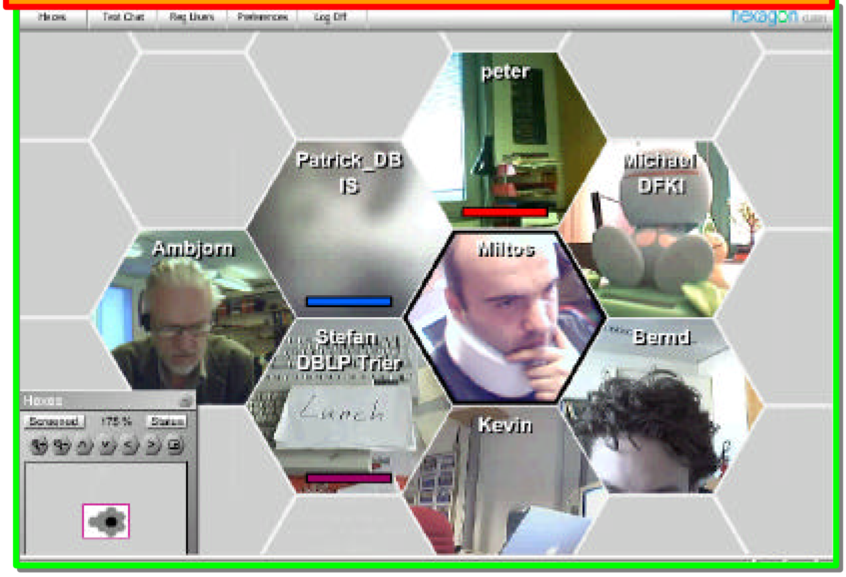
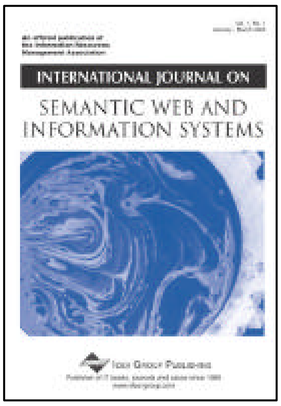
AIS SIGSEMIS
http://www.sigsemis.org
Semantic Web & Information Systems

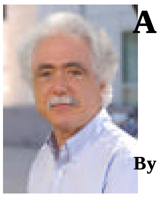
Bulletin
ISSN 1556-2301
Special Interest Group on Semantic Web and Information Systems

Vol. 2, (3&4) 2005
Editor in Chief
Dr. Miltiadis Lytras
Ó AIS SIGSEMIS

Special Issue on
Semantic Web and Social Networks




INTERNATIONAL JOURNAL ON
SEMANTIC WEB AND INFORMATION SYSTEMS



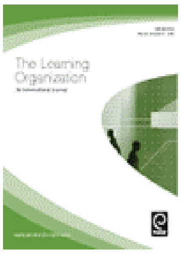
An Interview with Terry A. Winograd

By Dr. Morten Thanning Vendel

OPEN RESEARCH SOCIETY

http://www.open-research-society.org

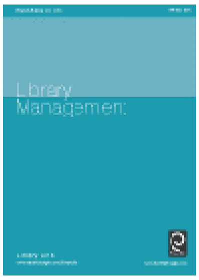
Join today!!


The Learning Organization

Electronic Government

BJET
British Journal of Educational Technology

IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING


Library Management

TECHNOLOGY MANAGEMENT

<mtsr>05

2006 ecis Göteborg

AMCIS 2006 ACAPULCO

AIS SIGSEMIS
http://www.sigsemis.org

Join:http://www.aisnet.org/sigs.shtml

Together we make the Semantic Web a Reality

Table of Contents

Vol 2 **Issue** 3&4

Jan.-Mar. 2005

The Official Quarterly Newsletter of AIS Special Interest Group on Semantic Web and Information Systems

AIS SIGSEMIS Ó 2005

Editorial

Vol 2 **Issue** 3&4

July -Dec 2005

The Official Quarterly Newsletter of AIS Special Interest Group on Semantic Web and Information Systems

AIS SIGSEMIS Ó *2005*

Dear friends,

We are on air again!! We are celebrating the two years of AIS SIGSEMIS Bulletin publication with a double issue. We must say a GREAT THANK YOU for your continuous support and your encouragement to continue our community contribution.

From our last contact several interesting things have happened. You will find a detailed update on the pages of this issue.

I will distinguish from these:

- The Completion of the First Volume of our IJ on Semantic Web and Information Systems (EIC: Prof. Amit Sheth). I will ask your support in building a significant subscription base. With the subscription rate only 55$ I think it is an excellent opportunity to support us.
- The organization of First Online Metadata and Research Conference. Due to its great success we extended it until 7/12/2005. For more info see the relevant section in the Bulletin or contact directly Miguel Angel Sicilia at msicilia@uah.es. A detailed report will be provided in the next issue.
- The preparation of the IEEE TKDE special issue. We were more than happy to get 58 submissions from all over the worlds. Gootfired (Vossen) Nick (Koudas) and me would like to thank all the contibutors for their contributions.

We are trying also to arrange an interview with TBL and many of you helped me to form an interesting and triggering question set. I am waiting for his reflection and lets hope to have a happy end:)

In the following table you can find some interesting statistics for our portal at www.sigsemis.org.

Summary by Month										
Month	**Daily Avg**				**Monthly Totals**					
	Hits	**Files**	**Pages**	**Visits**	**Sites**	**KBytes**	**Visits**	**Pages**	**Files**	**Hits**
Dec 2005	2893	2458	1117	374	727	328678	1499	4471	9832	11575
Nov 2005	2860	2289	885	308	2921	2191430	9267	26554	68688	85819
Oct 2005	3480	2846	1144	356	3380	2921514	11038	35471	88240	107907
Sep 2005	2708	2211	908	315	3079	2962576	9476	27262	66348	81256
Aug 2005	3300	2689	1150	316	3419	3465277	9812	35679	83371	102319

From time to time I receive mails from all over the world for encouraging us keep doing our work in SIGSEMIS. Of course from time to time I get very few mails complaining for the mails we send in various mailing lists for communicating our activities. I respect all these guys and of course i have to apologize to them but I have to say that we will continue with only one objective: To establish AIS SIGSEMIS as the leading Awareness and Learning Center for SW.

Towards this direction I am very happy that the Idea for the establishemtn of a SEMANTIC WEB ACADEMY is in good order as well as our efforts towards the establishment of a European Ecomonic Interest Groupin (EEIG) aiming to promote further the cross-border international collaboration on SW.

From time to time I put "SIGSEMIS" in Google. I would like to thank personal all of you that refer to our SIG and activities in your sites, publications, favorites and teaching materials. In a forthcoming issue i will provie a reflection to your posts. I like also Stephen Downes that is always worried why we dont exploit SW technologies for the publication of our Bulletin. We will do very soon Steve:) And please if any body wants to help drop a mail. We are OPEN!! We didnt developed SIGSEMIS for self-promotion. In this Issue, Leo Sauermann introduces his column on Semantic Desktop Grapewine and we are really happy for this. Leo as well as our other columnist disseminates excellent knowledge. We are planning to merge the contents of the columns in a nice edition available to you before the New Year.

In this issue we initiate our campaing towards the listing of AIS SIGSEMIS bulletin in JCR indexes (In the e-journals sections). We will make this true according to the follwoing criteria:

- On time publication and
- High Quality Short articles aiming to get citations

Both of these criteria can be achieved and this is a personal commitment to our community. We will make AIS SIGSEMIS Bulletin the first Sw-related e-journal listed in JCR indexes in two years time horizon.In this issue Lina Zhou and I worked in putting together a special issue on Semantic Web and Social Networks. The reflection to our invitation from PhD students was great and with the integration of articles from well known academics we hope that you will find this special issue interest. I contacted also Prof. Barabasi (http://www.nd.edu/~alb/) , the Guru of Social Networks and he agreedon an interview tthat we will host in the next issue. My limited time didnt allow me to complete this task on time.

I want to make something clear. We want to help PhD students and for this reason in every forthcoming issue a special section/issue will be devoted to PhDs short (ot RIP) articles. And given our other publication outlets (e.g. IJ on Semantic Web and Information Systems) and our SIGSEMIS sponsored events we can promise to help significantly towards the better quality of their publications targeting to IJSWIS publication.

I must say a big thank you to Amit Sheth for his hard work on IJSWIS. I am more than sure that this hard work is much appreciated by the SW, IS and CS community. We agreed with IGP for an ***Advaced Topics on Semantic Web and Information Systems Topics***. Volume 1 of the Seireis is under development and in late 2006 we will have the publication with chapters from IJSWIS Volume 1 published papers adjusted for teaching and learning purposes. I am sure that you will love the first volume of the series. In this issue there are several interesting pieces of news:

- News from AIS SIGSEMIS activities (Call for papers in special issues, sponsored minitracks and workshops)
- Eleven research articles
- Our regular columns

Dont miss the Interview of Prof. Terry A. Winograd to Dr. Morten Thanning Vendel. Dear Morten MY BEST WISHES AND A GREAT thank you for your kind offer and support.

And also have a look to our new initiative: OPEN RESEARCH SOCIETY (http://www.open-research-society.org). You will LOVE it!!!

Sas Efharisto!!! (Thank you all)

Dr. Miltiadis D. Lytras
Research Academic Computer Technology Institute, http://www.cti.gr,
KMR-Group, http://kmr.nada.kth.se, Email: Lytras@ceid.upatras.gr (use this mail for correspondence)

I am trying for an interview of TBL and asked your help

You proposed me these sample Questions and THANK you very much

1. William Woods wrote "Over time, many people have responded to the need for increased rigor in knowledge representation by turning to FOL as a semantic criterion. This distresses me, since it is already clear that FOL is insufficient to deal with many semantic problems,....". Lotfi Zadeh has similarly talked about the limitations of crisp logic. And Tom Gruber and Amit Sheth have been talking about "informal, semi-formal, formal ontologies", and "implicit, formal and powerful semantics," (e.g., a paper in IJSWIS 1(1) titled "Semantics for the SW: the implicit, the formal and the powerful"). Challenges have been seen not only when modeling NLP problems but also when modeling Life Sciences. However it seems, you are very bullish on DL and want to build at least initial SW on a DL based infrastructure. Is that so? If yes, why? If not, how do you anticipate that KR issues might progress?
2. There is a long standing vision that has focused on named relationships (e.g, Venevar Bush: "***The process of tying two items together is the important thing***.", William Woods: "What's in a link", or Amit Sheth" "Relationships at the heart of semantics"). DL provides reasoning based on sumbsmptive reasoning, but some see that to be highly limited in value and impact as it could not help much with exploiting named relationships, such as those in mining / discovery application that involve computing paths and "connecting the dots". Could you please share your views on computing with a focus on relationships and contexts?
3. What he thinks of Web 2.0. (www.web2con.com) - (The browser Flock, the email Zimbra, Ajax, etc)?
4. Has the Semantic Web lost an enormous opportunity of being part of the so-called Web 2.0 technologies?

5. I would like to hear Sir Tim Berners-Lee comment on is the layered architecture / model of the Semantic Web 5 years after its introduction by him at the XML 2000 conference. Does the model / architecture change given the Current technologies, and how?

6. The WWW has been comparatively much more successful than Artificial Intelligence. I see AI researchers are becoming very active in the Semantic Web. Does that mean the Semantic Web will not be a success (compared to the standards of the WWW)? Or, to put a positive spin on it, can the Semantic Web finally make AI a success?
7. In Sir Berners-Lee's opinion, when will commercial tools help the development of semantic web-based implementation? I.e., when will tools like DreamWeaver or FileMaker Pro support semantic web capabilities?
8. Over what time period do you expect the Semantic Web to enter mainstream technology and become as widely used as the existing web?
9. To what extent will intelligent agents play a role in harnessing the Semantic Web and how would you define an agent within this context?
10. What is the web after the Semantic Web?
11. It seems to be generally agreed that getting knowledge into machines in the Form of ontologies is a good thing. On the other hand, one prominent person in this field recently presented a slide showing a disconnect between ontologies and ROI. (Return on Investment).Can you [Tim] please point out some areas where ontologies have provided ROI, or are expected to do so? Would validating that instance data from several sources conforms to a schema specified in OWL perhaps be one such area?

