extraction and Generation of Knowledge from Web Documents. *Human Language Technology for the Semantic Web and Web Services, ISWC 2003*, Sanibel Island, Florida, USA.

[4] Alani, Harith and Kim, Sanghee and Millard, David and Weal, Mark and Hall, Wendy and Lewis, Paul and Shadbolt, Nigel. Web based Knowledge Extraction and Consolidation for Automatic Ontology Population. In *Proceedings Knowledge Markup and Semantic Annotation Workshop, 2nd Int Conf on Knowledge Capture (K-CAP 2003),* Sanibel Island, Florida, USA.

[5] Benjamins, V. Richard and Contreras, Jesús and Corcho, Oscar and Gómez-Pérez, Asunción. Six Challenges for the Semantic Web. *SIGSEMIS Bulletin*, April 2004.

[6] Chen-Burger, Yun-Heh and Hui, Kit-Ying and Preece, Alun and Gray, Peter and Tate, Austin. Supporting Collaboration through Semantic-based Workflow and Constraint Solving . *14th International Conference on Knowledge Engineering and Knowledge Management EKAW* 2004 5-8th October 2004 - Whittlebury Hall, Northamptonshire, UK .

[7] Chen-Burger, Yun-Heh and Robertson, Dave. *Automating Business Modelling.* Book Series of Advanced Information and Knowledge Processing, Springer Ver-Lag, October 2004.

[8] Chen-Burger, Yun-Heh and Stader, Jussi. Formal Support for Adaptive Workflow Systems in a Distributed Environment , Section I, Chapter of book: *Workflow Handbook 2003*, Editor: Layna Fischer. Published in association with Workflow Management Coalition. Publisher: Future Strategies Inc., USA, April 2003.

 [9] Ciravegna, Fabio. An Adaptive Algorithm for Information Extraction from Web-related Texts. In *Proceedings IJCAI-2001 Workshop on Adaptive Text Extraction and Mining*, Seattle.

[10] Domingue, John and Dzbor, Martin and Motta, Prof Enrico. Collaborative Semantic Web Browsing with Magpie. In Davies, John and Fensel, Dieter and Bussler, Christoph and Studer, Rudi, Eds. *Proceedings 1st European Semantic Web Symposium (ESWS)* Lecture Notes in Computer Science, Volume 3053, Hersonissos, Crete, Greece, 2004.

 [11] Glaser, Hugh and Alani, Harith and Carr, Les and Chapman, Sam and Ciravegna, Fabio and Dingli, Alexiei and Gibbins, Nicholas and Harris, Stephen and schraefel, m.c. and Shadbolt, Nigel. CS AKTiveSpace: Building a Semantic Web Application, in Bussler, Christoph and Davies, John and Fensel, Dieter and Studer, Rudi, Eds. *The Semantic Web: Research and Applications* (First European Web Symposium, ESWS 2004), pages pp. 417-432. Springer Verlag, 2004.

[12] Gray, Peter and Hui, Kit and Preece, Alun. An Expressive Constraint Language for Semantic Web Applications. In Preece, Alun and O'Leary, Prof Daniel, Eds. *Proceedings IJCAI-01 Workshop on E-Business and the Intelligent Web*, pages pp. 46-53, Seattle, USA, 2001.

[13] Kalfoglou, Yannis and Dasmahaptra, Srinandan and Chen-Burger, Yun-Heh. FCA in Knowledge Technologies: Experiences and Opportunities. In *Proceedings 2nd International Conference on Formal Concept Analysis* LNCS 2961, pages pp. 252-260, Sydney, Australia, 2004.

[14] Kalfoglou, Yannis and Domingue, John and Carr, Leslie and Motta, Enrico and Vargas-Vera, Maria and Buckingham Shum, Simon. *On the integration of technologies for capturing and navigating knowledge with ontology-driven services.* Technical Report 106, Knowledge Media Institute (KMi), The Open University. 2001.

[15] Lino, Ms. Natasha Queiroz and Tate, Austin and Siebra, Mr. Clauirton and Chen-Burger, Yun-Heh. Delivering Intelligent Planning Information to Mobile Devices Users in Collaborative Environments. In *Proceedings 18th International Joint conference on Artificial Intelligence (IJCAI), AI Moves to IA: Workshop on Artificial Intelligence, Information Access, and Mobile Computing*, Acapulco, Mexico, 2003.

[16] O'Hara, Kieron. *Plato and the Internet*. Icon Books, Cambridge, 2002.

[17] O'Hara, Kieron. *Trust: From Socrates to Spin*. Icon Books, Cambridge, 2004.

Methodologies for the Semantic Web Column by Dr. Matteo Cristani

Matteo Cristani is Assistant Professor in the University of Verona (Italy), Department of Computer Science. He was born in Verona (Italy) in 1966. He graduated in the University of Milan in 1991 and obtained a PhD from the University of Padova in 1995. He was employed as post-doctoral research associate in the University of Padova. He is employed by the University of Verona since 1997.

His first research interest has been Natural Language Processing, and then Temporal Reasoning, the theme on which he did the doctoral dissertation. His current main research interest is Artificial Intelligence, in particular Ontology on the Web, Spatial reasoning and Aesthetic Knowledge Representation. He has published in outstandingly International Conferences and Journals in the recent years, like the European Conference on Artificial Intelligence, the International Conference on Principles of Knowledge Representation and Reasoning, the International Joint Conference on Artificial Intelligence, the Journal of Visual Languages and Computing, Artificial Intelligence, the Journal of Artificial Intelligence Research, Spatial Cognition and Computation. Recently he developed an interest in the theme of Ontology methodology, from a Knowledge Management point of view. This has been delivered as a coordinated industrial project. He currently leads a long-term research activity in cooperation with industry about Ontology in Services and Industrial Production, and has an established long-term cooperation with outstanding research centres, including the University of Leeds, IRIT-Toulouse, Cambridge University, Napier University. He also published in I-KNOW 2004, and the Workshop on Terminology, Ontology and Knowledge Representation.

ABOUT THE COLUMN

The column will discuss and present the up-to-date situation of outstanding international research about Knowledge Management and Ontology Engineering as applied to the continuously growing field of Semantic Web, with a specific attention the applications to Information Systems.

There will appear alternatively the following types of articles:

1) *Research reviews*. This papers will briefly discuss a theme of interest of the communities of Semantic Web, Artificial Intelligence, Information Systems, Web Languages, Knowledge Management to which researchers have paid attention in the recent past. The papers will only review those investigations that have proven to be outstanding in terms of internationally recognised quality, such as international conferences, international journals, and books published by recognised international scientific publishers. The major concern of these papers will be to provide affordable up-to-date state-of-the-art summaries of major themes in the field of Methodologies for the Semantic Web. The columnist is responsible for the retrieval of the material and will be also providing the actual summary;

2) *Recent achievements updates*. This papers will introduce a new theme on which well-known authors have recently achieved results that still do not appear in outstanding forums, such as research reports, workshops, or other minor publications. The major concern of these papers will be to report from promising minor events and minor publications in order to provide a forum for new ideas to be discussed within the community;

3) *Public debate reports*. The columnist will report public debates commenced on the major mailing lists on the topic, like DBWORLD, SEWORLD and, obviously, Semantic Web on yahoo. These reports will focus both on self-arising debates, like the now well famous one about the achievements of the Semantic Web, but only about the selected topic. The public debates reported will be either appearing in some of those mailing lists, or being reported to the columnist, or provoked directly on such lists by

the columnist;

4) *Guest papers.* Outstanding guests will be occasionally invited to provide valuable opinions on the achievements obtained in the field.

The themes of the column are fundamentally four:

- *Distributed ontology systems.* We focus on the theme of Knowledge Management in a distributed environment;
- *Web ontologies.* Ontologies which can be shared on the Web;
- *Ontological Engineering.* Methods for building ontologies;
- *Methodologies for Semantic Web Technologies.* Methods for building ontologies on the web by means of knowledge-based languages, like OWL.

The column will briefly review the current state-of-the-art at least twice a year.

Methodologies for the Semantic Web: state-of-the-art of ontology methodology

Matteo Cristani (columnist) and Roberta Cuel (guest columnist)
Dipartimento di Informatica, Università di Verona, Italy
matteo.cristani@univr.it

1 Introduction

The Semantic Web community has grown faster and faster in the recent past, and several different viewpoint have been taken upon the nature of the web in the semantic era. The scope of this column is on the methodologies that are proposed within the various communities viewing the web as an object of interest from a semantical point of view. In particular we focus on the specific needs of the information systems communities, in particular Intelligent Information Systems and Knowledge Bases, which are the ones mostly interested in the Web as a semantic object.

This first column, in particular, focuses upon the methodologies that exist in literature which provide affordable frameworks for developing *Ontologies*. We are especially interested in Ontologies because that is the boundary topic between the Semantic Web research and Information Systems that mostly deserves a special attention to the methodological aspects. This is the argument that also urged several researchers over all the world to create outstanding projects that also involve methodological research.

This first column employs the results found in [28].

This paper is organised as follows: Section 2 introduces the general problems a methodology for ontology development has to deal with, and Section 3 reviews the existing methodologies that have been consolidated in recent ontological engineering research.

2 General

It is quite well established in recent investigation on Information Systems, that *formal ontologies* are a crucial problem to deal with, and in fact they received a lot of attention in several different communities, such as knowledge management, knowledge engineering, natural language processing, intelligent information integration, and so on [6]. Ontologies have been developed in Artificial Intelligence to facilitate knowledge sharing and reuse.

The viewpoint we adopt here is taken from the general considerations on the use of philosophical issues in Artificial Intelligence: "the systematic, formal, axiomatic development of the logic of all forms and modes of being" [27]. Another commonly accepted definition is that an ontology is an explicit specification of a shared conceptualization that holds in a particular context.

The actual topic of ontology is one of those themes that epistemology (theories on knowledge) dealt with in philosophical studies of Parmenides, Heraclitus, Plato, Aristotle, Kant, Leibnitz, Wittgenstein, and others.

Ontologies define the kind of things that exist in the world and, possibly, in an application domain. In other words, an ontology provides an explicit conceptualization which describes semantics of data, providing a shared and common understanding of a domain. From an AI perspective we can say that:

> "...ontology is a formal explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine readable. Shared reflect that ontology should capture consensual knowledge accepted by the communities[13]."

And moreover:

> "An ontology may take a variety of forms, but necessarily it will include a vocabulary of terms, and some specification of their meaning. This includes definition and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretation of terms" [15]

Nowadays, ontologies

- are used to allow communication among people and heterogeneous and widely spread application systems;
- are implied in projects as a conceptual model, to enable a content-base access on corporate knowledge memories, knowledge bases, archives;
- allow agents to understand each other when they need to interact, communicate and negotiate meanings;
- refer to a common piece of information and share common understanding of the information structure.

In other words, ontologies provide *qualitatively new* levels of services in several application domains such as the Semantic Web [3] or federated databases. They moreover enable reuse of domain knowledge, make domain assumption explicit, and separate domain knowledge from the operational knowledge.

One of the first steps in ontology creation is choosing domains and categories which allow the correct representation. In particular philosophers have tried to define very general and universal categories which are supposedly able to describe the real world. The main idea is to develop an understandable, complete and sharable system of categories, labels, relations, which represent, in an objective way, the real world.