12. The W3C is starting a Working Group on rule languages for the Semantic Web. What kinds of "rule languages" seem necessary on the SW and why? Are you considering reactive rules or trigger-like rules, i.e. of the kind if event then action, or deduction rules, i.e. rules of the kind if data then new-data?
    Are policies, e.g. for web service negotiations, a "rule issue" or an issue on its own for the W3C? What is new on the SW as far as rule languages are concerned? Will the W3C rule WG develop new rule languages? new rule processors? Is monotonic vs. non-monotonic negation a SW issue? An issue for rule languages on the SW?
13. What are the key challenges for next generation Web search? How will multimedia search on the Web be addressed in the future?
14. Basically Albert-Loszlo Barabasi in Notre Dame Uni argued that WWW follows scale-free network model, which means that Zipf's law applies. My question is would Semantic Web follow the same law?
15. A question I would like to see answered is related to TRUST - a lot has been done in terms of developing the SW infrastructure, better tools, languages and techniques but trust seems to be a bottleneck. Could you ask him to elaborate on this topic - present (federations, Verisign and the like, is it enough?) and Future (do we need new technology? will we ever get there?)
16. I think it's definitely worth touching on Tim's view of the role of the W3C. Is early central standardization relevant to web-centric technologies? One can certainly argue (and I do) that HTML and HTTP succeeded precisely because most of their evolution was conducted in a distributed ad-hoc manner. Beyond Tim's original seed, the W3C contributed primarily after the fact, and its formal definitions are still quite far removed from the realities of web distribution and applications. When the W3C has tried to blaze genuinely new ground, such as with XSchema and SOAP, developers and users tend to prefer simpler ad-hoc solutions. SOAP is being largely ignored in favour of plain HTTP and XML; will the semantic web be built with W3C standards, or with ad-hoc solutions emerging from user / developer communities?
17. "Massively Multiplayer Online Games are among the richest online environments wrt to interactivity and communication between users, how do you think the relationship between the future web and such games will be?"
18. What bothers me is that the assistance to newcomers in RDF / OWL is poorly organized, or rather not organized at all. Yes, we have our forum, but when I ask for solutions to problems the answers usually do not come from anyone of, say, the top-50 experts. There are a few occasional exceptions (Hendler, Manola, Hayes, Horrocks, Ayers, DeRoo) but that is about it (I may have forgotten one or two). Sometime, somewhere I have read a statement on the W3C site about the need to help implementers with their implementation (can't find it back), but in this case it may be a bit better organized. If SW really is required to take off, we better assist the one million-or-so new implementers. Not to be mistaken: I don't want to be a spoiled kid, because what has been accomplished so far is impressive, and free to use. I am grateful for that. My suggestion is to get this organized so that when someone really needs help, he / she gets it from an authoritive person (I'd rather have no response that a faulty one!). Perhaps they can render their services according a roster or so. I have "lived" for years on the forum for XML Schema, and found that one person (Henry Thompson) seemed to feel responsible for the provision of solutions in case nobody reacted..

19. Why W3C is promoting technology which is 30 years behind the leading edge from a simplicity point of view and "square wheel" unreasonable.
20. How come we're not getting there faster?"
21. What's his opinion on the Two Tower / One Tower-Rules quarrel going on? 2. Does he feel like the SemWeb is going in the direction he envisaged?
22. How did his vision change over time?
23. "We know that you believe that the semantic web vision will extend to machine-to-machine services, to enable automatic discovery, composition, invocation, and monitoring of families of web services to support end-user requests. What do you see as the sequence of major innovations and standardization activities that will bring this vision into reality, and how long do you think it will take?"

24. Do you think that dimensions such as (version of resources, country or region of resources, level of trust of resources) which are used in (Learning metadata) will be key elements in the semantic web? Do you think that researches in learning communities are complementary to the SW road map, and resources will finally be like Learning objects? Do you think that the semantic web is moving to a big Geographic Information System in which boundaries will appears again?
25. I would like to ask Sir. Timothy Berners Lee about his opinion on extending modelling of Semantic Web by ordering by relevance, preference, more or less semantically linked resources. This feature is present in Information Retrieval, search engines and also in multicriterial decision (just to mention a few). In his first vision of Semantic Web in the Scientific American paper it is present, but I am missing this in modelling standards recommended and/or developed by W3C.

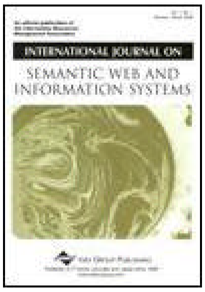
26. I'd be curious to know what he is doing at the moment (management? research? public relationship?) or/and what sort of issue he is in the process of tackling. What he regards as the current and short terms challenges in his capacity as a W3C chairman. So, that is a question centered on his activity. Now, more specifically and less about him, maybe, I would be curious to know whether he has any insight on current and short term researchable issues in the context of the semantic web.

27. Some Questions that are rarely asked of the W3C technical community:
28. Why is the Internet still so English language dependent and centric? One of the greatest advantages of W3C/semantic web technology is to provide the detailed context to make multiple language platforms translations more accurate. For example; the technology exists today to apply English queries to Chinese data domains; and the reverse.
29. Where is the future technical and market leadership to drive this multi language environment? Given the rapid growth of non-native English users and sources; particularly in Chinese, will the Internet split into language centric factions. For example; Chinese and English users now interact largely within only their own separate language data domains.
30. Do defacto standards like Google; with the inherent limitations in content and context understanding, drive the Internet into separate language factions?
31. Given the lack of Google and W3C focus on development of a multi-language architecture; does this mean that we are all doomed to live within our own single language domains.
32. What do you think of the capabilities, and the future, of the MKR language?

UPDATE: I have fwd to Janet Daly, Global Communications Officer, W3C, and waitng:)

AIS SIGSEMIS ACTIVITIES

International Journal on Semantic Web and Information Systems



SPECIAL OFFER: Subscribe to this journal for only **55$ per year**:

CFP: Special Issue on Multimedia Semantics @ IJSWIS

Special Issue Guest Editors:

William Grosky, University of Michigan-Dearborn, wgrosky@umich.edu
Farshad Fotouhi, Wayne State University, fotouhi@wayne.edu

Information is increasingly becoming ubiquitous and all-pervasive, with the World-Wide Web as its primary repository. The rapid growth of information on the Web creates new challenges for information retrieval. Recently, there has been a growing interest in the investigation and development of the next generation web – the *semantic web*. The semantic web enables programs / agents to automatically understand what data is about, and therefore, bridge the, so-called, semantic gap between the ways in which users request web resources and the real needs of those users, ultimately improving the quality of web information retrieval.

Multimedia information has always been part of the semantic web paradigm, but, in general, has been discussed very simplistically by the semantic web community. We believe that, rather than trying to discover a media object's hidden meaning, one should formulate ways of managing media objects so as to help people make more intelligent use of them. The *relationship* between users and media objects should be studied. Media objects should be interpreted relative to the particular goal or point-of-view of a particular user at a particular time. Media objects that would satisfy a user at one time may not satisfy him at other times. And, of course, media objects that would satisfy one user may not satisfy other users, even at the same time.

Content-based descriptors are necessary to this process. Recently, a major European wireless service provider managed to have all its digital media content providers supply metadata in RDF, and saw the revenues increase by 20% in three months. Major search engines are in the process of rolling out A / V search capabilities. At the same time, such descriptions are definitely not sufficient. Context is also important, and should be managed. The area of *emergent multimedia semantics* has been initiated to study the measured interactions between users and media objects, with the ultimate goal of trying to satisfy the user community by providing them with the media objects they require, based on their individual previous media interactions.

For this special issue on the Multimedia Semantic Web, we welcome all papers relevant to topics at the confluence of multimedia information management and the semantic web, such as multimedia ontologies, multimedia extraction and annotation, multimedia semantics, semantics-based search and integration of multimedia and digital content, emergent semantics, semantics enabled multimedia applications (including search, browsing, retrieval, visualization), semantics enabled networks and middleware for multimedia applications, semantic metadata for mobile applications, tool-based approaches utilizing such artifacts as RDF Schema and OWL, approaches using metadata standards such as MPEG-7, and industrial use cases and applications.

Submission Guidelines and Important Dates:

Submitting authors should follow the Style and Author Guidelines for regular IJSWIS papers available at http://www.idea-group.com/ijswis

Manuscripts should be submitted to William Grosky at wgrosky@umich.edu

Submission Deadline for Papers: March 1, 2006
Completion of 1st Round of Reviews: April 1, 2006
Major/Minor Revisions Due: April 15, 2006
Completion of 2nd Round of Reviews: May 1, 2006
Editorial Decisions Sent: May 15, 2006
Planned Publication: Issue 2(3) or 2(4)

Table of Contents & Abstracts of Papers Published on IJSWIS 1(4) 2005

Special Issue:
Semantic Web and Health Care Information Systems Interoperability

The contents of the latest issue of:

The contents of the latest issue of:

International Journal on Semantic Web and Information Systems (IJSWIS)
Official Publication of the Information Resources Management Association
Volume 1, Issue 3, July-September 2005
Published: Quarterly in Print and Electronically
ISSN: 1552-6283
EISSN: 1552-6291

Editor-In-Chief:
Amit Sheth, University of Georgia, USA and Semagix, Inc., USA
Executive Editor:
Miltiadis Lytras, Research Academic Computer Technology Institute and Athens University of Economics and Business, Greece and AIS SIGSEMIS (http://www.sigsemis.org)

Special Issue: Semantic Web and Health Care Information Systems Interoperability

EDITORIAL PREFACE:

Asuman Dogac, Middle East Technical University (METU), Turkey
Vipul Kashyap, Clinical Informatics R&D, Partners HealthCare System, USA

This special issue addresses some of the recent developments in the semantic interoperability in the e-health domain. The first two articles describe the research realized within the scope of a European Commission supported project titled "Artemis: A Semantic Web Service-Based P2P Infrastructure for the Interoperability of Medical Information Systems."

RESEARCH PAPERS

PAPER ONE:

"Archetype-Based Semantic Interoperability of Web Service Messages in the Health Care Domain"

Veli Bicer, Middle East Technical University (METU), Turkey
Ozgur Kilic, Middle East Technical University (METU), Turkey
Asuman Dogac, Middle East Technical University (METU), Turkey
Gokce B. Laleci, Middle East Technical University (METU), Turkey

In this article, the authors describe an infrastructure enabling archetype-based semantic interoperability of Web Service messages exchanged in the health care domain. They annotate the Web Service messages with the OWL representation of the archetypes. Then, by providing the ontology mapping between the archetypes, the authors show that the interoperability of the Web Service message instances can be achieved automatically. An OWL mapping tool, called OWLmt, has been developed for this purpose. OWLmt uses OWL-QL engine, which enables the mapping tool to reason over the source archetype instances while generating the target archetype instances according to the mapping patterns defined through a GUI.

To obtain a copy of the entire article, click on the link below.
http://www.idea-group.com/articles/details.asp?id=5342

PAPER TWO:

"A Distributed Patient Identification Protocol Based on Control Numbers with Semantic Annotation"

Marco Eichelberg, OFFIS, Germany
Thomas Aden, OFFIS, Germany
Wilfried Thoben, OFFIS, Germany

One important problem of information systems in health care is the localization and access to electronic patient records across health care institute boundaries, especially in an international setting. The complexity of the problem is increased by the absence of a globally accepted standard for electronic health care records, the absence of unique patient identifiers in most countries, and the strict data protection requirements that apply to clinical documents. This article describes a protocol that allows the identification of locations of patient records for a given patient and provides access to these records, if granted, under consideration of the legal and technical requirements. The protocol combines cryptographic techniques with semantic annotation and mediation and presents a simple Web-service-based access to clinical documents.

To obtain a copy of the entire article, click on the link below.
http://www.idea-group.com/articles/details.asp?id=5343

PAPER THREE:

"Family History Information Exchange Services Using HL7 Clinical Genomics Standard Specifications"

Amnon Shabo (Shvo), IBM Research Lab, Haifa
Kevin S. Hughes, Massachusetts General Hospital, Partners Health Care, USA

A number of family history applications are in use by health care professionals (e.g., CAGENE, Progeny, Partners Health care Family History Program) as well as by patients (e.g., the US Surgeon General's Family History Program). Each has its own proprietary data format for pedigree drawing and for the maintenance of family history health information. Interoperability between applications is essentially non-existent. To date, disparate family history applications cannot easily exchange patient information. The receiving application should be able to understand the semantics of the incoming family history and enable the user to view and/or to edit it using the receiving applications interface. The authors envision that any family history application will be able to send and receive an individual's family history information using the newly created HL7 Clinical Genomics Specifications through the Semantic Web, using services that will transform one format to the other through the HL7 canonical representation.

To obtain a copy of the entire article, click on the link below.
http://www.idea-group.com/articles/details.asp?id=5344

****
For full copies of the above articles, check for this issue of International Journal on Semantic Web and Information Systems (IJSWIS) in your Institution's library.
****

Note: For only $18.00, purchase an IJSWIS article or any of the 981 single journal articles available electronically by visiting www.idea-group.com/articles.

IJSWIS Regular CFP

CALL FOR PAPERS

International Journal on Semantic Web and Information Systems
Editor-in-Chief: Amit Sheth, Ph.D., University of Georgia, USA and Semagix, Inc., USA
Executive Editor: Miltiadis D. Lytras, Research Academic Computer Technology Institute & Computer Engineering and Informatics Department, University of Patras, Greece

ISSN: 1552-6283, E-ISSN: 1552-6291
Published: Quarterly
Institutional: US $195.00, Individual: US $85.00
Electronic Only: Institutional US $145.00

Key Points:

· Communicates high-quality research findings in the leading edge aspects of Semantic Web and Information Systems convergence

· Discusses the Semantic Web as an indissoluble whole of new generation of technologies, frameworks, concepts and practices for supporting intelligent, innovative, and effective global and networked information systems

· An official publication of the Information Resources Management Association

Complimentary Inaugural Issue http://www.idea-group.com/journals/free.sample.asp?ID=4625

Library Recommendation Form http://www.idea-group.com/recommend.asp?ID=4625

Description:
The **International Journal on Semantic Web and Information Systems** promotes a knowledge transfer channel where academics, practitioners and researchers can discuss, analyze, criticize, synthesize, communicate, elaborate, and simplify the promising vision of the Semantic Web in the context of information systems. **IJSWIS** establishes value-adding knowledge transfer and personal development channels in three distinctive areas: academia, industry, and government.

Submissions:
Interested contributors are asked to submit their manuscripts as an email attachment in Microsoft Word or RTF (Rich Text Format) to mdl@eltrun.gr or lytras@ceid.upatras.gr. Very soon an on-line submission system will be available. The main body of the e-mail message should contain the title of the paper and the names and addresses of all authors. Manuscripts must be in English. The author' s name should not be included anywhere in the manuscript, except on the cover page. Manuscripts must also be accompanied by an abstract of 100-150 words, precisely summarizing the mission and object of the manuscript The publisher will publish the journal in both print and electronic formats.

For more information on how to submit, go to www.idea-group.com/ijswis <http://www.idea-group.com/ijswis>. All submissions and inquiries should be directed to the attention of:

Dr. Miltiadis Lytras
Email: mdl@eltrun.gr or lytras@ceid.upatras.gr

Papers of the following areas are invited:

Full Research Papers
Reviews should focus on the following guidelines when submitting full research papers: The key objective is the presentation of research outcomes and the length should be 4,000-8,000 words. The evaluation factors include 20% theoretical background, 40%significance of propositions, 20% quality of writing, and 20% discussion of implications.

Research Papers Progress
The key objective is to outline interesting future research outlets, while keeping the length from 3,000-3,500. The evaluation factors include 30% theoretical background, 30% methodology outlined, 20% quality of writing, and 20% research problem description.

Case studies
Reviewers should focus on the objective: discussion of real world implementations, while keeping it at the length of 4,000-5,000 words. Evaluation factors include: Research Issues (30%), Promotion of theory & Practice (30%), Discussion of outcomes (20%), and Quality of writing (20%).

Literature Review Papers
When submitting literature review papers, please focus on the main objective: Intensive Critiques of literature / Gaps for possible research. The evaluation factors will include: theoretical background (40%), critical thinking (20%), discussion of gaps in theory (20%), and quality of writing (20%).

Critique of Clusters of SW projects
Please keep the key objective as the evaluation of outcomes when submitting a 5,000-7,000 word critique. The evaluations factors will include: methodologies used (50%), discussion of performance gaps (30%), and the quality of writing (20%).

Vision papers
Please keep the key objective as crafting roadmaps for the future when submitting a 4,000-6,000 word paper. The evaluation factors will include: innovation (50%), theory and technology exploitation (20%), and the quality of writing (20%).

Coverage:

· Semantic Web issues, challenges and implications in each of the IS research streams
· Real world applications towards the development of the knowledge society
· New semantic Web enabled tools for the citizen, learner, organization, business
· Semantic Web enabled business models, ROI matrices and measures, technology effectiveness, case studies, etc.
· Semantic Web enabled information systems, esp. involving ontologies and knowledge bases
· Integration with other disciplines: Semantic Web and Service Oriented Architectures (e.g., Semantic Web Services)
· Standards, Methodologies, Tools, Techniques and Architectures enabling realization of Semantic Web
· Semantics enabled business intelligence, e-services, e-commerce
· Multidisciplinary approaches to realize Semantic Web (e.g, involving Information Retrieval, Linguistics, Knowledge Management, AI, database management, library sciences)

· Beyond Semantic Web, e.g., extending meaning with perception and experience

Associate Editors

Editorial Board

Pls contact Miltiadis D. Lytras at mdl@eltrun.gr or Lytras@ceid.upatras.gr for further inquiries. We would be happy to receive abstracts of potential contributors for guidance.

MTSR' 05 First on-Line conference on Metadata and Semantics Research

[http://www.metadata-semantics.org

KEYNOTE SPEECHES:

Amit Sheth: Semantics Enabled Industrial and Scientific Applications: Research, Technology and Deployed Applications
Tom Gruber, Ontology of Folskonomy

Semantics Enabled Industrial and Scientific Applications: Research, Technology and Deployed Applications

Amit Sheth, LSDIS Lab, The University of Georgia and Semagix

Semantics can already be seen as the key enabler for the new breed of enterprise and scientific applications. I have had the unique vantage point of seeing Semantic Web research transition from academic research at the LSDIS lab to commercialization at Taalee and Semagix, and further on to powering applications serving Enterprise customers and scientific research partners.

In this talk, we will review several ontology-driven applications and information systems. For commercial applications, we will focus on Enterprise applications deployed by Semagix's customers. More specifically, these include *risk management* applications such as Anti-money Laundering and Case Management in Law Enforcement, and *compliance* application such as Security Clearance and Insider Threat Management. For scientific research applications, we will primarily look at clinical health-care and bioinformatics applications. For a Semantic Web platform that can support such applications, we will review the requirements, capabilities and state of the art technologies related to the following:

- Expressiveness of knowledge representations (ontology representation language),
- Development of large populated ontologies that are regularly updated,
- Automatic metadata extraction and annotation involving heterogeneous textual as well as scientific experiment data,
- High-performance and scalable query and rule processing
- Reasoning that computes semantic associations leading to identification or discovery of patterns or interesting/suspicious paths and complex relationships,
- Semantic visualization and semantic virtual interfaces for high-bandwidth user interactions with heterogeneous data, metadata and ontologies, and
- The role of standardards including RDF, RDFS, OWL, SPARQL, SWRL, etc.

We will also review our experiences in building practical ontologies that have involved hundreds of classes to a few million instances/assertions, and the approaches to deal with scalability and performance challenges in building real-world applications.

Background information for this talk can be found at:
Article in Data Engineering special issue on Making the Semantic Web Real (Dec. 2003)
http://wwwt.semagix.com/documents/SemanticWebTechinAction.pdf
Commercial Technology: http://www.semagix.com/download.html
Semantic Discovery Projects at UGA & UMBC: SemDis project
Active Semantic Document with application to Electronic Medical Records:
http://lsdis.cs.uga.edu/projects/asdoc/
Bioinformatics Ontologies and Applications: Glycomics project

Semantics for the Semantic Web: the Implicit, the Formal and the Powerful: http://lsdis.cs.uga.edu/library/download/SRT05-IJ-SW-IS.pdf

Ontology of Folksonomy: A Mash-up of Apples and Oranges

Tom Gruber

tomgruber.org and RealTravel.com

Summary

Ontologies are enabling technology for the Semantic Web. They are a means for people to state what they mean by formal terms used in data that they might generate or consume. Folksonomies are an emergent phenomenon of the social web. They are created as people associate terms with content that they generate or consume. Recently the two ideas have been put into opposition, as if they were right and left poles of a political spectrum. This piece is an attempt to shed some cool light on the subject, and to preview some new work that applies the two ideas together to enable Internet ecology for folksonomies.

Full article available at:
http://tomgruber.org/writing/ontology-of-folksonomy.htm

ECIS 2006: Semantic Web and Information systems Track

http://www.ecis2006.se/02_conferencetracks/semwebis.html

The 14th European Conference on Information Systems will be held in Göteborg, Sweden. It is organised by the IT University of Göteborg and this years theme is "Grand Challenges".

JUNE 9-11, 2006
Doctoral consortium

JUNE 12-14, 2006
ECIS 2006

Semantic Web and Information Systems Track

Track Chairs:

MILTIADIS LYTRAS,	ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS, GREECE
GOTTFRIED VOSSEN,	UNIVERSITY OF MÃœNSTER, GERMANY
AMBJÖRN NAEVE,	ROYAL INSTITUTE OF TECHNOLOGY, SWEDEN

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. Accepted Papers will be invited to a Special Issue in International Journal of Knowledge and Learning (http://www.inderscience.com/ijkl), EIC: Miltiadis Lytras and could be conditionally considered for publication in the International Journal of Semantic Web and Information Systems (http://www.idea-group.com/ijswis), EIC: Amit Sheth.

AMCIS 2006 SIGSEMIS Sponsored Tracks

http://amcis2006.aisnet.org/

Americas Conference on Information Systems (AMCIS) 2006
Acapulco, México August 4-6, 2006

SIG Semantic Web and Information Systems (3 sponsored mini-tracks)

Semantic Web and Information Systems

Miltiadis Lytras - Athens University of Economics and Business
Martin Hepp - University of Innsbruck
Amit Sheth - University of Georgia

AMCIS 2006 Track Proposal
Semantic Web and Information Systems
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.
Accepted Papers will be invited to a Special Issue in the International Journal of Knowledge and Learning (http://www.inderscience.com/ijkl), and extended versions of selected papers could be considered for refereed publication in the International Journal of Semantic Web and Information Systems (http://www.idea-group.com/ijswis).
Suggested topics:

- Semantic Web issues, challenges, and implications in each of the IS research streams
- Semantic Web technology in ERP-centric Information Systems environments
- 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
- Semantic Web services
- Semantic-based knowledge systems
- Semantic Technology-enhanced learning approaches
- Semantic learning organization perspectives
- Towards the development of the knowledge society
- Integration with other disciplines
- Intelligent Systems
- Standards
- Semantic-enabled business intelligence
- Semantic-based Business Process Management
- Enterprise Application Integration

- Metadata-driven (bottom-up) versus ontology-driven (top-down) SW development
- Contribution of semantic technology to the agility of IS evolution

Administrative contact:
Martin Hepp, Digital Enterprise Research Institute,
University of Innsbruck,
e-mail martin.hepp@deri.org, phone +43 512 507 6465

Semantic eBusiness Mini Track Call for Papers

Lakshmi S. Iyer - The University of North Carolina at Greensboro
Rahul Singh - University of North Carolina at Greensboro
A. F. Salam - University of North Carolina at Greensboro

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 eBusiness 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.

Possible topics for papers (theoretical or empirical) submitted to the special issue include but are not limited to:

- models explicating the various forms of knowledge-based processes in the digital economy and their value to the competitiveness of organizations

- the realization of the potential of emergent technologies in supporting knowledge-rich processes required for semantic e-business.

- initiatives for knowledge representation using ontologies and intelligent Agents for semantic processing of cross-enterprise business processes over heterogeneous systems.

- collaborative relationships in the digital economy, enabled and supported by emergent technologies
- the facilitation, initiation, nurturing, development and maintenance of various forms of knowledge-based exchange relationships in the digital environment
- competitive advantage afforded by knowledge-rich processes in semantic e-business
- agents and collaborative systems for intelligent knowledge sharing
- web enabled knowledge transfer and sharing

Keywords: Knowledge Processes, Semantic Web, Information Transparency, Collaborative Commerce, XML, and Ontology.