For instance, one of the interesting results achieved by Aristotle is the definition of general categories used to describe the main features of events, situations, and objects in the worlds: quality, quantity, activity, passivity, having, situated, spatial, temporal. Kant, figured out only four macro-categories used to describe the world: quantity, quality, relation, modality.

Unfortunately, in the "real world" or in "practical applications" (i.e. information systems, knowledge management systems, portals, and other ICT applications) these general and universal categories are not widely being used. In particular few problems are:

- it is difficult to implement a general ontology within specific domains;
- it is too expensive to create very complex, complete, and general ontologies;

Another important consideration is that, in the same project or domain, people might use different ontologies composed by various combinations of categories. This means that different ontologies might use different categories or systems of categories to describe the same kinds of entities; or even worse, they may use the same names or systems of category for different kinds of entities. Indeed, it might be that two entities with

different definitions are intended to be the same, but the task of proving that they are indeed the same may be difficult, if not impossible [22].

The basic assumption of this behavior is that what we know cannot be viewed simply as a *picture* of the world, as it always presupposes some degree of interpretation. Different categories represent different perspectives, aims, and degree of world interpretation. Indeed, depending on different interpretation schemas, people may use the same categories with different meaning, or different words to mean the same thing. For example two groups of people may observes the same phenomenon, but still see different problems, different opportunities, and different challenges. This essential feature of knowledge was studied from different perspectives, and the interpretation schemas were given various names, for example paradigms [16], frames [11], thought worlds [4], context [10], mental spaces [5], cognitive path [26]. This view, in which the explicit part of what we know gets its meaning from a (typically implicit, or taken for granted) interpretation schema leads to some important consequences regarding the adoption and the use of categories and ontologies.

It follows from what we said above, that an ontology is not a neutral organization of categories, but is the emergence of some interpretation schema, according to which it makes sense to organize and define things in that way. In short, an ontology is always the result of a sense-making process, and represents the point of view of those who took part in that process (see [1] for a in-depth discussion of the dimensions along which any representation – including an ontology – can vary depending on contextual factors).

3. Some methodologies

In the computer science, knowledge management, knowledge representation, and other fields, a lot of languages and tools are developed with the aim of helping people and system developer to create good and effective ontologies. In particular a lot of tools help people to create manually, or semi-automatic, categories, partonomies, taxonomies, and so on. Indeed, behind these tools and techniques different approaches, and different methods and techniques are used to develop a big quantity of heterogeneous ontologies. In this section we will describe some of them and we try to compare the more significant principle that are sustaining these approaches.

In this subsection some methodologies, for knowledge discovery and ontology creation, will be presented. Until now, few domain-independent methodological approaches have been reported for building ontologies. In particular Uschold's methodology (who propose codification in a formal language) [24] and Methontology (which express the idea as a set of intermediate representation and then generation the ontology using translators) [12], [8] are the most representative.

3.1 DOLCE: Descriptive Ontology for Linguistic and cognitive Engineering

The main authors' idea is to develop not a monolithic module but a library of ontologies (WonderWeb Foundation Ontologies Library) which allows agents to understand one another despite of enforce them to interoperate by adoption of a single ontology [20]. They developed few starting points for building new ontologies which are:

1. determine what things there are in the domain to be modelled;
2. develop easy and rigorous comparisons among different ontological approaches;
3. analyses, harmonize, and integrate existing ontologies and meta-data standards;
4. describe a foundational ontology on paper, using a full first order logic with modality;
5. isolate the part of the axiomatization that can be expressed in OWL, and implement it;
6. add the remaining part in the form of KIF comments attached to OWL concepts

In addition, they intend the library to be:

- minimal: the library is as general as possible, including only the most reusable and widely applicable upper-level categories;
- rigorous: ontologies are characterized by means of rich axiomatizations;

- extensively researched: modules in library are added only after careful evaluation by expert and consultation with canonical works.

One of the first module of their foundational ontologies library is a Descriptive Ontology for Linguistic Cognitive Engineering (DOLCE). DOLCE is an ontology of particulars and refers to cognitive artifacts which depend on human perception, cultural imprints, and social conventions.

Their ontology derives from harm chair research in particular, referring to enduring and perduring entities, from philosophical literature.

Finally basic functions and relations (according to the methodology introduced by Gangemi [9]) should be:

- general enough to be applied to multiple domains;
- sufficiently intuitive and well studied in the philosophical literature;
- hold as soon as their relata are given, without mediating additional entities.

3.2 Ontology Development 101

This methodology has been developed by authors involved in these ontology editing environments: Protégé-2000, Ontolingua, and Chimaera [21]. They propose a very simple guide, based on iterative design, that helps developer to create an ontology using these tools. The sequence of the steps to develop an ontology is described in the following table:

Name of the Phase	Input	Phase Description	Output
determine domain and scope or the ontology	nothing. It is the first step	definition of what is the domain that the ontology will cover, for what ontology will be used, for what types of question the ontology should provide answers (competency questions are very important in this domain which allow the designer to understand when ontology contains enough information, and when it achieve the right level of detail or representation), and who will use and maintain the ontology.	the resulting document may change during the whole process, but at any time this documentation helps to limit the scope of the model.
consider reusing existing ontologies	documents with the domain and the scope of the ontology	looking for other ontologies which are defining the domain. There are libraries of reusable ontologies on the web and in literature (e.g. Ontolingua ontology library, DAML ontology library, UNSPSC, RosettaNet, and DMOZ)	one or more domain ontologies, or part of them with their description
enumerate important terms in the ontology	documents with the domain, the scope of the ontology, and libraries on the domain	write a list of all terms used within the ontology, and describe the terms, their meaning, and their properties	terms and important aspect to model into the ontology
define the classes and the class hierarchy	important terms in the ontology, domain and scope description	there are several possible approaches in developing a class hierarchy [25]: top down development process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts; bottom up development process goes on the opposite direction; a combination development process is a combination of the top-down and bottom-up approaches	classes and class hierarchy
define the properties of classes-slot	the taxonomy, and the domain and scope description	add all the necessary properties and information which allow the ontology to answer the competency questions	classes, class hierarchy and properties
define the facets of the slots	slots and classes	there are different facets describing the value type, allowed values, the number of the values, and other features of the values the slot can take: slot cardinality, slot-value type, domain and range	ontology
create instances	the ontology	create individual instances of classes in the hierarchy, that means: choosing a class, creating an individual instance of that class, and filling in the slot values.	ontology and the modelled domain

3.3 OTK Metholdology

The methodology developed within the On-To-Knowledge project is called OTK Methodology, and it is focused on application-driven development of ontology during the introduction of ontology-based knowledge management systems [23], [7], [17].

It is based on the following steps:

- feasibility study: to identify problem/opportunity areas and potential solutions, and put them into a wider organizational perspective. In general, a feasibility study serves as a decision support for economical, technical and project feasibility, in order to select the most promising focus area and target solution. The process of study the feasibility of the organization, the task, and the agent model, proceed in the following steps:
  - carry out a scoping and problem analysis study, consisting of two parts:
    - identifying problem/opportunity areas and potential solutions, and putting them into a wider organizational perspective.
    - deciding about economic, technical and project feasibility, in order to select the most promising focus area and target solution.
  - carry out an impacts and improvements study, for the selected target solution, again consisting of two parts:
    - gathering insights into the interrelationships between the business task, actors involved, and use of knowledge for successful performance, and what improvements may be achieved here.
    - deciding about organizational measures and task changes, in order to ensure organizational acceptance and integration of a knowledge system solution.
- kick-off phase: it starts with an ontology requirements specification document (ORSD). The ORSD describes what an ontology should support, sketching the planned area of the ontology application and listing, e.g., valuable knowledge sources for the gathering of the semi-formal description of the ontology. The ORSD should guide an ontology engineer to decide about inclusion and exclusion of concepts and relations and the hierarchical structure of the ontology. In this early stage one should look for already developed and potentially reusable ontologies. The result will be a document containing:
  - goal, domain and scope of the ontology;
  - design guidelines;
  - knowledge sources;
  - (potential) users and usage scenarios;
  - competency questions;
  - supported applications;
- refinement phase: the goal of the refinement phase is to produce a mature and application-oriented "target ontology" according to the specification given by the kick-off phase. This phase is divided into different sub-phases:
  - a knowledge elicitation process with domain experts based on the initial input from the kick-off phase. The knowledge about the domain is captured from domain experts in the previous mentioned competency questions or by using brainstorming techniques;
  - a formalization phase to transfer the ontology into the target ontology expressed in formal representation language like DAML+OIL. To formalize the initial semi-formal description of the ontology we firstly form a taxonomy out of the semi-formal description of the ontology and add relations other than the "is-a" relation which forms the taxonomical structure.

  The refinement phase is closely linked to the evaluation phase. If the analysis of the ontology in the evaluation phase shows gaps or misconceptions, the ontology engineer takes these results as an input for the refinement phase. It might be necessary to perform several (possibly tiny) iterative steps to reach a sufficient level of granularity and quality.

- evaluation phase: it serves as a proof for the usefulness of developed ontologies and their associated software environment.

- application and evolution phase: it is based on defining strict rules for the update, insert, and delete processes within ontologies.

3.4 Methontology

One of the most famous ontology design environment is "Methontology". It try to define the needed activities that people need to carry out when building an ontology[8]. In other words a flow of ontology development process for three different processes: management, technical, supporting.

The ontology development process is composed by the following steps:

- project management activities include:
  - planning: it identifies which tasks are to be performed, how they will be arranged, how much time and what resources are needed for their completion;
  - control: it guarantees that planned tasks are completed in the manner that they were intended to be performed;
  - quality assurance: it assures that the quality of each and every product output is satisfactory;
- development-oriented activities include:
  - specification: it states why the ontology is being to built and what are its intended uses and who are end-users;
  - conceptualization: structures the domain knowledge as meaningful models at the knowledge level;
  - formalization: transforms the conceptual model into a formal o semi computable model;
  - implementation: builds computable models in a computational language.
- support activities: include a series of activities, performed at the same time as development oriented activities:
  - knowledge acquisition;
  - evaluation: it makes a technical judgment of the ontologies, their associated software environment, and documentation with respect to a frame of reference;
  - integration;
  - documentation.

In below we provide a table listing the single phases of the above mentioned methodology. It is worth now also depict one fundamental fact about methontolgy: its specific explicit avoidance to commit to the users' interest in the development of tools which automates all the phases. We look for a systematic help, not necessarily an automation of the process.

Name of the phase	Input	Description	Output
Planning	nothing: fist step	plan the main tasks to be done, the way in which they will be arranged, the time and resources that are necessary to perform these tasks	a project plan
Specification	a series of question such as: "why this ontology is being built and what are its intended uses and end-users"	identify ontology's goals	ontology requirements specification document, specifying purposes and scopes. Its goal is to produce either an informal, semi-formal or formal ontology specification document written in natural language, using a set of intermediate representations or using competency question respectively. The document has to provide at least the following information: the purpose of the ontology (included its intended uses, scenarios of use, end-users, ...); the level of formality used to codify terms and meanings (high informal, semi informal, semi formal, rigorously formal ontologies); the scope; its characteristics and granularity. Properties of this documents are: concision:; partial completeness: coverage of terms, the stopover problem and level of granularity of ache and every term; consistency of all terms and their meanings.