Authors of accepted papers from the Semantic eBusiness mini-track will be encouraged to submit an improved version of their paper for possible publication to the *International Journal of Semantic Web and Information Systems.*

Mini Track Co-Chairs:

Dr. Lakshmi S. Iyer
Information Systems and Operations Management (ISOM) Department
Bryan School of Business and Economics
The University of North Carolina at Greensboro.
Greensboro, NC 27402, USA
Email: lsiyer@uncg.edu; Office Telephone: (336) 334-4984

Dr. Rahul Singh
Information Systems and Operations Management (ISOM) Department
Bryan School of Business and Economics
The University of North Carolina at Greensboro.
Greensboro, NC 27402, USA
Email: rahul@uncg.edu; Office Telephone: (336) 256-0260; Fax: (336) 334-4083

Dr. A. F. Salam
Information Systems and Operations Management (ISOM) Department
Bryan School of Business and Economics
The University of North Carolina at Greensboro.
Greensboro, NC 27402, USA
Email: amsalam@uncg.edu; Office Telephone: (336) 334-4991

Social Intelligence in Information Systems

Mini Track Chair: Linda Zhou - University of Maryland

Social intelligence in information systems is enabled by the Internet and more recent Semantic Web. As more and more information and services are connected via the Internet, the potential of social intelligence becomes increasingly recognized. Social intelligence presents excellent opportunities for research and development of the Semantic Web. Yet, many challenging issues related to social intelligence in terms of fundamental technologies and impacts on individual and organizations remain to be addressed.

The objectives of the mini-track is to explore technologies for social intelligence, examine the impacts of social intelligence, and improve the ways of applying social intelligence to enhance the efficiency and/or effectiveness of individual and organizational activities. We welcome empirical research that applies quantitative or qualitative methodologies.

The topic areas that are of interest to this mini-track is listed but not limited to the following:

- Social intelligence technologies such as social network analysis and visualization
- Theories and models for improving the understanding of social intelligence
- Semantic representation for social intelligence
- Applications of social intelligence technologies
- Impact of social intelligence on humans
- Impact of social intelligence on physical or virtual organizations
- Best practice of social intelligence
- Culture issues in adopting social intelligence
- Success factors for applying social intelligence

Detailed guidelines for paper submission are available online at http://amcis2006.aisnet.org/.
Questions regarding to the mini-track can be directed to Lina Zhou at zhoul@umbc.edu

Update on AIS SIGSEMIS Sponsored IEEE Transactions on Knowledge and Data Engineering

"Knowledge and Data Engineering in the Semantic Web Era" Special Issue Early 2007

GUEST EDITORS

Gottfried Vossen
University of Muenster, Germany
Email: vossen@helios.uni-muenster.de

Miltiadis Lytras
Research Academic Computer Technology Institute, Hellas and AIS SIGSEMIS
Email: Lytras@ceid.upatras.gr

Nick Koudas
AT&T Labs Research, USA
Email: koudas@research.att.com

We are very happy for the great number of submissions. We received 58 papers and we are running an intensive review process. We will let you know on further developments.

No	Title	Author(s)
1.	A Case-based Model for Ontology Reuse on the Semantic Web	Yuxin Mao , Zhaohui Wu, Huajun Chen, Mengya Tang
2.	A Component Model and Infrastructure for a Fluid Web	André Santanchè, Claudia Bauzer Medeiros
3.	A Context-based Approach for Semantic Inter-Portlet Communication in Enterprise Knowledge Portals	Thorsten Priebe
4.	A Flexible Ontology Reasoning Architecture for the Semantic Web	Jeff Z Pan
5.	A Multilevel Approach to Semantic Annotation for Interoperability	Federica Schiappelli, Michel Missikoff Nunzia Osimi, Francesco Taglino
6.	A Probabilistic Ontology Algebra	Octavian Udrea, Yu Deng, Edward Hung, V.S. Subrahmanian, Nazif Cihan Tas
7.	A Relation-Based Search Engine in Semantic Web	Yufei Li, Yuan Wang, Xiaotao Huang
8.	A Requirement Driven Framework for Benchmarking Semantic Web Knowledge Base Systems	Yuanbo Guo, Abir Qasem, Zhengxian Pan, Jeff Heflin
9.	A Semantic Web Based Approach toKnowledge Management for Grid Applications	Liming Chen, Carole Goble, Nigel Shadbolt
10	A Semantic Web Usage Mining Approach for Discovering Periodic-based Personal Web Access Patterns	Baoyao Zhou, Siu Cheung Hui, Alvis C.M. Fong
11	A Tool for Empirical Evaluation of Semantic Similarity in Ontologies	Valerie Veltri Cross, Youbo Wang
12	AlViz: Alignment Visualization for Ontology Management	Monika Lanzenberger; Jennifer Sampson
13	An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval	Pablo Castells, Miriam Fernàndez David Vallet
14	An Algebra for Ontology Composition	Prasenjit Mitra, Gio Wiederhold
15	Automatic Asymmetric Merging of Ontologies	L. Venkata Subramaniam, Amit A Nanavati, Sougata Mukherjea
16	Bottom-up Extraction and Trust-based Refinement of Ontology Metadata	Ernesto Damiani, Paolo Ceravolo, Marco Viviani
17	Building RuleML Learning Object Ontologies on the Semantic Web	Yevgen Biletsky, Harold Boley, David Hirtle
18	Building trusted personal Webs: the myGnosiWeb(s) approach	Konstantinos I Kotis
19	Constrained-based Ontology Induction from Online Customer Reviews	Thomas Lee
20	Contented-based Routing for Semantic Web	Haifeng Liu, Milenko Petrovic, Hans-Arno Jacobsen
21	CRCTOL: A Semantic Based Domain Ontology Learning System	Xing Jiang, Ah-Hwee Tan
22	Defining Apparent Ontologies, and a Method for their Evaluation	Charles Turnitsa, Andreas Tolk
23	DR-Prolog: A System for Reasoning with Rules and Ontologies on the Semantic Web	Grigoris Antoniou, Antonis Bikakis

24	Evaluation of Ontological Representations of Learning Resources on the Semantic Web	Miguel Angel Sicilia
25	Formalization of data-intensive P2P systems	Zoran Majkic
26	From Wrapping to Knowledge	Rafael Corchuelo, José Arjona David Ruiz-Cortés
27	Improving Real-World Semantics in OWL	Palash Bera, Anna Krasnoperova Yair Wand
28	Intention-Driven Query Tool for Semantic Web Services Systems	Chiung-Hon Lee, Alan Liu
29	Interoperability support for between the MPEG-7 MDS and OWL in the DS-MIRF Framework	Chrisa Tsinaraki, Panagiotis Polydoros Christodoulakis
30	Knowledge & Data Reasoning in Semantic Web	Jin Song Dong , Yuzhang Feng, Yuan Fang Li, Hai Wang
31	Knowledge Flow Management Supporting Complex Problem Solving: Learning Spectrum and Its Infrastructure	Wanchum Dou
32	Logical Deficiencies in XML	Han Reichgelt, Adrian Gardiner, Vladan Jovanovic
33	Making the Semantic Web Accessible to the Casual User: Emirical Evidence on the Usefulness of Semiformal Query Languages	Abraham Bernstein, Esther Kaufmann
34	Managing Change in Large Scale Data Sharing Systems	Peter Mork, Steven D. Gribble, Alon Y. Halevy
35	Media Transformation via Semantic Information Processing	Kaoru Sumi, Mizue Nagata, Tanaka Katsumi
36	Mining Generalized Associations of Semantic Relations for Knowledge Discovery in Textual Web Contents	Jiang Tao, Tan Ah-Hwee
37	Modeling in the Era of the Semantic Web	Andrew W. Crapo, William A. Wallace, Thomas R. Willemann
38	Ontia iJADE: An Intelligent ontology-based Agent Framework for Semantic Web Service	Toby H.W. Lam, Tony W..A. Ao Ieong, Raymond S.T. Lee
39	Ontology Based QoS-Aware Service Discovery and Measurement	Chen Zhou, Liang-Tien Chia, Bu-Sung Lee
40	Ontology Engineering from a Database Perspective	Bodo Hüsemann, Gottfried Vossen
41	Ontology Extraction Based on Interpretations of Table Sturctures	Masahiro Tanaka, Ishida Toru
42	Planning as Model Checking for Semantic Web Agents	Paul Benjamin Lowry; James V. Hansen, Bonnie Brinton Anderson
43	Ranking Gossip for P2P Information Dissemination and an Implication: Networking with Measuring Space and Semantic Space	Hai Zhuge, Xiang Li, Xue Chen
44	RDFS(FA): Connecting RDF(S) and OWL DL	Jeff Z Pan, Ian Harrocks
45	Semantic Annotation Based on Extraction Ontologies	Yihong Ding, David W. Embley, Stephen W. Liddle
46	Semantic Content Management Systems	Kilian Stoffel, Ciorascu Claudia, Ciorascu Iulian
47	Semantic Integration of Tree-Structured Data Using Partially Specified Tree-Pattern Queries	Dimitri Theodoratos, Theodore Dalamagas, Antonis Koufopoulos
48	Semantic Knowledge Integration for eBusiness Processes: An Ontological Analysis	Rahul Singh, Lakshmi S Iyer, Fergle Daubeterre
49	Semantic Similarity Methods for Information Retrieval in Medical Information Systems and the Web	Giannis Varelas, Angelos Hliaoutakis, Epimenidis Voutsakis, Euripides G.M Pektrakis, Paraskevi Raftopoulou, Evangelos Milios
50	Semantic Web Personalization	Magdalini Eirinaki, Michaelis Vazirgiannis
51	Spatial Integrity Maintenance for Geographic Ontologies on the Semantic Web	Alia I Abdelmoty, Phil D Smart Baher A El-Geresy Chris B. Jones
52	Supporting scientific Collaboration in a Network of Excellence through a semantically indexed Knowledge Map	Paola Velardi, Allessandro Cucchiarelli, Michael Pétit
53	Temporal RDF	Claudio Gutierrez, Carlos Hurtado, Alejandro Vaisman
54	The Design and Implementation of a Multi-Agent Based Intelligent Context-Aware News Reporting System –iV iJADE News Reporter	Eddie Chan, Tony Ao Ieong, Raymond Lee
55	The m^2 Distance Metric for Learning	Guihua Wen, Lijun Jiang, Nigel R. Shadbolt
56	Unsupervised Semantic Annotation of Glossaries using Regular Expressions: an Experiment in the Cultural Heritage Domain	Pierluigi D'Amadio, Roberto Navigli Paola Velardi
57	Using Semantic Web Technologies to Enable Interoperability of Disparate Information Systems	Marwan Sabbouh, Joseph K. Derosa Susan A. Powers, Scott R. Bennett
58	What to Ask to a Peer: Query Reformulation	Diego Calvanese, Guiseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati

Special Issue on
Digital Libraries in the Knowledge Era: Knowledge Management and Semantic Web Technologies

At Library Management Journal Vol 26(4/5) 2005

Articles available at:
http://www.emeraldinsight.com/Insight/viewContainer.do?containerType=Issue&containerId=22447

Editor in Chief: Steve O' Connor
Chief Executive Officer
Caval Collaborative Solutions
4 Park Drive
BUNDOORA. Vic 3083

Special Issue Guest Editors
Miltiadis Lytras, Research Academic Computer Technology Institute, and AIS SIGSEMIS
Miguel-Angel Sicilia, University of Alcalá, Spain
John Davies, BTExact Next Generation Web Research, UK
Vipul Kashyap, Partners HealthCare System, Clinical Informatics R&D, USA

Special Issue on
Semantic and Social Aspects of Learning in Organizations

At http://thesius.emeraldinsight.com/vl=1232840/cl=68/nw=1/rpsv/tlo.htm
(Official Publication date: 29/7/2005)

Managing Editor
Nancy Rolph
E-mail: nrolph@emeraldinsight.com

Editor
Colin Coulson-Thomas
Professor of Direction and Leadership,
University of Lincoln,
UK
All correspondence to:
Mill Reach, 12 Mill Lane, Water Newton, Cambridgeshire, UK, PE8 6LY
E-mail: colinct@tiscali.co.uk

Articles Available at:

http://www.emeraldinsight.com/Insight/viewContainer.do?containerType=Issue&containerId=22676

Articles

The semantic learning organization
Miguel-Ángel Sicilia, Miltiadis D. Lytras (pp. 402-410)
Keywords: Computer based learning, Internet, Learning organizations
ArticleType:Conceptual paper
View HTML | View PDF (61 KB)

Semantic networks and social networks
Stephen Downes (pp. 411-417)
Keywords: Information networks, Internet, Social networks
ArticleType:Conceptual paper
View HTML | View PDF (57 KB)

Social networking on the semantic web
Tim Finin, Li Ding, Lina Zhou, Anupam Joshi (pp. 418-435)
Keywords: Information networks, Internet, Social networks
ArticleType:Research paper
View HTML | View PDF (586 KB)

Western and Eastern views on social networks
Patricia Ordóñez de Pablos (pp. 436-456)
Keywords: Eastern hemisphere, Knowledge management systems, Social networks, Western hemisphere
ArticleType:Research paper
View HTML | View PDF (123 KB)

Aspect-oriented re-engineering of e-learning courseware
Victor Pankratius, Wolffried Stucky, Gottfried Vossen (pp. 457-470)
Keywords: Computer based learning, Learning
ArticleType:Technical paper
View HTML | View PDF (307 KB)

An ontological representation of learning objects and learning designs as codified knowledge
Salvador Sánchez-Alonso, Dirk Frosch-Wilke (pp. 471-479)
Keywords: Computer based learning, Knowledge management, Learning
ArticleType:Conceptual paper
View HTML | View PDF (696 KB)

Combining ontologies and peer-to-peer technologies for inter-organizational knowledge management
Heiner Stuckenschmidt, Wolf Siberski, Wolfgang Nejdl (pp. 480-491)
Keywords: Information networks, Internet, Knowledge management, Learning organizations
ArticleType:Conceptual paper
View HTML | View PDF (273 KB)

Strategic industrial alliances in paper industry: XML- vs Ontology-based integration platforms
Anton Naumenko, Sergiy Nikitin, Vagan Terziyan, Andriy Zharko (pp. 492-514)
Keywords: Communication technologies, Internet, Strategic alliances
ArticleType:Research paper
View HTML | View PDF (445 KB)

(Official Publication date: Spring 2006)

Special Issue on

Exploiting Knowledge Management for Ubiquitous E-Government in the Semantic Web Era.

Electronic Government, an International Journal

ISSN (Online): 1740-7508 - ISSN (Paper): 1740-7494

Editor in Chief : Prof. Binshan Lin, Dept of Management & Marketing, College of Business Administration, *Louisiana State University, USA*

Special Issue Editors

Dr. Miltiadis Lytras, Lakshmi Iyer, Athanassios Tsakalidis

Editorial: Exploiting Knowledge Management for Ubiquitous E-Government in the Semantic Web Era

Miltiadis D. Lytras

Department of Management Science and Technology, ELTRUN – The Research Center, Athens University of Economics and Business, 47A Evelpidon Str., 113 62 Athens, GREECE
E-mail: mdl@eltrun.gr

Lakshmi S. Iyer,

Information Systems & Operations Management, Bryan School of Business and Economics, The University of North Carolina at Greensboro, USA, Email: lsiyer@uncg.edu

Athanasios Tsakalidis,

Department of Computer Engineering and Informatics, University of Patras & Research Academic Computer Technology Institute, Greece, Email: tsak@cti.gr

In the last years the semantic web and knowledge management technologies have provided a new context for transforming policies and social priorities to systems, services and tools aiming to foster the capacity of citizens to pursue a unique fulfilment of personal needs.

In the W3C Semantic Web Activity, a number of critical milestones have been set as the key priorities for the promotion of human centric information systems, even though most people consider SW, as a technical oriented issue. Objectives like information flow and collaborative life, ontological evolution, Creating a Policy Aware Infrastructure and Web of Trust require a multi-fold approach that brings together research communities like computer science, semantic web, information systems, social science etc.

Our AIS SIGSEMIS (http://www.sigsemis.org) - SIG on Semantic Web and Information systems in the Association for Information Systems, has a key objective and a unique characteristic. To provide how significant changes can be enhanced from the application of the Semantic Web to specific contexts that are related to critical social objectives and emerging information systems.

This special issue is timely since the exploitation of the Semantic Web and Knowledge Management Technologies for E-government is a hot topic in the agenda of most governments worldwide.

With an acceptance rate of about 20% and with a key concern to include research works from all over the world we do believe that the final outcome is of high quality and addresses all the various aspects of the key theme.

We tried through a specific editorial strategy to combine articles in three general directions:

- Introduction to the convergence of the Knowledge Management and the Semantic Web Technologies for E-Government
- Demonstration, through Technical Research papers, of the capacity of the Semantic Web to support new approaches to key E-government research issues
- Discussion of social aspects that affect the effectiveness of Semantic E-Government applications

In the first paper Dr. Miltiadis Lytras, Officer of the AIS SIGSEMIS, provides a rich picture on "*The Semantic Electronic Government: Knowledge Management for Citizen Relationship and new assessment scenarios*". Citizen Relationship Management is a Knowledge Intensive task which requires an in depth analysis of knowledge infrastructures, knowledge flows and dynamic transformations. The emphasis on semantic web provides an introduction for the readers of the E-Government Journal.

In the second article entitled "*Knowledge Management for Government-to-Government (G2G) Process Coordination*", Lakshmi S. Iyer, Rahul Singh, Al F. Salam and Fergle D'Aubeterre from the University of North Carolina at Greensboro, based on the foundations of Semantic Web, including ontologies, knowledge representation, multi-agent systems and web-services; Knowledge Management (KM); and G2G processes, they present a vision for knowledge management for G2G process coordination.

Christian Wagner, Karen Cheung and Rachael Ip, from the Department of Information Systems, City University of Hong Kong and Stefan Böttcher from the Department of Computer Science, University of Paderborn (Germany) in "*Building Semantic Webs for E-government with Wiki Technology* propose the design of a two-layer semantic wiki web, which consists of a content wiki, largely identical to the traditional web, and a semantic layer, also maintained within the wiki, that describes semantic relationships.

Wajee Teswanich, Chutiporn Anutariya and Vilas Wuwongse, from Thailand in the fourth paper: "A Knowledge Management System Framework for Governmental Regulating Processes" propose a framework for the modelling and management of as well as reasoning with four types of knowledge (*terminological* (or *ontological*), *factual*, *empirical* and *regulatory)* by means of prominent Semantic Web languages, namely RDF and OWL.

Ljiljana Stojanovic, Nenad Stojanovic and Dimitris Apostolou in the fifth paper entitled "*Change Management in eGovernment: OntoGov Case Study*", they discussing on methods aiming to improve the usability of eGovernment services. Particularly they focus on new methods for the semantic service annotation as well as for semantic service discovery.

Jennifer Blechar, Ioanna D. Constantiou and Jan Damsgaard contributed the sixth paper on *Understanding Behavioural Patterns of Advanced Mobile Service Users* Through an exploratory analysis of the results of a field study on advanced mobile service use in Denmark, this paper suggests that pricing of services and consumers references to already established service delivery platforms can be important elements influencing consumers' and citizens behaviours.

The last short article of the special issue by Saggi Nevo and Henry Kim from Canada is entitled *"How to compare and analyse risks of Internet voting versus other modes of voting"* challenges reader to explore further the theme of the special issue on specific contexts like I-voting.

We do believe that in the next five years Semantic Web will affect critically the design, implementation and support of E-Government services. This special issue is only the beginning of an exciting journey. Join us!!

Acknowledgments

Our deepest appreciation and respect to Professor BLIN, Editor in Chief of E-Government journal who gave us the opportunity to serve our community, and for his continuous commitment and contribution to a mutual vision. We wish him health, prosperity, creativity and well-being.

We would like to thank the renowned academics and practitioners who contributed the articles to this special issue. Their knowledge, expertise, imagination and inspiration are evident in every line of this edition. We are sure that this book was only the first part of a long-term collaboration. Last but not least we are grateful to the members of AIS SIGSEMIS who served as reviewers.

Special Issue on

Advances of Semantic Web for E-learning: Expanding learning frontiers.

British Journal of Educational Technology

Edited by: Nick Rushby , **Print ISSN:** 0007-1013, **Online ISSN:** 1467-8535

URL: http://www.blackwellpublishing.com/journal.asp?ref=0007-1013

Special Issue Editors

Ambjorn Naeve, ***Miltiadis Lytras***, ***Wolfgang Nejdl***, ***Joseph Hardin, Nicolas Balacheff,***

UPDATE:

We completed the Special Issue preparation and the contents of the special issue will be forwarded to the EIC of BJET on 10/12/2005 with a publication schedule mid 2006.

Special issue on Knowledge Management within the Health, Pharmaceutical and Clinical Sectors: Towards patient-centric health care systems

International Journal of Technology Management

Edited by: Mohammed Dorgham, **Print ISSN:** 0267-5730, **Online ISSN:** 1741-5276

URL: http://www.inderscience.com/ijtm

Special Issue Editors

Miltiadis Lytras, Research Academic Computer Technology Institute, ELTRUN, AUEB, Greece and AIS SIGSEMIS, Email: Lytras@ceid.upatras.gr

Ambjorn Naeve, Head of the Knowledge Management Research group, Computer Science and Communication (NADA), Royal Institute of Technology (KTH), Sweden, Email: amb@nada.kth.se

Constantin Makropoulos, Director - Division of Applied Technologies
National Centre for Scientific Research (NCSR) "Demokritos", Greece
Email: cmakr@dat.demokritos.gr

Vipul Kashyap, Clinical Knowledge Management, Partners HealthCare Systems, Clinical Informatics R&D, USA, Email: vkashyap1@partners.org

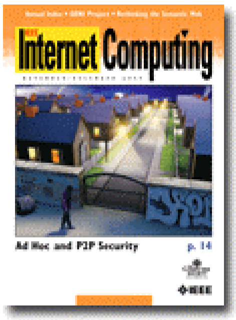
Important Dates

Send manuscripts to Lytras@ceid.upatras.gr and cc: amb@nada.kth.se

30th November 2005	Submission of manuscripts
15th February 2006	Notification to authors
15th June 2006	Final versions due
Late 2006	Publication

Update:

Submission of papars completed and we are in tha review process. We will let you know on further developments.



http://www.computer.org/portal/site/internet/

Editor in Chief : Robert E. Filman

About IEEE Internet Computing

IEEE Internet Computing targets the technical and scientific Internet user communities as well as designers and developers of Internet-based applications and enabling technologies. *IC* publishes refereed articles on the latest developments and key trends in Internet technologies and applications.

A crossroads between academic researchers and software professionals, the magazine presents novel content from academic and industry experts on a wide range of topics, including architectures, data mining, middleware, security, standards, and more.

IC's content reaches more than 7,000 subscribers internationally, comprising leading researchers, developers and engineers (76% industry, 24% government/academia).

An AIS SIGSEMIS proposal for a special issue on SW.

Dear All,
I would like to share with you an excellent piece of news: We are discussing with Editor in Chief of IEEE IC, Prof. Robert Filman, an AIS SIGSEMIS Sponsored special issue on **Semantic Knowledge Management**.

I am more than happy that **Prof. Amit Sheth** (IEEE Fellow, LSDIS Leader and distinguished personality of SW community) as well as **Prof. Thomas Davenport** (Distinguished Consultand, Professor and Director of Research, Babson Executive Education, Babson College, KM Guru) accepted my invitation to suppport this proposal and form the guest editorial team in case of acceptence. We will let you know on this in next issues. I would like personally to thank EIC Prof. Filman who accepted our invitation to serve on the Advisory Board of AIS SIGSEMIS.

Sincerely, Miltiadis Lytras

Journal of Web Engineering

http://www.rintonpress.com/journals/jwe/index.html

The Journal of Web Engineering (JWE) aims to provide a forum for advancing the scientific state of knowledge in all areas of Web Engineering. Original articles, survey articles, reviews, tutorials, perspectives, and general correspondence are all welcome. Appropriate submissions should address significant issues and problems, and potential solutions, and will be reviewed in accordance with peer review conventions.