Conceptuali-zation	a good specification document	conceptualize in a model that describes the problem and its solution. To identify and gather all the useful and potentially usable domain knowledge and its meanings	A complete glossary of terms (including concepts, instances, verbs, and properties). Then a set of intermediate representations such as concepts classification trees, verb diagram, table of formulas and table of rules. The aim is to allow the final user to: ascertain wether or not an ontology is useful and; compare the scope and completeness of several ontologies, their reusability, and shareability.

Name of the Phase	Input	Phase Description	Output
formalization	conceptual model	transform conceptual model into a formal or semi-compatible model, using frame oriented or description logic representation systems	formal conceptualization
Integration	existing ontologies and the formal model	processes of inclusion, polymorphic refinement, circular dependencies and restriction. For example, select meta ontologies that better fit the conceptualization	
implementation	formal model	select target language	create a computable ontology
maintenance		including, modifying definition in the ontology	guidelines for maintaining ontologies
Acquisition		searching and listinknowledge sources through non structured interviews with experts, informal text analysis, formal text analysis, structured interviews with expert to have detailed information on concepts, terms, meanings, and so on	a list of the sources of knowledge and a rough description of how the process will be carried out, and of the techniques are used.
Evaluation	computable ontology	technical judgment with respect to a frame of reference	a formal and correct ontology
documentation			Specification document must have the property of concision

3.4.1 A multilingual domain ontologies

The authors [18] use the methodology defined by [8] METHONTOLOGY, and enrich this one stressing on specific actions for supporting the creation process for ontology-driven conceptual analysis.

The domain ontology is built using two different knowledge acquisition approaches:

- acquisition approach 1: creation of the core ontology. A small core ontology with the most important domain concepts and their relationships is created from scratch. This stage is basically comprised of the first three steps of METHONTOLOGY development activities: requirement specification, conceptualization of domain knowledge, formalization of the conceptual model in a formal language. The goal of this step is to define a list of frequent terms and a list of domain specific documents to analyze;

- acquisition approach 2: deriving a domain ontology from thesaurus. A thesaurus consists of descriptive keywords linked by a basic set of relationships. The keywords are descriptive in terms of the domain in which they are used. The relationships may either descrive a hierarchical relation or an inter-hierarchical relation. The goal of this step is to refine a RDFS ontology model to develop a pruned ontology and a list of frequent terms;

- ontology merging: merging the manually created core ontology and the derived ontology using thesaurus terms;

- ontology refinements and extension: the frequent domain terms are used as possible candidate concepts or relationships for extending the ontology. These terms have to be assessed by subject specialists and checked for relevance to the ontology.

3.5 TOVE

Toronto Virtual Enterprise (TOVE) is a methodology for ontological engineering which allows the developer to build ontology following these steps:

- motivating scenarios: the start point is the definition of a set of problems encountered in a particular enterprise;
- informal competency questions: based on the motivating scenario, it is the definition of ontology requirements described as informal questions that an ontology must be able to answer;
- terminology specification: the objects, attributes and relations of the ontology are formally specified (usually first order logic);
- formal competency question: the requirements of the ontology are formalized in terms of the formally defined terminology
- axiom specification: axioms that specify the definition of terms and constraints on their interpretations are given in first order logic;
- completeness theorems: an evaluation stage which assesses the competency of the ontology by defining the conditions under which the solutions to the competency question are complete.

The most distinctive aspect of TOVE is the focus on maintenance, using a formal techniques to address a limited number of maintenance issues.

3.6 A natural language interface generator (GISE)

In [19] the authors developed three steps process to build a domain ontology:

- first step: building and maintenance of:
  - the general linguistic knowledge. It includes:
    - linguistic ontology that covers the syntactic and semantic information needed to generate the specific grammars;
    - general lexicon that includes functional and domain and application independent lexical entries;
  - the general conceptual knowledge: it includes both the domain and application independent conceptual information and the meta-knowledge that will be needed in the following steps.
- second step: definition of the application in terms of the conceptual ontology. Both the domain description and the task structure description must be built and linked to the appropriate components of the conceptual ontology;
- third step: definition of the control structure. It includes:
  - the meta-rules for mapping objects in the domain ontology with those in the task ontology,
  - the meta-rules for mapping the conceptual ontology onto the linguistic ontology and those for allowing the generation of the specific interface knowledge sources, mainly the grammar and the lexicon.

3.7 Business object ontology

The authors Izumy and Yamaguchi [14] have used this methodology to develop an ontology for business coordination. They constructed the business activity repository by employing WordNet as a general lexical repository. They have constructed the business object ontology in the following way:

- concentrating the case-study models of e-business and extracting the taxonomy;
- counting the number of the appearance of each noun concept
- comparing the noun hierarchy of WordNEt and the taxonomy obtained and adding the number counted for the similar concepts;
- choosing the main concept with high scores as upper concepts and building upper ontologies by giving all the nouns the formal is-a relation;
- merging all the noun hierarchy extracted from the whole process.

4 Some consideration comparing these methodologies

Although there are considerable differences between the methodologies described above, a number of points clearly emerge:

- many of the methodologies take a task as a starting point. >From one point of view it focuses the acquisition, provides the potential for evaluation and provides a useful description of the capabilities of the ontology, expressed as the ability to answer well defined competency question. On the other side, it seems to provide limitations to the re-use of the ontology, and to the possible interactions among ontologies;
- there are two different types of methodology models: the stage-based models (represented for example by TOVE) and evolving prototype models (represented by Methontology). Both approaches have benefits and drawbacks: the first one seems more appropriate when purposes and requirements of the ontology are clear, the second one is more useful when the environment is dynamic and difficult to understand;
- most of the time there are both informal description of the ontology and formal embodiment in an ontology language. These are often developed in separated stages, and this separation increase the gap between real world models and executable systems.

The common point in these methodologies are the starting point for creating an ontology which could arise from different situation [25]:

- from scratch;
- from existing ontologies (whether global or local);
- from corpus of information sources only;
- a combination of the latter two approaches.

Normally, methods to generate ontology could be summarized as [3]:

- bottom-up: from specification to generalization;
- top-down: from generalization to specification such as KACTUS ontology;
- middle-out: from the most important concepts to generalization and specialization such as Enterprise ontology and Methondology;

There are also a number of general ontology design principles that are proposed:

- Guarino proposed a methodology to design a "Formal ontology" in particular defining domain, identity a basic taxonomic structure and explicit the roles [20];
- Uschold and Gruninger (1996) proposed a skeletal methodology for building ontologies via a purely manual process: identify purpose and scope, ontology capture (identification of key concepts and relationships and provision of definitions), ontology coding (committing to the basic terms for ontology), language, integrating existing ontologies, evaluation, documentation, guidelines [24];
- Reich (1999) separates the construction and the definition of complex expression from its representation;

Other authors proposed a number of desirable criteria for the final generated ontology to be open and dynamic, scalable and interoperable, easily maintained, and context independent.

There is no one correct way to model a domain, there are always viable alternatives. Most of the time the best solution depends on the application that the developer has in mind, and the tools that she use to develop the ontology. In particular, we can notice some emerging problems: [2] [12]:

- the correspondence between existing methodologies for building ontologies and environments for building ontologies, causes this consequences:
    - conceptual models are implicit in the implementation codes and a re-engineering process is usually required to make the conceptual models explicit;
    - ontological commitments and design criteria are implicit in the ontology code;
    - ontology developer preferences in a given language condition the implementation of the acquired knowledge. So, when people code ontologie directly in a target language, they are omitting the minimal encoding bias criterion defined by [13];
- most of the tools only give support for designing and implementing the ontologies, but they do not support all the activities of the ontology life-cycle;

- ontology developers may find it difficult to understand implemented ontologies, or even to build a new ontology.

References

[1] M. Benerecetti, P. Bouquet, and C. Ghidini, *Contextual Reasoning Distilled*, Journal of Theoretical and Experimental Artificial Intelligence **12** (2000), no. 3, 279–305.

[2] M. Blázquez, M. Fernández, J. M. García-Pinar, and A. Gómez-Pérez, *Building Ontologies at the Knowledge Level using the Ontology Design Environment*.

[3] Y. Ding and S. Foo, *Ontology Research and Developement, Part 1 – A Review of Ontology Generation*, 2002.

[4] D. Dougherty, *Interpretative barriers to successful product innovation in large firms*, Organization Science **3** (1992), no. 2.

[5] G. Fauconnier, *Mental spaces: aspects of meaning construction in natural language*, MIT Press, 1985.

[6] D. Fensel, *Ontologies: A silver bullet for knowledge management and electronic commerce*, Springer, 2000.

[7] D. Fensel, F. van Harmelen, M. Klein, and H. Akkermans, *On-To-Knowledge: Ontology-based Tools for Knowledge Management*.

[8] M. Fernández, A. Gómez-Pérez, and N. Juristo, *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*, In Working Notes of the AAAI Spring Symposium on Ontological Engineering. Stanford University, AAAI Press. (1997), Stanford, CA.

[9] A. Gangemi, D. M. Pisanelli, and G. Steve, *Ontology Integration: Experiences with Medial Terminologies*.

[10] C. Ghidini and F. Giunchiglia, *Local Models Semantics, or Contextual Reasoning = Locality + Compatibility*, Artificial Intelligence **127** (2001), no. 2, 221–259.

[11] I. Goffman, *Frame analysis*, Harper & Row, New York, 1974.

[12] A. Gómez-Pérez, *A proposal of infrastructural needs on the framework of the semantic web for ontology construction and use*.

[13] T. R. Gruber, *A translation approach to portable ontology specifications*, Knowledge Acquisition **5** (1998), 199–220.

[14] N. Isumi and T. Yamaguchi, *Semantic Coordination of Web SErvices Based on Multi-Layered Repository*.

[15] R. Jasper and M. Ushold, *A Framework for Understanding and Classifying Ontology Applications*.

[16] T. Kuhn, *The structure of scientific revolutions*, University of Chicago Press, 1979.

[17] T. Lau and Y. Sure, *Introducing Ontology-based Skill Management at a large Insurance Company*.

[18] B. Lauser, T. Wildemann, A. Poulos, F. Fisseha, J. Keizer, and S.Katz, *A Comprehensive Framework for Building Multilingual Domain Ontologies: Creating a Prototype Biosecurity Ontology*, Proc. Int. Conf. on Dublin Core and Metadata for e-Communities (2002), 113–123, Firenze University Press.

[19] G. Marta and R. Horacio, *A domain-restricted task-guided Natural Language Interface Generator*.

[20] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari, *Wonderweb deliverable d17*, intermediate report 2.0, ISTC-CNR, 2002.

[21] N. F. Noy and D. L. McGuinnes, *Ontology development 101: a guide to creating your first ontology*.

[22] J. F. Sowa, *Knowledge Representation. Logical, Philosophical and Computational Foundations*, Brooks/Cole, 2000.

[23] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, and D. Wenke, *OntoEdit: Collaborative Ontology Development for Semantic Web*.

[24] M. Ushold, *Creating, integrating and maintaingin local and global ontologies*, Prooceedings of the First Workshop on Ontology Learning (OL-2000) in conjunction with the 14th European Conferenc on Artificial Intelligence (ECAI 2000), 2000.

[25] M. Ushold and M. Gruninger, *Ontologies: principles, methods, and applications*, Knowledge Engineering Review **11** (1996), no. 2, 93–155.

[26] E.K. Weick, *The social psychology of organizing*, McGraw-Hill, Inc., 1979.

[27] Cocchiarella, N.B. Formal Ontology. In H. Burkhardt and B. Smith (eds.), *Handbook of Metaphysics and Ontology*. Philosophia Verlag, Munich, 1991.