An AIS SIGSEMIS proposal for a special issue on JWE.

Dear All,
In our concern that SIGSEMIS must lead the dissemination, awareness and education aspects of Semantic Web in various communities we are discussing with EICs of JWE, Martin Gaedke and David Lowe for further collaboration and joint publications.

Martin and David accepted our invitation to serve on our SIGSEMIS Advisory board and thank you for this.
Sincerely,

Miltiadis Lytras

RESEARCH PAPERS IN THIS ISSUE

Special issue on Semantic Web and Social Netowrks

TABLE OF CONTENTS

Social Networks and the Semantic Web

Prof. Lina Zhou
Department of Information Systems, University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore,Maryland 21250
zhoul@umbc.edu
http://userpages.umbc.edu/~zhoul/

Reference to this article should be made as follows:
Zhou L. (2005), Social Networks and the Semantic Web, AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 34-35

A social network is an explicit representation of a set of social relationships such as friendship, professional peers, or information exchange, between a group of people (or organizations or other social entities) [1, 2]. A social network can be visualized as a graph with nodes representing people and links representing social relationships. The real power of social networks is demonstrated in the spread of the relationships. Social network analysis can discover knowledge (or pattern) from the links between people within networks. Social networks have been used to examine organizational structure, member reputation, marketing effect, terrorist network, disease spreading, online dating, and so on.

Social networks are rooted in the concept of small word effect discovered about 40 years ago [3]. There is renewed wide interest in both research and practice in social networks recently. This time, the traditional sense of physical social network has been extended to online social networks enabled by the Web, which is in turn attributed to the penetration of the Internet technology. In this case, people who are connected in the network may have never met each other in person or they could use online social networks to enhance the physical social networks that have already existed.

The semantic Web allows resources on the Web to be processed automatically by bringing structure to the meaningful content on the Web [4]. Ontologies and knowledge representation are among the enabling technologies of the semantic Web. One of the first widespread applications of RDF [5], a standard semantic Web language, is the representation of social networks [2].

The semantic Web brings new opportunities and challenges to social networks. They can be illustrated with the following perspectives:

- Information representation. Personal profiles have been widely used to share information about oneself with others in a social network. The profile information is explicitly elicited from members during the registration process. Different social networks tend to represent personal profiles differently. Due to information heterogeneity, members who share similar interests but who are in different social network would not able to link to each other. Semantic web offers a promising solution to the heterogeneity problem by providing ontologies (a set of well agreed concept, properties, relationships, axioms) such as FOAF to represent personal profiles.
- Trust relationship. Trust is an important issue in human-to-human as well as human-to-computer (information systems) interactions. The issue can also be extended to indirect human-to-human relationship created via computers. There are two opposing views on one property of trust relationship in social networks: transitive and intransitive. For example, if person A trusts person B and personal B trusts person C, we can derive that A trust C based on the transitive view and A may not trust C based on the intransitive view. The semantic Web offers a potential solution to the above problem by enhancing reputation management with the capability of semantics-based reasoning. As a result, individuals can decide whether an indirectly linked person in the network is trustworthy or not.
- Individual privacy. Social networks bring up some interesting social issues, especially in the online environment. Members of a social network are committed to publicizing their personal profiles and

social relationships to other members. The personal information is subject to misuse by other members. The semantic Web is potentially useful to address such a concern by specifying privacy policies with ontologies in sharing personal information. Based on the privacy policy, personal information can be classified based on different levels of privacy and shared with other members with the correponding strength of ties.

- Network of network. It is likely that one person belongs to more than one social network. The person plays an important role in social networks by connecting or transcending the boundaries of different communities. As social networks currently stand, different social networks are relatively isolated. The semantic Web provides an opportunity to create a network of networks by using ontologies and ontology mapping to bridge different social networks.

People were used as nodes of social networks in the above discussion. Similar issues apply to social networks consisting of other types of nodes such as organizations. In addition, the relationship between two nodes can be multi-dimensional, which serves as the basis of different social networks. The social network mechanism could lead to a better understanding of the social structure in different applications. The semantic Web provides a machine-friendly infrastructure for online social networks. Ultimately, agents can not only support relationship building in human socirty by analyzing social networks but also become part of the social networks.

[1] L. Garton, C. Haythornthwaite, and B. Wellman, "Studying Online Social Networks," *JCMC*, vol. 3, 1997.

[2] T. Finin, L. Ding, L. Zhou, and J. A., "Social networking on the semantic Web," *The Learning Organization*, vol. 12, pp. 418-435, 2005.

[3] S. Milgram, "The Small World Problem," *Psychology Today*, pp. 60 - 67, 1967.

[4] T. Berners_Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, vol. 284, pp. 34-43, 2001.

[5] W3C, "RDF Primer," http://www.w3.org/TR/rdf-primer/ .

Semantic Social Networks for Email Filtering:

A Prototype and Analysis

Jennifer Golbeck
University of Maryland, College Park
College Park, Maryland
golbeck@cs.umd.edu

Reference to this article should be made as follows:
Golbeck J., (2005), Semantic Social Networks for Email Filtering: A Prototype and Analysis,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 36-40

1 Introduction

The fact that spam email has become such a ubiquitous problem has lead to much research and development of algorithms that try to identify spam and prevent it from reaching the user's mailbox. Many of those techniques have been highly successful, catching and filtering a majority of spam messages that a person receives.

Though work still continues to refine these methods, some focus has shifted to new mechanisms for blocking unwanted messages and highlighting good, or valid, messages. "Whitelist" filters are one of these methods. In these systems, users create a list of approved addresses from which they will accept messages. Any messages from whitelisted senders are delivered to the user's inbox. In pure-whitelist systems, all other messages are filtered into a low-priority folder. These systems do not claim that all of the filtered messages will be spam, but rather that a whitelist makes the inbox more usable by only showing messages that are definitely not spam. Though whitelists are nearly 100% effective at blocking unwanted email and delivering only messages from known people, there are two major problems cited with them. Firstly, there is an extra burden placed on the user to maintain a whitelist, and secondly, valid emails will almost certainly be filtered into the low-priority mailbox. If that box contains a lot of spam, the valid messages will be especially difficult to find. Even in less strict whitelist systems, where the whitelist is used to prevent valid messages from being marked as spam, and all non-spam is delivered to the inbox, it can be difficult to sort important messages from unimportant ones.

Semantic Web-based social networks offer an interesting alternative to whitelists. We propose using the social network data accessible from on the Semantic Web as a sort of social whitelist. We augment basic social connections with trust ratings, to enhance sorting and filtering messages according to the trustworthiness of the sender. The prototype email client, TrustMail, is presented here, along with an analysis of its potential efficacy based on the Enron email corpus.

2 TrustMail

2.1 System Introduction

The Friend-of-a-Friend (FOAF) project is one of the largest initiatives on the Semantic Web, with the personal and social networking information of well over 8,000,000 people. In the approach to email filtering we will present here, the entire FOAF social network is used as a sort of whitelist. Users maintain lists of connections to people they know. That connects them into a larger network, creating paths to a large number of other, unknown users. By following these paths in the network, the social whitelist includes a much larger number of people, with a relatively small burden on the user.

However, it is often the case that people have social connections to people they do not trust. If these social paths are used to filter email, it is useful to have some assurance that we trust someone to send us relevant, worthwhile messages; the fact that the person solicited a social connection online is often not enough to ensure they are trustworthy in this respect. To incorporate trust information into the FOAF social network, we have developed the FOAF Trust Module[1], an ontology that extends FOAF and allows people to rate how

[1] http://trust.mindswap.org/trustOnt.shtml

much they trust people either in general, or about a specific topic. The ontology does not set a scale for making trust ratings, but in this work we use a scale of 1(low trust) to 10 (high trust). Currently, our crawl of FOAF files shows over 2,000 unique users who have assigned trust ratings to people they know. Figure 1 depicts a visualization of this network.

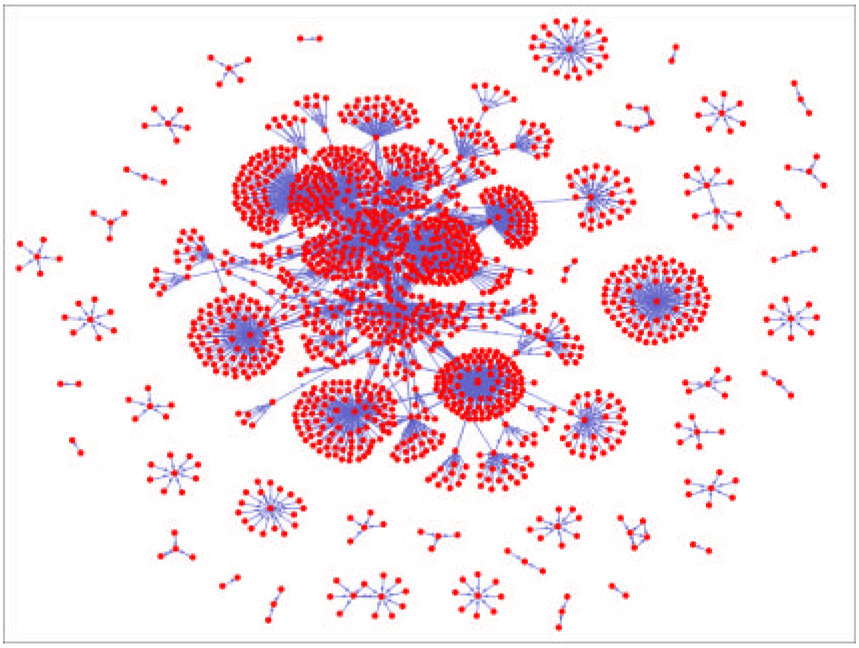


Figure 1. A visualization of the social network formed by users with trust ratings.

Users directly assign trust ratings to people they know. In order to give the user insight into the trustworthiness of people in the network who are *unknown*, we have developed an algorithm to recommend how much to trust an unknown person. Our algorithm, TidalTrust (Golbeck, 2005), looks at paths in the network from the user to the unknown person, and composes the trust values along those paths to create a recommendation.

If there are two people (say Alice and Bob) in the network who are not directly connected to one another, how can they know how much to trust each other? The TidalTrust algorithm finds paths connecting the individuals in the network, and computes a recommended trust rating based on the trust ratings assigned along the way. For example, if Alice wanted to know how much to trust Bob, the algorithm would look at all of Alice's friends, and ask for their trust rating of Bob. Alice's recommended trust rating for Bob is computed as the average of Alice's friends ratings, weighted by Alice's trust for those friends. This algorithm is applied recursively – if Alice's friends do not have a direct trust rating for Bob, they would poll their friends, compute the weighted average, and return that to Alice. Further details about the TidalTrust algorithm can be found in (Golbeck, 2005).

When integrated into email, these trust ratings are used as a score for messages. When a message is received, the email client computes the recipient's recommended trust rating for the sender. This rating is displayed next to each message; messages from more trusted people are shown with higher scores, while messages from untrustworthy people will receive lower scores. This system makes the inbox more usable by making “good” messages prominent. It is important to note that scores will appear next to messages from people with whom the user has never has contact before since, if they are connected through a path of mutual acquaintances in the reputation network, a rating can be inferred.

It is important to note that the goal of this scoring system is not to give low ratings to bad senders. We do not strive to identify spam. Rather, the main premise is to provide *higher* ratings to *non-spam* senders, so users are able to identify messages of interest that they might not otherwise have recognized.

Because of this focus, this algorithm is not intended to be a solution to spam by itself; rather, it is a technique for use in conjunction with a variety of other anti-spam mechanisms. There are some spam issues that particularly effect this algorithm. Forged email headers, where the "From:" line of a message is altered to look like a valid address is one such issue. This work is not designed to address this problem, and some other technique must deal with forged headers. Because this technique is designed to identify good messages that make it past spam filters, it also do not address the case where a person has a virus sending messages from their account. Other spam-detection techniques will be required to flag these messages.