[28] Cristani, M. and Cuel R. *A comprehensive guideline for building a domain ontology from scratch*, Proceedings of I-Know 2004, IDEA Group Publishing, Graz, 2004.

Semantic Web Basics Column by Dr. Madhu Therani

Therani Madhusudan (Madhu) is an Assistant Professor at the MIS Department, University of Arizona, Tucson, AZ, USA. He holds Ph. D. (1998) and M. S. degrees (1994) in Robotics & Industrial Administration from Carnegie-Mellon University and a B.Tech in Mechanical Engineering (1990) from the Indian Institute of Technology, Madras, India. Prior to joining the University of Arizona in Fall 2000, he was a lead systems architect for Engineering Knowledge Management at Honeywell International, South Bend, IN. His research focuses on the development of knowledge-based tools to support the design and management of complex hardware and software systems. Primary approaches include the development of deep domain models to support the utilization of AI planning, machine learning, case-based reasoning and sequential decision making technologies in implementing robust and useful realworld systems. Specific areas of research include: Information Integration, Intelligent Business Process Management, Product Lifecycle Management and Engineering design automation. He has published over 20 refereed research articles in conferences and journals in these areas.

Semantic Web Basics

Scope: Provide a monthly overview of key issues, new problems, new applications and advances in Semantic Web technology.

Target Audience:

IS researchers in process management, information and data integration, systems analysis & design, SW enggr, Info. Retrieval, data management, KM, econ/IS

Description:

The development of the Semantic Web rests on the convergence of ideas and technologies from multiple areas of research, namely AI and knowledge representation, distributed systems and messaging technologies, data management, process management, software engineering, application domain modeling and finally, the interaction with, acceptance and diffusion of relevant applications in the end-user community.

From an IS research community viewpoint, an entry researcher into this area needs to cover a large background before one can effectively contribute to this growing research area and utilize the applications build thereof in an impactful manner. Currently, multiple research outlets exist and important contributions span both research and industrial publications. Keeping abreast is quite difficult.

The intent of the monthly column is to provide an updated resource for use by the IS research audience (both technical and behavioral), outlining ongoing trends in the individual research areas mentioned above. For each possible IS research area, research questions, solutions, methodologies etc. will be reviewed and summarized from the different research communities mentioned above. Such, a synthesised perspective shall provide a viable platform for further dialogue and research from the community. The overall theme of the column will be to keep the community updated of recent, important progress in the area.

The SemanticWeb - Research Issues from an IS perspective

Madhusudan, Therani
MIS Department,
University of Arizona,
USA 85721
July 24, 2004

The paradigm of the Semantic Web for delivery of applications and services is gaining major momentum in the academic and industrial communities, by exploiting the infrastructure of the World Wide Web. The notion of the Semantic Web cuts across academic research disciplines and technological developments, as indicated in the introductory articles and interviews on the SIG Website [1]. This short introductory column provides an IS researcher's perspective on potential research areas that may fulfill the promise of the Semantic Web, and also how IS research may complement research pursued in other academic areas.

This column lays out the different threads of possible research. Future columns will provide more in-depth discussions, summarize progress, and provide pointers to the latest in these individual areas. The underlying conceptual framework for discussion is provided in Figure 1 (derived from [7]). The figure illustrates three main components of the overall system: a) the real world, b) the Semantic Web with its internal Semantic Web Layer Cake (courtesy of Tim Berners Lee), and c) the users of the Semantic Web. Entities and phenomena from the real world are encoded into the Semantic Web. Hopefully, users will also develop appropriate mental models of the same. Interactions between these three components may occur in one of the following ways: a) Users may interact with the Semantic Web (requesting information or triggering actions) and the Semantic Web responds to the same. The Semantic Web may also initiate interactions with the real world during its processing of the user request. Changes made to the real world or information provided (by the Semantic web) about the real world must be consistent with what is being observed or expected by the users (based on their mental models); b) User's may also interact with the real world directly and the SemanticWeb is updated appropriately (possibly autonomously), so that the Semantic Web may maintain consistent knowledge about the current state of the real world; and c) The Semantic Web may interact with the real world and notify the users appropriately and users may simply observe the effects.

Developing tools, technologies and evaluation methodologies to effectively support the above modes of interaction in variety of domains provide rich avenues for IS research.
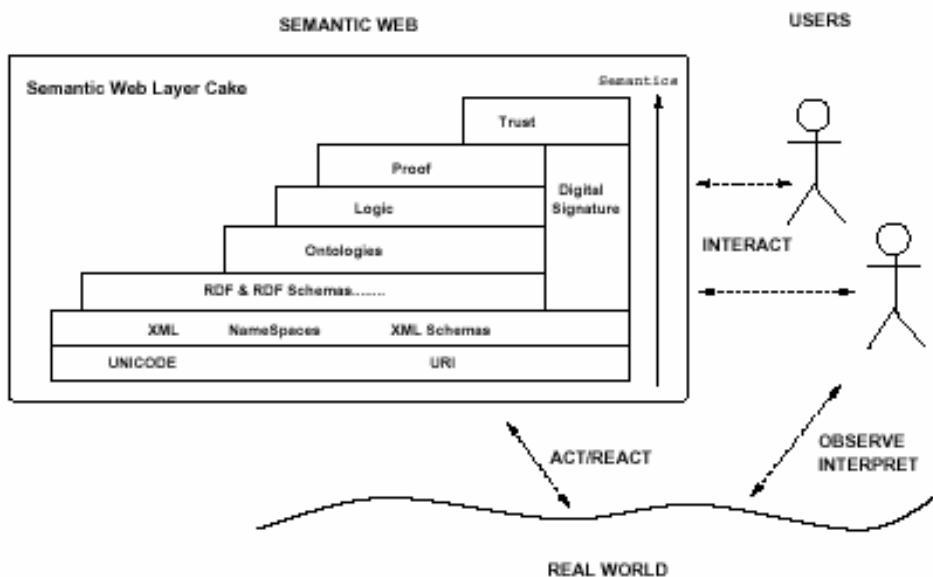

Figure 1: Conceptual Architecture of the Semantic Web

Much of the recent attention has focused on tools for structuring the different layers of the cake. In the following paragraphs, I summarize some of the key avenues (there may be more) of research and exemplars of the same.

Research on the establishing internal structure of the Semantic Web (along the lines of the layered cake) is an active area with involvement from a variety of communities including AI, distributed systems, domainspecific areas etc. Note that the bottom layers of the cake focus on aspects of data modeling whereas the higher layers focus on semantics. The addition of semantics provides rich functionality (both generic and domain-specific). For example, consider the Trust layer. Enabling the same requires models of users, along with models of the domain situated in specific in business contexts, organizational policies etc. Exemplar research on computational issues in developing this internal structure include: a) combining syntax and semantics (including the work on conceptual data modeling, data management) [5], b) ontologies in different application domains (including support for information retrieval and extraction (such as digital libraries)) [3], c) logics of different kinds for different kinds of reasoning [2], d) trust models (both application specific and generic) [4], and e) distributed systems and agent system infrastructures (for efficient effective communication, data and process management) [6]. A key issue from an applied research perspective is modeling domains effectively and developing semantic rules that guide inference at the top layers. Areas of related work in IS is performed in the systems analysis and design, process modeling, data modeling and systems communities.

Managing user interaction with the Semantic Web (and its possible application specific instantiations) is a rich avenue for qualitative and social science research. Work on group systems, online communities, effectiveness of applications, HCI , visualization etc. all have a bearing on the evolution of the Semantic Web. The layering of semantics (the top layers of the cake) supports richer modalities of man-machine interaction (including individuals and teams), thus enabling richer functionality. However, a key aspect of such research also needs to consider issues such as user training and modeling. From an organizational management perspective, implications of adopting Semantic Web technologies within an organization are yet to be understood. Strategic issues include technology adoption and evaluation and managing technology evolution. For example, how does SemanticWeb technology change current approaches towards application integration and systems management? Moreover, what are the implications for intellectual property and knowledge management?

IS researchers are well-positioned to investigate the complete dynamics of the interaction between the three components of Figure 1 in the larger social, economic, technical and cultural context. Methodologies and guidelines for such integrative and applied research need to be developed (respecting the more fundamental areas). In conclusion, the Semantic Web may provide a unifying theme (even for IS research) in the long term. Please send your comments and suggestions to me (madhu@email.arizona.edu). In future columns, I intend to highlight exemplars of research in the wider community along the lines discussed above.

References

1. Special Interest Group on Semantic Web and Information Systems. http://www.sigsemis.org/, 2004.

2. Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2003.

3. J. Davies, D. Fensel, and F. van Harmelen, editors. Towards the Semantic Web:Ontology-driven Knowledge Management. Wiley, 2002.

4. T. Grandison and M. Sloman. A survey of trust in internet applications. IEEE Communication Surveys, (4), 2000.

5. P.F Patel-Schneider and J. Simeon. The Yin/Yang Web: A Unified Model for XML Syntax and RDF Semantics. IEEE Transactions on Knowledge and Data Engineering, 15(4):797–812, 2003.

6. Andrew Tanenbaum and Maarten Van Steen. Distributed Systems: Principles and Paradigms. Prentice- Hall, 2001.

7. W3C. Semantic Web. http://www.w3.org/2001/sw, 2004.

Semantic Web Calendar Column

Forthcoming Special Issues in International Journals

Educational Technology & Society, ISSN 1436-4522
Published by International Forum of Educational Technology & Society
Endorsed by IEEE Learning Technology Task Force

Special Issue (October 2004)

Ontologies and the Semantic Web for E-learning

The *Semantic Web* is the emerging landscape of new web technologies aiming at web-based information and services that would be understandable and reusable by both humans and machines. We argue that *Ontologies*, generally defined as a representation of a shared conceptualisation of a particular domain, is a major component of the Semantic Web. It is anticipated that Ontologies and Semantic Web technologies will influence the next generation of e-learning systems and applications. To this end, key developments such as

- *Formal taxonomies* expressed, e.g., with the help of the web ontology languages RDFS and OWL, and
- *Rules* expressed, e.g., with the help of the web rule language RuleML,

are expected to play a key role in enabling the representation and the dynamic construction of shared and re-usable learning content.

The aim of this special issue is to explore topics related with the new opportunities for e-learning created by the advent of Ontologies and the Semantic Web. We aim at a balanced composition of conceptual, technological and system evaluation work and invite submissions dealing with the following topics:

- Ontologies for e-learning systems
- RDFS/OWL-based educational metadata languages and technologies
- Architectures for ontology-based e-learning systems
- Rules and formal logic for e-learning systems
- Semantic web services for e-learning systems
- Supporting personalized and adaptive e-learning with Semantic Web technologies
- Supporting flexible e-learning systems with Semantic Web technologies
- Innovative Case Studies

Special issue guest editors

Demetrios G Sampson, Dept of Technology in Education and Digital Systems, University of Peiraias, and Informatics and TelematicsInstitute(ITI), Center for Research and Technology - Hellas (CERTH)
Email: sampson@iti.gr, URL: http://www.iti.gr/db.php/en/people/Demetrios_Sampson.html

Paloma Diaz, Computer Science Department, Universidad Carlos III de Madrid, Spain
Email: pdp@inf.uc3m.es or pdp@dei.inf.uc3m.es
URL: http://www.dei.inf.uc3m.es/english/members/pdp.html

Miltiadis D. Lytras, ELTRUN, the E-Business Center, Department of Management Science & Technology, Athens University of Economics and Business, Athens, Greece
Email: mdl@aueb.gr, URL: http://www.eltrun..gr

Gerd Wagner, Faculty of Technology Management, EindhovenUniversity of Technology
Email: G.Wagner@tm.tue.nl, URL: http://tmitwww.tm.tue.nl/staff/gwagner/

Semantic e-Business
Americas Conference on Information Systems (AMCIS)
August 5th – 8th, 2004, New York, USA

The emergence of collaborative processes as an effective means for organizations to deliver their value propositions to their customers, and ultimately to consumers, places an increased onus on organizations to develop systems incorporating emergent technologies. These systems should support the seamless availability of information and knowledge, content and know-how, among partners in the organizations' value chains. Rapidly increasing volume of available information and growing competition in the digital economy are forcing organizations to find efficient ways to gain valuable information and knowledge to improve the efficiency and effectiveness of their business processes.

The realization of representing these knowledge-rich processes is possible through the broad developments in the 'Semantic Web' initiative of the World Wide Web Consortium. But significant amount of research is needed to understand how conceptualizations that comprise business processes can be captured, represented, shared and processed by both human and intelligent agent-based information systems to create transparency in service and supply chains. The developments in on-demand content and business logic availability through technologies such as web-services offer the potential to allow organizations to create content-based and logic or intelligence driven information value chains enabling the needed information transparencies for semantic e-business processes.

Developments on these dimensions are critical to the design of knowledge-based and intelligence driven processes in the digital economy. Research is needed in the development of business models that can take advantage of emergent technologies to support collaborative, knowledge-rich processes in the digital economy. Equally important is the adaptation and assimilation of emergent technologies to enable business processes that contribute to organizations' value propositions. This mini track invites original research contributions that investigate the development of innovative business models to support knowledge-rich business models that enhance collaborations in the digital economy.

<u>Mini Track Chair</u>

Dr. Lakshmi S. Iyer
Information Systems and Operations Management (ISOM) Department
Bryan School of Business and Economics
The University of North Carolina at Greensboro.
Email: lsiyer@uncg.edu
Office Telephone: (336) 334-4984

<u>Co-Chairs</u>
Dr. Rahul Singh & Dr. A. F. Salam
Information Systems and Operations Management (ISOM) Department
Bryan School of Business and Economics
The University of North Carolina at Greensboro, USA
Email: rahul@uncg.edu , amsalam@uncg.edu

NOVEMBER 2004

Third International Semantic Web Conference
(ISWC2004), 7-11 November 2004,
Hiroshima Prince Hotel, Hiroshima, Japan

Conference Web Site: http://iswc2004.semanticweb.org

Organized by

The Japanese Society for Artificial Intelligence and The Semantic Web Science Association

Conference Chair: Frank van Harmelen, Vrije Universiteit Amsterdam, Frank.van.Harmelen@cs.vu.nl

Program Chairs

Sheila McIlraith, Department of Computer Science, University of Toronto, sheila@cs.toronto.edu

Dimitris Plexousakis, Department of Computer Science, University of Crete, And Institute of Computer Science, Foundation for Research and Technology (FORTH), dp@ics.forth.gr

Local Chair

Riichiro Mizoguchi

The Institute of Scientific and Industrial Research, Osaka University, miz@ei.sanken.osaka-u.ac.jp

Important Dates

Submission deadline for Research Track and Industrial Track extended by 2 weeks to Friday April 30, 11:59pm (Hawaii Time). The deadline for submission of abstracts is also Friday April 30, 11:59pm (Hawaii Time).

Semantic Web Research Community: A column dedicated to presentation of Research Groups Worldwide

By Lina Zhou and Gerd Wagner
(July additions)

1. Semantic Web Group at ILRT, UK, http://www.ilrt.bris.ac.uk/projects/semantic_web

The Semantic Web Group at ILRT is primarily interested in transforming mostly human-readable information on the web to there a critical mass of structured data via practical tools, applications and documentation for getting your data on to the Semantic Web. RSS and Calendaring are some of their key application interests, and they have produced tools for storage and query of RDF data.

2. **Semantic Computing Research Group** (SeCo) at University of Helsinki and HIIT, Finland, http://www.cs.helsinki.fi/group/seco/

The focus of SeCo is on machine-processable semantics. They investigate techniques for representing data and knowledge in such a way that machines can "understand" its meaning, and develop algorithmic methods for creating intelligent applications based on such representations.

3. Geospatial ontology research group (OntoGeo) at National Technical University of Athens, Greece, http://ontogeo.ntua.gr/

OntoGeo has focused on the application of ontology and semantics in geography, including spatio-temporal modeling, ontology engineering, semantic interoperability, geographic knowledge representation, and so on.

4. Knowledge-as-Media Research Group (KasM), National Institute of Informatics, Japan, http://www-kasm.nii.ac.jp

The aim of KasM group is to discuss and investigate knowledge sharing issues from various aspects that includes community engineering, ontology engineering, and metadata engineering. Knowledge is considered as a unique media that to interact other people and our environment. Their research investigates interaction among people and develops systems to support such activities.

5. Knowledge Representation Laboratory (KRLAB), Asian Institute of Technology, Thailand, http://kr.cs.ait.ac.th/

The current research in KRLAB focuses on information representation and modeling, the Semantic Web, and software engineering. One of their current researches is XML Semantic Query.

6. China Knowledge Grid (CKG) Research Group, Institute of Computing Technology, Chinese Academy of Sciences, http://kg.ict.ac.cn/

The aim of CKG is to establish a worldwide resources (including knowledge, information, and service) sharing and management model and to develop the corresponding software platform. Their final aim is to establish an intelligent and cooperative platform on the Internet for problem-solving, knowledge management, and decision support. They have proposed a Resources Space Model RSM and related theory and method for the first time.

7. DataBase Systems Lab, Information and Communication University, Korea, http://dblab.icu.ac.kr/
IUC DB Lab carries out a variety of research and development projects. They continue to explore advanced information/knowledge management techniques and apply them to a broad range of applications of the present and future. Some of the projects it has been involved includes: Development of Semantic-aware

Metadata Transformation Engine and Development of Semantic Web based Digital Library System.

8. Semantic Web Laboratory (SemWebLab) at the NRC Institute for Information Technology (NRC-IIT), Canada, http://iit-iti.nrc-cnrc.gc.ca/projects-projets/sem-web-lab-web-sem_e.html
SemWebLab aims to develop Semantic Web tools and applications and to coordinate with similar efforts in Canada and worldwide. At the basic layer, SemWebLab develops ontologies consisting of taxonomies that classify Web objects along with rules, typed by taxonomies, for integrity checking and knowledge inference. SemWebLab also studies agents that use ontologies to support, e.g., the similarity retrieval and composition of learning objects. SemWebLab has a focus on metadata extraction to cope with the vast number of Web objects that are natural language documents.

Key Research Centers (April List)

- The World Wide Web Consortium (W3C), http://www.w3.org/

  The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding.

- W3C SemanticWeb.org, http://www.w3.org/2001/sw/

  The *Semantic Web* provides a common framework that allows *data* to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (*RDF*), which integrates a variety of applications using XML for syntax and URIs for naming.

- Web-Ontology (WebOnt) Working Group, http://www.w3.org/2001/sw/WebOnt/

  The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

- Transatlantic Research Center for the Semantic Web and XML Technologies, http://www.semanticwebcenter.org.uk/

  The Centre provides leading European and American researchers and developers in the area of XML Technologies and the Semantic Web with unique opportunities for effective and flexible transatlantic collaboration aimed at achieving world-class results.

  The Center conducts research into a wide range of emerging leading-edge technologies. Specific research topics are defined in a particular Research Project. Each Project is being curried out by a Research Group, specially formed for this purpose. Every Research Project is aimed at publishing a world-class research monograph or research-based dictionary in order to make the major results of the project available to the world's scientific community.

- Competence Center Semantic Web (CCSW) at DFKI, http://ccsw.dfki.de/

  This site is part of the German research center for artificial intelligence. The focus of the center is on distributed information management with Web-based standardized object representations, ontologies, and rule systems.

- The Information Management Group at University of Manchester, UK, http://img.cs.man.ac.uk/cgi-bin/index.pl?groupsGo=groupsShow&group=semweb&groupsType=Project&strReturn

  The group concerns with Ontologies Knowledge Representation Hypermedia. It uses knowledge representation language to represent conceptual models in machine-amenable formats, while allowing

agents to reason and compute over those models. The group is linked to projects such as OilEd, OntoWeb, WonderWeb, and so on.

- The Knowledge Management Group at **University of Karlsruhe, Institute AIFB**, Karlsruhe, Germany, http://www.aifb.uni-karlsruhe.de/WBS/

  The group has a strong focus on Semantic Web and related areas. Core Semantic Web infrastructure technologies such as Ontobroker, OntoEdit and KAON are developed in collaboration with other groups in Karlsruhe. The group is involved into projects such as SEKT, Knowledge Web, AceMedia, OntoWeb, WonderWeb, SWAP and so on.

- The Knowledge Management Group (WIM) at the **Research Center for Information Technologies (FZI)**, Karlsruhe, Germany, http://www.fzi.de/wim/eng/

  The research group develops techniques and applications for the acquisition, representation & modeling, extraction, storage, access and application of knowledge. A wide range of knowledge intensive systems are based on different core techniques. The group is involved in projects such as DIP, SWWS, KAON, and so on.

- On-To-Knowledge, http://www.ontoknowledge.org/

  On-To-Knowledge-Project aims to develop tools and methods for supporting knowledge management relying on sharable and reusable knowledge ontologies. The technical backbone of On-To-Knowledge is the use of **ontologies** for the various tasks of information integration and mediation.

- **Knowledge Systems Laboratory** at Stanford University, http://www.ksl.stanford.edu/projects/DAML/

  They are developing semantic markup and agent-based technologies to help realize the vision of semantic web. DAML-Enabled Web Services Project had the goal of developing next generation semantic web tools and technology.

- The MINDSWAP Group at the University of Maryland, http://www.mindswap.org/

  It is Maryland Information and Network Dynamics Lab Semantic Web Agents Project. Simple HTML Ontology Extensions (SHOE) is one of its first research projects on Semantic Web. It is also involved with trust and security on the Semantic Web and automatic ontology mapping.

- **eBiquity Research Group** at **University of Maryland, Baltimore County, USA,** http://ebiquity.umbc.edu/v2.1/research/area/id/9/

  The group has been involved with a variety of projects related to the Semantic Web. Among others, Spire, a Personal application for the Semantic Web, explores the use of semantic web technologies in support science in general and the field of ecoinformatics in particular. Securing the Semantic Web investigates distributed trust management as an alternative to traditional authentication and access control schemes in dynamic and pen computing environments such as multiagent systems, web services and pervasive computing. Semantic Discovery focuses on the design, prototyping, and evaluation of a system, called SEMDIS that supports indexing and querying of complex semantic relationships and is driven by notions of information trust and provenance.

- **OntoWeb** , http://ontoweb.aifb.uni-karlsruhe.de/

  Ontoweb is a thematic network funded by the European commission. Its goal is to bring together activities in the area of ontology-based methods and tools for the Semantic Web, bypassing communication bottlenecks between the various and heterogeneous groups of interest.