2.2 Related Work

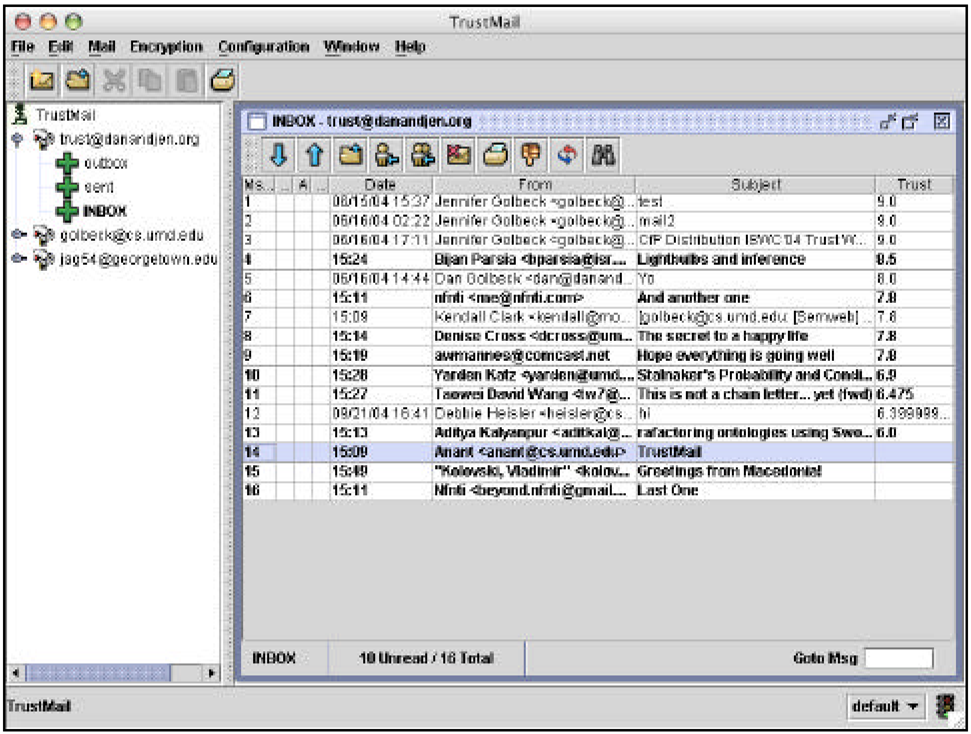
Other approaches have used social networks for message filtering. In Boykin and Roychowdhury (2004) create a social network from the messages that a user has received. Using the structural properties of social networks, particularly the propensity for local clustering, messages are identified as spam, valid, or unknown based on clustering thresholds. Their method is able to classify about 50% of a user's email into the spam or valid categories, leaving 50% to be filtered by other techniques.

Our approach takes some of the basic premises of whitelisting and social network based filtering and extends them. Unlike Boykin and Roychowdhury's technique that builds a social network from the user's own email folders, the trust-based technique uses a network that connects users.

In our application, trust is integrated into the email client to serve as a tool for ranking and filtering messages according to their presumed importance. This is not the first technique developed for this task. Maxims (Lasharki et al., 1994) is an agent integrated with an email client that learns how to filter, delete, and archive messages based on user actions. While my work takes a social network-based approach to the problem of message filtering instead of an agent-based approach, the two methods are not contradictory; they could, in fact, be integrated into a system as complementary in the task of easing email overload.

2.3 The TrustMail Application

TrustMail is a prototype email client that adds trust ratings to the folder views of a message. This allows a user to see their trust rating for each individual, and sort messages accordingly. Potentially important messages can be highlighted, even if the user does not know the sender. Because the algorithm that computes trust values uses the user's perspective on the trust network, the scores are personalized for each user. The social network depicted in Figure 1 is used as the basis for making the trust computations shown next to each message in Figure 2.



Ms...	...	A	...	Date	From	Subject	Trust
1				06/15/04 15:37	Jennifer Golbeck <golbeck@...	test	9.0
2				06/16/04 02:22	Jennifer Golbeck <golbeck@...	mail2	9.0
3				06/16/04 17:11	Jennifer Golbeck <golbeck@...	CfP Distribution ISWC'04 Trust W...	9.0
4				15:24	Bijan Parsia <bparsia@isr....	Lightbulbs and inference	8.5
5				06/16/04 14:44	Dan Golbeck <dan@danand...	Yo	8.0
6				15:11	nfrdi <me@nfrdi.com>	And another one	7.8
7				15:09	Kendall Clark <kendall@mo...	[golbeck@cs.umd.edu: [Semweb] ...	7.8
8				15:14	Denise Cross <dcross@um...	The secret to a happy life	7.8
9				15:19	awmannes@comcast.net	Hope everything is going well	7.8
10				15:28	Yarden Katz <yarden@umd....	Stalnaker's Probability and Condi...	6.9
11				15:27	Taowei David Wang <tw7@...	This is not a chain letter... yet (fwd)	6.475
12				09/21/04 16:41	Debbie Heisler <heisler@cs...	hi	6.399999...
13				15:13	Aditya Kalyanpur <aditkal@...	rafactoring ontologies using Swo...	6.0
14				15:09	Anant <anant@cs.umd.edu>	TrustMail	
15				15:49	"Kolovski, Vladimir" <kolov...	Greetings from Macedonia!	
16				15:11	Nfrdi <beyond.nfrdi@gmail...	Last One	

Figure 2: The TrustMail Interface. In this window, messages are sorted according to the trust rating of the sender, with the most trusted appearing highest in the list.

The ratings alongside messages are useful, not only for their value, but because they basically replicate the way trust relationships work in social settings. For example, today, it would sensible and polite for a student emailing a professor she has never met to start the email with some indication of the relationships between the student and the two professors, e.g., “My advisor has collaborated with you on this topic in the past and she suggested I contact you.” The professor may chose to verify the validity of this statement by contacting the student's advisor or finding information that verifies the claim.

TrustMail replaces the process of searching for information about a recipient by utilizing the data in web-based social networks. Because calculations are made from the perspective of the email recipient, high trust ratings will have necessarily have come through people the recipient trusts. This allows the trust network-based system to complement spam filters by identifying good messages that might otherwise be indistinguishable from unwanted messages, and carrying the validation of a rating drawn from the user's own network of trusted acquaintances.

One question that is important to consider is how many messages will actually be scored using the TrustMail system. The next section uses a real email corpus to gain some insight into this question.

3 Case Study: The Enron Email Corpus

To gain some insight into how TrustMail may impact a user's mailbox, a large network with many users is required. Although the trust network shown in Figure 1 had about two-thousand members, it is not ideal for this type of analysis because it only connects a small community of users, and thus it would only be possible to analyze the mailboxes of a few users. The ultimate application of TrustMail would involve a much larger network or a better connected community. Since this type of social network with trust ratings was not available to test TrustMail, it had to be generated from other existing data.

The Enron email dataset[2] (Klimt, Yang, 2004) is a collection the mail folders of 150 Enron employees, and it contains over 1.5 million messages, both sent and received. There are over 6,000 unique senders in the corpus, and over 28,000 unique recipients. These numbers are much greater than the number of users whose mailboxes have been collected because they represent everyone who has sent a message to the users, everyone who has been cc-ed on a message to the users, and everyone the users have emailed. The collection was made available by the Federal Energy Regulatory Commission in late 2003 in response to the legal investigation of the company. Because the messages represent a single community, they are ideal for analyzing the potential of TrustMail.

To create a social network for analysis, each message in the corpus was read. A social networking connection was added from the *sender* to each of the *recipients*. This produced an initial social network, although the connections are weak. To be more sure that the links between people represented a relationship and not just a one-time message, we revised the network to only include social connections when the sender had emailed the recipient at least twice.

The resulting social network is obviously lacking trust values. While the strength of relationships could be derived from the corpus of messages, this measure would not correlate directly to trust as we have considered it. Instead of creating trust data, we ignore that component in this analysis. The question we are seeking to answer by looking at this email corpus is, assuming a trust network existed, how many messages will actually receive scores in the TrustMail interface.

From the corpus, we first extracted a list of all individuals who sent mail to a given user. Then, the social network was searched for a path from the recipient to each sender. These calculations allow us to determine what percentage of senders could be given trust ratings if there were actually a trust network supporting the Enron users.

An analysis of the Enron network showed the following statistics:

- 37% of recipients had direct connections to people who sent them email in the social network; in other words, 37% of the time the recipient had emailed the sender of a message.
- 55% of senders who were not directly connected to the recipient could be reached through paths in the social network.
- Thus, a total of 92% of all senders can be rated if trust values were present in the social network

These numbers indicate that users in a community similar to Enron, an application like TrustMail would be able provide information about a vast majority of the incoming messages. While the Enron corpus is a close community of users, we believe that if users are properly supported in making trust ratings as part of their email client, a similarly high percentage of senders and messages would receive ratings.

[2] http://www.cs.cmu.edu/~enron/

4 Conclusion

In this paper, we described TrustMail, an email client that uses a trust network and algorithms for inferring trust to score each email message. Users are able to sort their mail folders according to the trustworthiness of the sender. An analysis of the structure of the Enron corpus showed that users can expect a majority of their messages to receive ratings if a well-connected trust network is available to support the application.

While the Enron email corpus represents a natural community of users, social networks on the Semantic Web have great potential to serve as an effective network for filtering messages. The most important factor in their success will be how frequently they are updated and maintained. This behavior is promoted by applications, like TrustMail, that take advantage of the network. When users feel they are receiving a benefit from using the trust ratings next to email messages, this encourages them to create more and more accurate trust ratings of people they know. This, in turn, improves the accuracy of the inferred ratings and the number of messages that can be rated. If this cycle is encouraged with a user interface that allows users to easily make trust ratings, trust rated email may become an effective method of email filtering.

References

Boykin, P. O., V. Roychowdhury, (2004) Personal email networks: an effective anti-spam tool. Preprint, <http://www.arxiv.org/abs/cond-mat/0402143>.

Golbeck, Jennifer. (2005) "Computing and Applying Trust in Web-Based Social Networks," Ph.D. Dissertation, University of Maryland, College Park.

Klimt, B., Y. Yang. (2004) Introducing the Enron Corpus, *Proceedings of the First Conference on Email and Anti-Spam*, Mountain View, California.

Lasharki, Y., M. Metral,, and P. Maes. (1994) Collaborative interface agents, Proceedings of the National Conference on Artificial Intelligence, MIT Press, Cambridge, MA.

Applying Mined Social Networks to Knowledge Management Scenarios

Victoria Uren[1], Jianhan Zhu[1] and Alexandre L. Gonçalves[2]

[1]Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK,
{j.zhu, v.s.uren}@open.ac.uk
[2]Stela Institute, Florianópolis, Brazil, {a.l.goncalves}@stela.org.br

Reference to this article should be made as follows:
Uren V., Zhu J., Goncalves A., (2005), Applying Mined Social Networks to Knowledge Management Scenarios,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 41-44

1. Introduction

The study of social networks is very rich. Mathematicians produce compelling insights into the structure and development of social groups by modelling them as graphs, e.g. (Watts 1999). Author co-citation analysis is one of the core methods of bibliometrics (White & McCain 1989), in which citation behaviour is used to map the intellectual links between individuals and hence to study the structure of scientific literatures. In the semantic web area we are starting to see social networking systems, such as FLINK (Mika 2005), that exploit semantic sources about people's interactions, e.g., FOAF profiles.

Our own interest is in the exploitation of this kind of social network data to support complex knowledge management methods. In particular, we are interested in methods which can mine relationship data from ordinary text of the sort that abounds in many scenarios, as opposed to special such as those used by the bibliometrics community or FOAF. We believe that such methods need to be able to operate on relatively small datasets, such as an organizational intranet, and be robust to data sparseness issues.

To try to address this problem, we have recently been working on an algorithm which we call CORDER (COmmunity Relation Discovery by named Entity Recognition). The first version draws on ideas of co-occurrence from information retrieval practice. Starting with a collection of text documents we use standard information extraction techniques to locate named entities (NEs) in the text, e.g., people's name, organization's names and subject areas of interest. The CORDER algorithm is then used to calculate the strength of the relationship between given pairs of entities. From this process we can build up a matrix of relationship strengths for the community concerned and derive rankings for a given entity. For example, for a particular person we can rank the people, organizations or topic areas with which they have the strongest association. In this paper we will outline the current CORDER algorithm (section 2) and describe a number of application scenarios we have been exploring using CORDER rankings (section 3). These scenarios have pointed towards the need for an algorithm which produces probabilities of relationship strength rather than rankings. We are currently in the process of running experiments to compare a range of alternative algorithms (section 4).