- Large Scale Distributed Information Systems Lab (LSDIS) at the University of Georgia, http://lsdis.cs.uga.edu/

  The LSDIS lab has extensive research, training, and technology transfer program in the areas of Semantic (Web) technologies. The SemDis project focuses on knowledge discovery and semantic analytics, and have developed a very large populated ontology testbed SWETO for evaluating (million

object and relationship) that is being made available for all non-commercial usage. The METEOR-S project on Semantic Web Processes has researched and is developing tools/systems that utilize semantics in complete Web Service and Web Process lifecycle (annotation, discovery, composition, orchestration/execution). The Bioinformatics for Glycan Expression is applying semantic techniques for integration, analysis and discovery activities in the area of Glycomics, and has developed GLYCO, a comprehensive ontology covering some of the significant areas in the field. Example of commercialization of LSDIS lab's research is Semagix Freedom that has been used to develop semantic web applications for some of the world's biggest companies.

- [Semantic Web enabled Web Services (SWWS) at HP](http://www.hpl.hp.com/semweb/swws.htm), http://www.hpl.hp.com/semweb/swws.htm

  HP Labs Bristol has overall responsibility for two of case studies, which will concentrate on different aspects of procurement, to support developing SWWS platform. SWWS (Web Web Services) is a European 5th Framework project whose goal is to demonstrate how Semantic Web technology can be used to enable an open and flexible approach to web services. More specifically its goals are: 1) provide a comprehensive web services description framework; 2) define a web service discovery framework; and 3) provide a scalable web service mediation platform.

- [Protégé Research Group at the Stanford University](http://protege.semanticweb.org/), http://protege.semanticweb.org/

  Protégé-20000 is on ontology editor and a knowledge-based editor. It provides support for editing Semantic Web ontologies.

Projects Corner: A column dedicated to dissemination of project outcomes

PROJECT ONE: METEOR-S: Semantic Web Services and Processes, provided by Amit Sheth

METEOR-S: Semantic Web Services and Processes (a brief review of current architecture and implementation)

Amit Sheth, John Miller, Kunal Verma, Rohit Aggarwal and the METEOR-S team[12]
Large Scale Distributed Information Systems (LSDIS) lab, the University of Georgia

The METEOR project at the LSDIS Lab, University of Georgia, focused on workflow management techniques for transactional workflows. Its follow on project, which supports Web-based business processes within the context of Service Oriented Architecture (SOA) and the semantic Web technologies and standards, is called METEOR-S [METEOR-S]. Rather than reinvent from scratch, METEOR-S attempts to build upon existing SOA and Semantic Web standards whenever possible (using extensibility features) where appropriate, or seeks to influence existing standards to support and exploit semantics.

A key feature in this project is the usage of semantics for the complete lifecycle of Semantic Web processes, which represent complex interactions between Semantic Web services. The main stages of creating semantic Web processes have been identified as process creation, Web services deployment/annotation, discovery, composition and orchestration. A key research direction of METEOR-S has been in exploring different kinds of semantics, which are present in these stages. We have identified data, functional, Quality of Service and execution semantics as different kinds of semantics and are working on formalizing their definitions [Sheth, 2003; Aggarwal et al., 2004]. Ontologies are the primary mode of expressing and reasoning on the various kinds of semantics. The architecture of the system is shown in Figure 1. A detailed description of the architecture is present in [Aggarwal et al., 2004]. This *brief* article describes the METEOR-S project as designed and implemented so far. Key components of recently completed METEOR-S v0.8 are as follows

1. Abstract Process Designer

In METEOR-S, we define an abstract process as a BPEL4WS (called BPEL) process with semantic annotations. The semantic annotations allow late binding of services to BPEL as opposed to early binding by most other BPEL tools. The abstract process designer allows users to do the following:
1. Create the control flow of the process using BPEL constructs
2. Annotate each call to Web services which require late binding. Alternately, users can bind known services to the process
3. Specifying process constraints/objectives for local and global optimization.

2. Semantic Web Service Developer / Annotator

A basic tenet of Web services is that any service requestor, based on the description in the WSDL files, can invoke them. WSDL (Web Services Description Language) provides information about the service such as the operations present, the expected inputs and outputs for an operation. As also argued by other advocates of semantic Web services (e.g., OWL-S and WSMO projects), METEOR-S also advocates semantic annotation of WSDL. We have pursued semantic extensions to WSDL in two ways-- annotated WSDL 1.1 [Sivashanmugam et al., 2003] and WSDL-S files [Miller et al., 2004]. Annotated WSDL 1.1 is a WSDL document with semantic

---

features added to it via permissible extensibility elements present in the language. METEOR-S contains tools for manual annotation of WSDL or Java source code [Rajasekeran et al., 2004] as well schema matching based approach for semi-automatic annotation of WSDL [Patil, 2004]. WSDL-S (an input to W3C in June 2004 for consideration w.r.t. to the next release of WSDL standard) is based on the upcoming WSDL 2.0 standard and proposed using OWL types for inputs and outputs.

METEOR-S uses the semantic extensions to enhance discovery and dynamic composition. At the same time, as the generated Annotated WSDL 1.1 file adheres to the current industry standard, it can be also be used outside the METEOR-S framework by service requestors unaware of semantics. This flexibility demonstrates the light-weight approach of the methodology used.
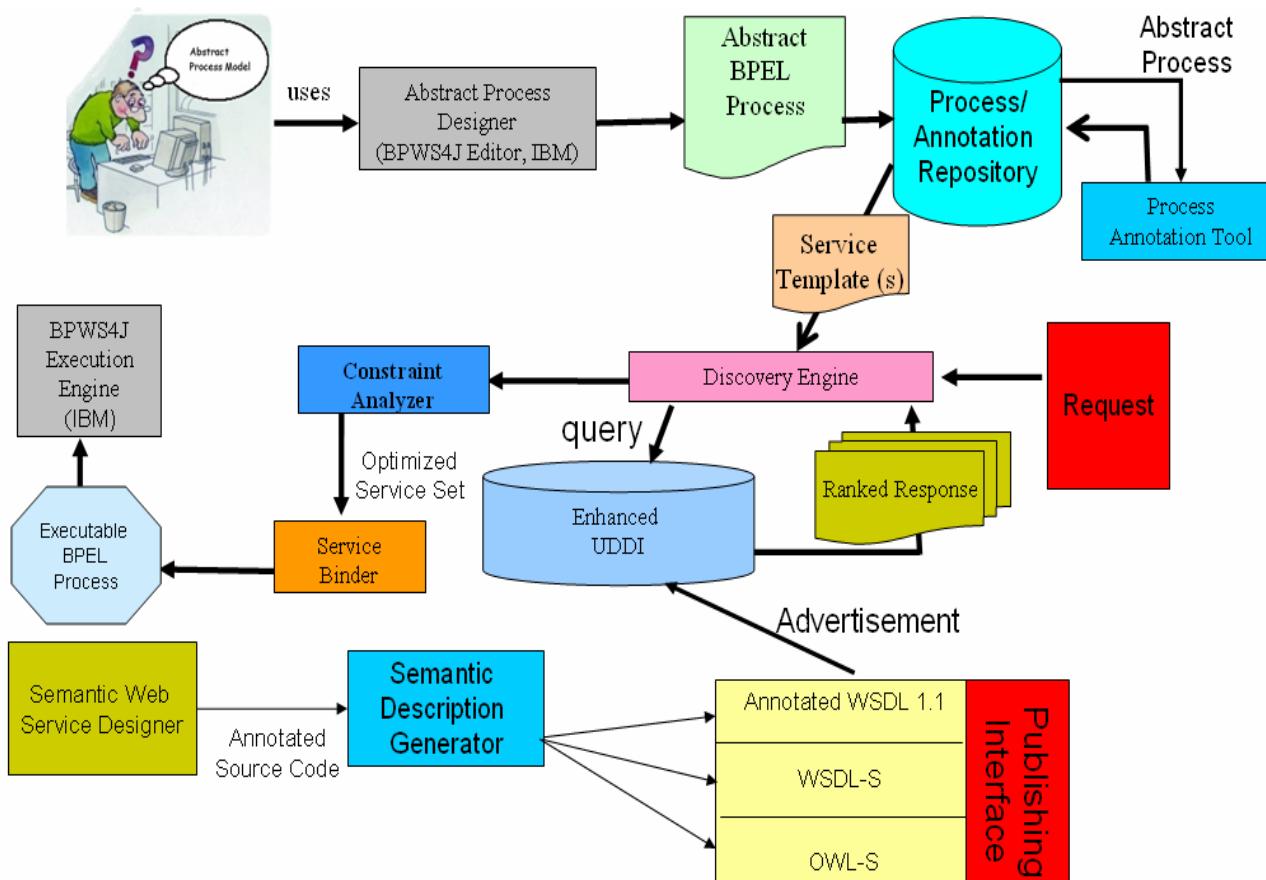

Figure 1: METEOR-S Architecture (v0.8, July 2004)

3. Semantic Publication and Discovery Engine

The discovery engine in METEOR-S is based on adding semantic extensions to UDDI [Verma et al., 2004a]. The ontology based semantic annotations are used to provide semantic matching based on subsumption and property matching. This tool allows users to publish semantically annotated Web services. Users can also use a templates based GUI or the discovery API for querying the engine for matching services.

4. Constraint Analyzer

The constraint analyzer dynamically selects services from candidate services, which are returned by the discovery engine. This selection is done on the basis of global QoS constraints and objectives for the process as well domain constraints. The QoS optimization [Aggarwal et al., 2004] uses an Integer Linear Programming solver [LINDO] and the SWR algorithm [Cardoso et al., 2004]. Domain constraints and inter-service dependencies [Verma et al., 2004b] are handled using an inference engine.

5. Execution Environment

The execution environment consists of a binder and an execution engine. The binder performs actual late binding of services returned by the constraint analyzer and converts abstract BPEL to executable BPEL. IBM's BPWS4J Engine is then used to execute the process.

Release Information
METEOR-S v0.8 (alpha completed July 2004, undergoing testing for beta release August 2004, see demo and download pages of [METEOR-S]) includes tools that allow users to do the following.
- Publish Semantic Web services
- Create Semantic Web processes based on BPEL
- Dynamically discover and bind services to BPEL at design time based on user's criteria (current support for design time binding)
- Aggregate Quality of Service metrics for the entire process
- Optimize selection of services based on user constraints
- Create executable BPEL from semantic definitions and execute it using the BPWS4J Engine.

Other METEOR-S related tools, resources, and contributions include:
- WSDL-S (a proposal to W3C in June 2004 for consideration w.r.t. to the next release of WSDL standard),
- QoS and other inputs to Semantic Web Services Initiative Architecture Committee (SWSA),
- Eclipse Editor for WSDL-S (due early August 2004),
- MWSAF: METEOR-S Web Service Annotation Framework for Semantic Annotation of Web Services (release due late July 2004),
- 50+ semantically annotated Web Services in WSDS1.1 (early July 2004) and in WSDL-S (late July 2004), and more.

REFERENCES

[Aggarwal et al, 2004] Aggarwal, K. Verma, A. Sheth, J. Miller, W. Milnor, Meteor-S Dynamic Composition Environment, To appear in the proceedings of 2004 IEEE International Conference on Services Computing, September 2004.

[Cardoso et al., 2004] J. Cardoso, A. Sheth, J. Miller, J. Arnold, and K. Kochut, Quality of Service for Workflows and Web Service Processes, Journal of Web Semantics, Elsevier, 1 (3), 2004, pp. 281-308.

[LINDO] LINDO API version 2.0, Lindo Systems Inc. http://www.lindo.com/

[METEOR-S] METEOR-S: Semantic Web Services and Processes, http://lsdis.cs.uga.edu/Projects/METEOR-S/

[Miller et al., 2004] J. Miller, K. Verma, P. Rajasekaran, A. Sheth, R. Aggarwal, K. Sivashanmugam, WSDL-S: A Proposal to W3C WSDL 2.0 Committee, LSDIS Lab, June 2004.

[Patil et al., 2004] A. Patil, S. Oundhakar, A. Sheth, K. Verma, METEOR-S Web service Annotation Framework, The proceedings of the 13th International World Wide Conference, (2004).

[Rajasekaran et al, 2004] Enhancing Web Services Description and Discovery to Facilitate Orchestration, Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition. (SWSWPC 2004), July 2004, pages 34-47.

[Sheth, 2003] A. Sheth, "Semantic Web Process Lifecycle: Role of Semantics in Annotation, Discovery, Composition and Orchestration," Invited Talk, WWW 2003 Workshop on E-Services and the Semantic Web, Budapest, Hungary, May 20, (2003).

[Sivashanmugam et al., 2003] K. Sivashanmugam, K. Verma, A. Sheth, J. Miller: Adding Semantics to Web Services Standards, Proceedings of 1st International Conference of Web Services, 395-401, (2003).

[Verma et al., 2004a] K. Verma, K. Sivashanmugam, A. Sheth, A. Patil, S. Oundhakar and J. Miller, METEOR–S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services, Journal of Information Technology and Management (in print), (2004).

[Verma et al., 2004b] K. Verma, R. Akkiraju, R. Goodwin, P. Doshi, J. Lee, On Accommodating Inter Service Dependencies in Web Process Flow Composition, AAAI Spring Symposium PP: 37-43 on Semantic Web Services.

Students Corner: A column dedicated to dissemination of students work concerning Semantic Web

Constraint Driven Web Service Composition in METEOR-S

By Rohit Aggarwal

Creating Web processes using Web service technology gives us the opportunity for selecting new services which best suit our need at the moment. Doing this automatically would require us to quantify our criteria for selection. In addition, there are challenging issues of correctness and optimality. We present a Constraint Driven Web Service Composition tool in METEOR-S, which allows the process designers to bind Web Services to an abstract process, based on business and process constraints and generate an executable process. Our approach is to reduce much of the service composition problem to a constraint satisfaction problem. I have achieved Web service composition based on constraints, starting with an abstract process. I was also able to bind an optimal set of services to the abstract process to create an executable process. This work was done as part of the METEOR-S framework, which aims to support the complete lifecycle of semantic Web processes.

A demonstration and more information about this project are available at http://swp.semanticweb.org/. METEOR-S betav0.8 is scheduled to be released in August 2004.

Rohit Aggarwal is a Master's student at the University of Georgia working in the Large Scale Distributed Information Systems (LSDIS) Lab, directed by Dr. Amit P. Sheth. He is thesis research is co-advised by Drs. John A. Miller and Amit P. Sheth. More information regarding his work, including publications, is available at http://lsdis.cs.uga.edu/~rohit.

Title: A Semantic Web Approach to Intellectual Property Rights Management
By Roberto García
The objective is to make a new contribution to the Intellectual Property Rights (IPR) management research field. There are different initiatives trying to solve the problem of interoperability between Digital Rights Management (DRM) systems. They have started from isolated and proprietary initiatives. However, they are lately clearly moving to a web-broad application domain and thus facing interoperability problems.
There are many harmonisation initiatives but, basically, all have one thing in common, they work at the syntactic level. Their approach is to make a formalisation of some XML DTDs and Schemas that define rights expression languages (REL). In some cases, the semantics of these languages, the meaning of the expressions, are also provided but formalised separately as rights data dictionaries (RDD). Rights dictionaries list terms definitions in natural language, solely for human consumption and not easily automatable.
However, the syntactic approach does not scale well in really wide and open domains like the Internet. An automatic processing of a huge amount of metadata coming from many different sources requires machine understandable semantics. The syntax is not enough when unforeseen expressions are met. Here is where semantics come to help their interpretation to achieve interoperation.
The idea is to facilitate the automation and interoperability of IPR frameworks integrating both parts, the Rights Expression Language and the Rights Data Dictionary. These objectives can be accomplished using ontologies, which provide the required definitions of the rights expression language terms in a machine-readable form. Thus, from the automatic processing point of view, a more complete vision of the application domain is available and more sophisticated processing can be carried out.
The Semantic Web approach has been taken because it is naturally prepared for the Internet domain and thus web ontologies are used. The modularity of web ontologies, constituted by concept and relation definitions openly referenceable as URIs, allows their easy extension and adaptation to meet evolvability and interoperability.

Once the approach is decided, the ontology creation process starts. Starting from previous work on the IPR domain and the analysis of IPR regulations, the IPR core ontology is being developed. It is being founded on upper ontologies in order to check it and facilitate interoperability. All this is building up IPROnto (Intellectual Property Rights ONTOlogy, http://dmag.upf.edu/ontologies/ipronto).
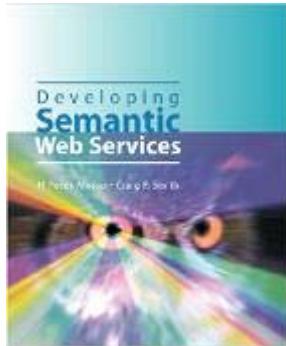
Work continues on and IPROnto is being enriched by top-down and bottom-up processes. For the former by connecting it to different upper ontologies. For the latter by relating it to the DRM standards that are appearing, although they are syntactic ones. Moreover, a test platform based on a Semantic Web portal and agent technologies is being developed, http://dmag.upf.edu/newmars.

Roberto García is a PhD student at the Universitat Pompeu Fabra (UPF), Barcelona, Spain. He is also a Research Assistant at the Technology Department of the UPF. His main research line on IPR and Semantic Web is carried out in the Distributed Multimedia Applications Group. His research interests can be summarised as a "redefing obsession". In other worlds, to try to move things to the Semantic Web in order to interrelate them and see what emerges from the resulting massively connected information space. More details at http://www.tecn.upf.edu/~roberto.

Book Corner: A column dedicated to the presentation of interesting books

Developing Semantic Web Services"

by

H. Peter Alesso
Craig F. Smith
ISBN: 1-56881-212-4

Amazon - Barnes & Noble

The inventor of the World Wide Web, Tim Berners-Lee, is also the originator of the next generation Web architecture, the Semantic Web. Currently, his World Wide Web consortium (W3C) team works to develop, extend, and standardize the Web's markup languages and tools. W3C has developed a new generation of open standard markup languages which are now poised to unleash the power, flexibility and above all - logic – of the next generation of Web, as well as, open the door to the next generation of Web Services.

There are many ways in which the two areas of Web Services and the Semantic Web could interact to lead to the further development of Semantic Web Services. Berners-Lee has suggested that both of these technologies would benefit from integration that would combine the Semantic Web's meaningful content with Web Services' business logic Areas, such as UDDI and WSDL are ideally suited to be implemented using Semantic Web technology. In addition, SOAP could use RDF payloads, remote RDF, and interact with Semantic Web business rules engines, thereby laying the foundation for Semantic Web Services.

Currently, Web Services using the .NET and J2EE frameworks are struggling to expand against the limitations of existing Web architecture and conflicting proprietary standards. With software vendors battling for any advantage, Semantic Web Services, offer a giant leap forward to the first developer to successfully exploit its latent potential to deliver semantic search, e-mail and collaborative work processing.

Developing Semantic Web Services presents the complete Language Pyramid of Web markup languages, including; Resource Description Framework (RDF), Web Ontology Language (OWL) and OWL-Services (OWL-S) along with examples and software demos. In addition, it describes semantic software development tools; including design and analysis methodologies, parsers, validators, editors, development environments and inference engines. The source code for the "Semantic Web Author," an Integrated Development Environment for Semantic Markup Languages is presented and available for download at http://www.web-iq.com

Reviews:

Developing Semantic Web Services is "well-informed about work on WS (Web Services) and the SemWeb (Semantic Web), and in particular ... understand (s) OWL-S ...very well.   ... good job of accurately expressing ... the work. Also, the book ... fill (s) a need that, to my knowledge, hasn't been met at all. "—Dr. David Martin, editor, DAML-S/OWL-S.org.

Semantic Web for Beginners Corner:

Courses taught for SW:

By Amit Sheth

I teach two graduate courses that might be relevant here:

Semantic Web: http://lsdis.cs.uga.edu/SemWebCourse/index.htm

Semantic Web Services and Processes: http://lsdis.cs.uga.edu/SemWebProcess/

Cheers,
Amit

Two brief articles which give an introduction to Ontologies and Semantic Web:

Y. Sure. **Fact Sheet on Semantic Web.** In: KTweb -- Connecting Knowledge Technologies Communities, available at http://www.ktweb.org/doc/Factsheet-SemanticWeb.pdf

Y. Sure. **Fact Sheet on Ontologies.** In: KTweb -- Connecting Knowledge Technologies Communities, available at http://www.ktweb.org/doc/Factsheet-Ontologies-0306.pdf

Job vacancies

JOB DESCRIPTION

Job: Researcher in formal ontologies

Funds: This job opening is one of the workpackages of the project enIRaF - Enhanced Information Retrieval and Filtering for Analytical Systems. Project is funded by EU as Marie Curie Host Fellowship for the Transfer of Knowledge, 6. Framework Programme
Job Description: The goal of the job is to transfer the knowledge on formal ontologies. Successful candidate will work together with local faculty, prepare workshops to teach them the concept of ontologies, and classes on formal models for graduate students. As an outcome, the faculty should be familiar with formal ontologies and choose the model for further implementation of the project enIRaF.

Type of Contract: Temporary, 3 months, full-time

Country: POLAND

City: Poznan, Wielkopolska

Company/Institute: The Poznan University of Economics

Benefits:
- o Working with young, ambitious, open-minded team, having good scientific backgrund and very good communication skills in English.
- o We co-operate with many international organizations and are involved in many EU-funded projects. There is a good opportunity to establish new relations.
- o The Poznan University of Economics is the second best business school in Poland.
- o Dept. of Management Information Systems is organizing international conference on Business Information Systems - BIS (http://bis.kie.ae.poznan.pl). The conference is gaining importance in Europe. There is a good opportunity to make new contacts.

E-Mail: eniraf@kie.ae.poznan.pl
Website: www.kie.ae.poznan.pl

APPLICATION DETAILS

Job Starting Date: 01/10/2004 (later date can be negotiated)

Application Deadline: 20/08/2004

Research area: The actual research sub field should be in the area of ontologies or semantic web

Eligibility:
Required experience: PhD or 4 year of working experience as researcher
Mobility criteria: any nationality except Polish

Salary: typical Marie Curie fellowship, information on request by e-mail

Application: Interested candidates should forward their resume by email to Witold Abramowicz (W.Abramowicz@mis.ae.poznan.pl),who is chair of the Dept. of Management Information Systems
More details about the Human Resources and Mobility Actions can be found at
http://europa.eu.int/mariecurie-actions.
More details about this position can be found on EU Mobility Portal
http://europa.eu.int/eracareers/index_en.cfm?l1=1&l2=1&l3=1&IdJob=2397846

Semantic Web Challenge 2004,
http://challenge.semanticweb.org

By **Ubbo Visser**

People from academia and from industry are invited to submit applications that illustrate the possibilities of the Semantic Web. The applications should integrate, combine, and deduce information from various sources to assist users in performing specific tasks. The submissions should at least satisfy the minimal requirements for a Semantic Web Application and preferably exhibit some of the additional desires.

Although we expect that most applications will use RDF, RDF Schema, and OWL, this is not an official requirement. This is the specific goal for the Semantic Web Challenge 2004: we encourage people to show the benefits of the inference capabilities of the SW-languages used within the application.

What is the Semantic Web Challenge?

How would you explain the Semantic Web to your grandparents? We as scientists working in various research areas pertaining to many aspects of the Semantic Web should be able to give this explanation.
Recently, many interesting issues were presented and discussed at various Semantic Web events (ISWC, IJCAI, AAAI & ECAI workshops). However, these events leave us with the impression that an attractive integrated example of what the Semantic Web can provide does not yet exist. Many of the presentations address only a small aspect of the Semantic Web or are reformulations of other research results.
Although these contributions are positive, the drawback still remains; we as a scientific community cannot yet illustrate what we will deliver to the society in the future. We should come up with these convincing examples. This is why a "Semantic Web Challenge" has been initiated, which will serve several purposes:

- Help us illustrate to society what the Semantic Web can provide
- Give researchers a benchmark to "compare" research results
- Stimulate current research to a higher final goal

What is the goal?

The overall objective of the challenge is to apply "Semantic Web Techniques" in order to build an online application that integrates, combines, and deduces information needed to assist users in performing tasks. The challenge will continue for at least five years and will be updated annually, according to the development of the Semantic Web.

How is the challenge organized?

The challenge does not intentionally define specific data sets because the potential applicability of the Semantic Web is very broad.
Therefore, a number of minimal criteria have been defined which allow people to submit any type of ideas in the form of an application. In addition to the criteria, a number of specific desires have been formulated. The more desires are met by the application, the higher the score will be.

A Semantic Web Challenge Advisory board will define an additional goal every year. The board consist of experts working at universities and in private industry. It will also act as a jury and award the best applications at the ISWC conference.

The members are:

Mike Dean, Stefan Decker , Jérôme Euzenat, Frank van Harmelen, Ian Horrocks, Michel Klein, Nicholas Kushmerick, Deborah McGuiness, Mike Uschold, Ubbo Visser

SIG Board Members

The Following Renowned Academics and practitioners are members of the SIG Board.(Names are displayed in alphabetical order)

Karl Aberer
Distributed Information Systems Laboratory (LSIR) Institute for Core Computing Science (IIF)
School for Computer and Communication Science (I&C)
EPFL Lausanne, Switzerland
E-mail: karl.aberer@epfl.ch
URL: http://lsirwww.epfl.ch/

Richard Benjamins
Intelligent Software Components, S.A.

E-mail: rbenjamins at isoco.com
Web Site: http://www.isoco.com

Christof Bussler
Science Foundation Ireland Professor
Executive Director
Digital Enterprise Research Institute (DERI)
National University of Ireland, Galway
Galway, Ireland

Jesus Contreras
Research Manager
Intelligent Software Components S.A.
Spain

Oscar Corcho
Research Manager
Intelligent Software Components S.A.
Spain

Ming Dong
Assistant Professor
Computer Science Department
College of Science, Wayne State University, USA

E-mail: mdong@cs.wayne.edu
Web Site: www.cs.wayne.edu/~mdong/

Dieter Fensel
Scientific Director of the Digital Enterprise Research Institute (DERI)
Leopold Franzens University of Insburg, Austria
National University of Ireland,
Galway, Ireland

John Davies
Next Generation Web Research Broup
BT Exact, UK
E-mail: john.nj.davies@bt.com

Jorge Marx Gomez
Otto-von-Guericke-Universitat Magdeburg
Faculty of Computer Science
Institute of Technical and Business Information Systems
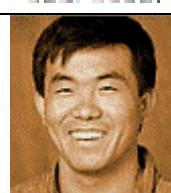Magdeburg, Germany

Farshad Fotouhi
Department of Computer Science
Wayne State University, Detroit, USA
E-mail: fotouhi@cs.wayne.edu
Web Site: www.cs.wayne.edu/fotouhi

William I. Grosky
Department of Computer and Information Science
University of Michigan, USA
Email: wgrosky@umich.edu
Web Site: www.engin.umd.umich.edu/~wgrosky

Asunción Gómez-Pérez
Laboratorio de Inteligencia Artificial (LIA), Facultad de Informatica (FI)
Universidad Politecnica de Madrid (UPM), Spain
E-mail: asun@fi.upm.es
Web Site: delicias.dia.fi.upm.es/miembros/ASUN/asun_CV_Esp.html

Henry M. Kim
Schulich School of Business
York University, Toronto, Canada
E-mail: hkim@schulich.yorku.ca,
Web Site: www.yorku.ca/hmkim

James Hendler
Maryland Information and Network Dynamics Laboratory
Semantic Web and Agents Research,
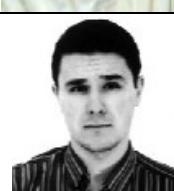University of Maryland at College Park
E-mail: hendler@cs.umd.edu
URL: http://www.cs.umd.edu/~hendler/

Ralf Klischewski
Department for Informatics
University of Hamburg, Germany
E-mail: klischewski@informatik.uni-hamburg.de
Web Site: swt-www.informatik.uni-hamburg.de/people/rk.html

Lakshmi S. Iyer
Information Systems & OperationsManagement
Bryan School of Business and Economics
The University of North Carolina at Greensboro

Miltiadis D. Lytras
ELTRUN, the E-Business Center
Department of Management Science & Technology,
Athens University of Economics and Business, Athens, Greece,
E-mail: mdl@aueb.gr
Web Site: www.eltrun.aueb.gr

Kinshuk
Advanced Learning Technologies Research Centre, Information System Department, Massey
University, New Zealand
Chair, IEEE Learning Technology Task Force
E-mail: kinshuk@massey.ac.nz,
Web Site: infosys.massey.ac.nz/~kinshuk

Ram Ramesh
School of Management
SUNY at Buffalo
E-Mail: rramesh@acsu.buffalo.edu
Web Site: www.mgt.buffalo.edu/CFDOCS/Forms/faculty/bios/faculty.cfm?fac=rramesh

Rajiv Kishore
School of Management
SUNY at Buffalo, USA
E-mail: rkishore@buffalo.edu
Web Site: www.mgt.buffalo.edu/CFDOCS/Forms/faculty/bios/faculty.cfm?fac=rkishore

Demetrios Sampson
Dept of Technology in Education and Digital Systems, University of Piraeus, Greece
Head of Unit, Advanced e-Services for the Knowledge Society (ASK) Research Unit
Informatics and Telematics Institute (ITI)
Center for Research and Technology - Hellas (CERTH)
E-mail: Sampson@iti.gr
Web site: www.ask.iti.gr/

Henrik Legind Larsen
Department of Computer Science
Aalborg University
Esbjerg, Denmark
E-mail: legind@cs.aue.auc.dk
URL: http://www.cs.aue.auc.dk/~legind/

Miguel-Angel Sicilia
Computer Science Department, University of Alcala
Alcala de Henares (Madrid)
Spain

Shiyong Lu
Multimedia Information Systems Group
Department of Computer Science
Wayne State University, USA
E-mail:shiyong@cs.wayne.edu
Web Site: www.cs.wayne.edu/~shiyong/

York Sure
Institut AIFB
Universität Karlsruhe (TH), Karlsruhe Germany
E-mail: sure@aifb.uni-karlsruhe.de
Web Site: http://www.aifb.uni-karlsruhe.de/WBS/ysu/
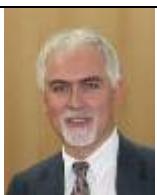
Ambjorn Naeve
KMR Group, Interactive Learning Environments, Centre for user-oriented IT Design (CID)
Department of Numerical Analysis and Computer Science (NADA)
Royal Institute of Technology (KTH) Stockholm, Sweden.
E-mail: amb@nada.kth.se

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

Lisa Neal
Editor-in-Chief of ACM eLearn Magazine
E-mail: lisa@acm.org
URL: http://www.elearnmag.com

Gottfried Vossen
Institut für Wirtschaftsinformatik
Universität Münster, Germany
E-Mail: vossen@uni-muenster.de
Web Site: dbms.uni-muenster.de

Al Salam
Information Systems & Operations Management
Bryan School of Business and Economics
The University of North Carolina at Greensboro

Lina Zhou
Department of Information Systems
University of Maryland, USA
E-mail: zhoul@umbc.edu
Web site: userpages.umbc.edu/~zhoul/

Amit P. Sheth
Large Scale Distr. Info. Sys. Lab
Department of Computer Science
University of Georgia, Athens GA
Email: amit@cs.uga.edu (UGA)
Web Site: http://lsdis.cs.uga.edu/~amit/

Bhavani Thuraisingham
NSF. - Directorate for Computer and Information Science and Engineering
USA
E-mail:bthurais@nsf.gov

Rahul Singh
Information Systems & Operations Management
Bryan School of Business and Economics
The University of North Carolina at Greensboro

Gerd Wagner
Department of Technology Management
Eindhoven University of Technology Eindhoven, The Netherlands
E-mail: G.Wagner@tm.tue.nl
Web Site: is.tm.tue.nl/staff/gwagner

Ubbo Visser
Center for Computing Technologies
University of Bremen
Germany
E-mail: visser@tzi.de
URL: http://www.tzi.de/~visser/

IN THE FORTHCOMING ISSUE (JULY):

- A Call for Papers for the *International Journal on Semantic Web and Information Systems*
- A Call for Contributions for the *Encyclopaedia of Semantic Web Research*
- A special section of LSDIS Lab, University of Georgia
- Announcement of the SIGSEMIS Sponsored Award for the **Best PhD Student Work on SW**

JOIN SIGSEMIS

Support our activities and Join our SIG in Association for Information Systems. A 10$ fee is required and the main reason for this is to sponsor the student awards.

http://www.aisnet.org/sigs.shtml

GET INVOLVED...

❑ **Provide Content** ❑ **Post Articles** ❑ **Become a Columnist** ❑ **Announce Forthcoming Events** ❑ **Ask for guidance** ❑ **Initiate Discussions** ❑ **Browse Newsletter** ❑ **Enter to our Journal** ❑ **Browse Interviews** ❑ **Use E-learning Facility (soon)** ❑ **Find Members Details** ❑ **Learn about SIGSEMIS Sponsored Tracks, Books, Symposiums and Events** ❑ **Apply for the PhD student award (soon)** ❑ **Contribute to our Encyclopaedia of Semantic Web Research** ❑ **Contact SIG SEMIS Board and Members and Committees members**

http://www.sigsemis.org

CONTACT AIS SIGSEMIS NEWS Editorial Board

Please provide any comments, inquiries, ideas, etc to Miltiadis D. Lytras at mdl@aueb.gr

DISTRIBUTE THIS DOCUMENT

Feel free to forward this issue to your colleagues, mailing lists and anyone who may be interested in. Thanking you in advance.

Thank you...

SEE YOU IN ATHENS OLYMPICS!!!