2. CORDER Algorithm

Here we present a brief description of the CORDER algorithm. A full description of the algorithm and an initial evaluation is presented in Zhu et al. (2005a). The process CORDER follows comprises the steps of:

1. *data selection*, in which the Web pages that will represent the organization are identified,
2. *named entity recognition*, in which the pages are preprocessed to identify entities, and
3. *relation strength and ranking*, in which co-occurrence data is processed and the relation strengths of associated NEs with the target are established.

The relation strength calculation is based upon the intuition that the more often two entities occur on the same web page and the closer they are to each other the stronger the relation between them. Its general form for two entities E1 and E2 is

$$\text{RelationStrength} = \text{Cooccurrence} \times \sum \frac{\text{PageRelevance}(\text{Frequency}E1 \times \text{Frequency}E2)}{\text{MinimumDistance}E1 - E2}$$

Looking at each of these components in more detail:

Co-occurrence: Two NEs are considered to co-occur if they appear in the same Web page. We use Resnik's method (Resnik 1999) to compute a relative frequency of co-occurrences of *E1* and *E2* as:

$$\hat{p}(E1, E2) = \frac{Num(E1, E2)}{N},$$

where $Num(E1,E2)$ is the number of pages in which $E1$ and $E2$ co-occur, and N is the total number of pages.

Distance: We calculate the distance between two entities as the difference between their offsets. If *E1* occurs once and *E2* occurs multiple times in the Web page, the distance between *E1* and *E2* is the difference between the offset of *E1* and the offset of the closest occurrence of *E2*. When both *E1* and *E2* occur multiple times in the Web page, we average the distance from each occurrence of *E1* to *E2* and define the logarithm distance between *E1* and *E2* in the *i*th Web page as

$$\overline{d}_i(E1,E2) = \frac{\sum_j (1 + \log_2(\min(E1_j\ ,E2)))}{Freq_i\ (E1)},$$

where $Freq_i(E1)$ is the number of occurrences of *E1* in the *i*th Web page and $\min(E1_j, E2)$ is the distance between the *j*th occurrence of *E1*, $E1_j$, and *E2*.

Frequency: An NE is considered to be more important if it has more occurrences in a Web page. We define frequency as $f(Freq_i(E1)) = 1 + \log_2(Freq_i(E1))$, $f(Freq_i(E2)) = 1 + \log_2(Freq_i(E2))$, where $Freq_i(E1)$ and $Freq_i(E2)$ are the numbers of occurrences of *E1* and *E2* in the *i*th Web page respectively.

Page relevance: Optionally, it might be desired to assign different levels of trust to evidence from different kinds of pages e.g., a high relevance weight might be set to homepages and a low relevance weight to blog pages. w_i is the weight showing the relevance of the *i*th Web page to *E1*,

Relation strength: The final form of the relation strength equation is then:

$$R(E1, E2) = \hat{p}(E1, E2) \times \sum_i \left(\frac{w_i \times f(Freq_i(E1)) \times f(Freq_i(E2))}{\overline{d}_i(E1,E2)} \right)$$

3. Applications

The CORDER approach is proving flexible in a number of applications for which strength of association between entities is of interest. Here we report its use to find experts on a particular topic, to support literature search, to enhance traditional IR vector spaces, and for corpus based ontology selection.

3.1 Finding Experts

The CORDER-BuddyFinder application addresses the problem of finding the right contact from an instant messaging buddy list to answer a particular query (Zhu et al. 2005b). This addresses a scenario in which the user does not necessarily know every person on their buddy list well. For example, in an enterprise wide situation everyone on the pay-roll may be on the list, or in a large collaborative project it may include people from other organizations whose skills are not known to the user. BuddyFinder also assumes that user will be unwilling to spend much time updating profiles, such as FOAF (Friend Of A Friend) files, describing their skills and contacts and so such profiles are also likely to get out of date.

Since CORDER requires only ordinary web pages as input corpus representing the users on a particular list can be built using standard web search tools. In response to a query for say "Java and C++" the BuddyFinder tool can respond with a ranked list with the Buddy's who have the strongest links to the search terms at the top.

3.2 Information Retrieval

One of the inspirational use cases for CORDER has been the database of researchers' resumes held by the Brazilian National Research Council (CNPq) as part of the Lattes Platform (http://lattes.cnpq.br). This has the characteristics of being large and being concerned largely with people, their research interests and experience, and the organizations they have worked for or collaborated with. In short, exactly the kind of entities that CORDER handles.

One of us (Alex Gonçalves) has been investigating whether CORDER outputs can enhance the vector representations of resumes on the Lattes platform to improve search performance. Document vectors are expanded by adding weights for entities which, although they do not appear directly in a given resume have a strong indirect link to it. These indirectly linked entities are co-occurring entities found using CORDER. In this way the vector for a single document can be enhanced by incorporating knowledge derived from the whole document collection. Initial results suggest that the enhanced vectors can give improved clustering results compared to k-means and SOM.

3.3 Bibliographic Search

With its emphasis on the relations between people, and between people and topics, CORDER is well placed to support citation search activities. As part of the Knowledge Web project (http://kmi.open.ac.uk/projects/kweb/) we are deploying CORDER as a web service to help students search for relevant papers within a knowledge portal. The CORDER web service will be deployed on the REASE portal (http://rease.semanticweb.org/ubp) via KMi's Magpie tool which can highlight entities in webpages according to a given ontology (Dzbor 2004). For a highlighted entity, say a person's name, the student will be able to access a list of papers authored by that person, people they are related to, topics they are expert on etc.

3.4 Ontology selection

The semantic web assumes a situation in which a large number of ontologies are publicly available. Consequently, a problem facing the builders of the semantic web is how to match resources with the publicly available ontologies that would yield the best annotation. A natural extension of CORDER would be to use it to help learn an ontology from a text collection. Such a learned ontology could be mapped against a collection of ontologies to find the one (or ones) which could usefully be used to mark up that data.

We have been conducting some initial clustering experiments to determine whether the related entities discovered by CORDER can be used as clustering features to generate conceptual taxonomies which could be applied to this task (Thorne et al. 2005). Ward's agglomerative hierarchical clustering algorithm using the single linkage criterion (Murty et al. 1999) was used with 1) a set-based representation of the data and 2) a vector-space representation of the data. Tests were run using the ACM classification of computer science research areas as our gold standard. So far the most promising results were given by the vector-space representation.

4. Exploring Alternative Algorithms

An interesting outcome of the clustering work is the realisation that more powerful post-processing methods would be available to us if the CORDER algorithm were grounded out theoretically (the current version rests on certain, strong, intuitions about the nature of the problem). In particular, we need a measure in which the ranking values are probabilistic, in order that the values produced for the rankings of one entity can be directly compared to the values for another. We are therefore undertaking a systematic trial of a range of probabilistic and information theoretic measures to find some which give similar ranking performance to CORDER. We intend to consider a range of general purpose statistical measures that include corpus frequency, mutual information, ?^{2}, Z score etc. in comparative tests using a range of window sizes where applicable.

Acknowledgements

This research was partially supported by the Designing Adaptive Information Extraction from Text for Knowledge Management (Dot.Kom) project, Framework V, under grant IST-2001-34038 and the Advanced Knowledge Technologies (AKT) project. AKT is an Interdisciplinary Research Collaboration (IRC), which is sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/N15764/01. Alexandre L. Gonçalves is supported by CNPq, Brazil, with a doctoral scholarship.

We acknowledge the support of many colleagues who have collaborated on the development of CORDER and its applications including, Enrico Motta, Roberto Pacheco, Marc Eisenstadt, Chris Denham and Joerg Diederich.

References

M. Dzbor, Motta, E., Domingue, J.B. (2004) "Opening Up Magpie via Semantic Services", In Proc. of the 3rd Intl. Semantic Web Conference, November 2004, Japan.

P. Mika (2005) "Flink: semantic web technology for the extraction and analysis of social networks", Journal of Web Semantics, 3(2).

M.N. Murty, Jain, A.K., Flynn, P.J. (1999) Data clustering: Review. ACM Computing Surveys, 31(3).

P. Resnik (1999) “Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language”, Journal of Artificial Intelligence Research, 11, 95-130.

Thorne C., Zhu J., Uren, V. (2005) *kmi-05-14*.

D.J. Watts (1999) “Small Worlds”, Princeton University Press.

H.D. White, McCain, K.W. (1989) “Bibliometrics”. In Annual Review of Information Science and Technology (ARIST), 24, 119-186.

J. Zhu, Gonçalves, A.L., Uren, V., Motta, E. Pacheco, R. (2005a) Mining Web Data for Competency Management. In: Proceedings of 2005 IEEE / WIC / ACM International Conference on Web Intelligence 2005 (WI'2005), Compiegne University of Technology, France.

J. Zhu, Eisenstadt, M., Gonçalves, A.L., Denham, C. (2005b) BuddyFinder-CORDER: Leveraging Social Networks for Matchmaking by Opportunistic Discovery, To appear in Proc. of International Semantic Web Conference, (ISWC2005) Workshop on Semantic Network Analysis, November 7, 2005, Galway, Ireland.

Study of Design Issues on an Automated Semantic Annotation System

Yihong Ding
Departments of Computer Science
Brigham Young University
ding@cs.byu.edu

Reference to this article should be made as follows:
Ding Y., (2005), Study of Design Issues on an Automated Semantic Annotation System,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 45-51

Abstract

The semantic annotation process turns ordinary HTML web pages into machine-understandable semantic web pages. We have proposed a semantic annotation system to automate annotation of web pages by ontologies. In this paper, we present a study of five design issues on this system, which include: (1) the covering of web pages; (2) the general paradigm of annotation process; (3) the measurements of annotation performance; (4) the compatibility to semantic web standards; and (5) the resolution for lack of ontologies. To address these issues, our proposed system intends to automate annotation for data-rich web pages; simply the annotation process and improve the degree of automation by adapting the ontology-based data recognizer; measure to obtain accurate annotations, to run fast, and to be resilient to page layouts through a two-layer annotation process; be compatible to the semantic web standards by using OWL ontologies; and benefit an interactive ontology creation process by automatically choosing relevant ontology components according to an annotation task from an ontology knowledge base.

1. Introduction

The semantic web provides a machine-understandable environment [2]. Machines, however, do not really understand web page content unless its meaning is *explicitly* specified in a *formal*, *unambiguous* way. To establish machine understandability, people use *ontologies*, which are explicit, formal specification of conceptualizations [8]. A *semantic annotation* process is thus to label web page content explicitly, formally, and unambiguously using ontologies.

After the emergence of the semantic web, more and more people have accepted it to be the next-generation of World Wide Web. With machine-understandable semantic web pages, we can develop more practical and interoperable web applications. For example, when the semantic web comes to reality, users can directly search web content using database-like queries. To establish the semantic web, however, is difficult. There are billions of pages in the ordinary web and only very few of them have been converted to become machine-understandable. It is impractical to ask web developers to rewrite all their web pages with respect to new semantic-web standards, especially if it involves tedious manual labeling of documents. Hence automatic semantic annotation becomes the important bridge that links the ordinary web to the fascinating semantic web.

By the study of existing semantic annotation approaches (e.g., [1, 3, 9, 10, 12, 13]), we have summarized a typical process for current automated semantic annotation systems. It usually takes an annotation task and a domain ontology as inputs, where an annotation task is a set of web pages waiting to be annotated. The system sends these inputs to an automated data recognizer, which is an adapted information extraction (IE) tool, to extract data instances from web pages. After extracting, the system usually performs "a set of heuristics for post-processing and map-ping of the IE results to an ontology." [10] An annotation generator takes these mappings to eventually create explicit annotations that ontology-aware machine agents can process. These annotations may be stored either within the original web pages or in separate files.

Although this typical annotation process works, there are some problems that limit its practi-cability. For example, Kiryakov et. al. have pointed out that the requirement of the "post-processing and mapping" between the extracted results and ontologies is, as their words, "the main drawback" for these existing automated annotation systems [10]. To solve this problem and along with several other improvements, we have proposed a new automated semantic annotation system using extraction ontologies [4]. As Figure 1 shows, this proposed system takes web pages and domain ontologies as inputs. When there are no input ontologies, the system embeds an ontology assembler that provides an interactive ontology creation process by automatically choosing relevant ontology components according to the annotation task from an ontology knowledge base. Depending on the number of web pages and the degree of domain diversity, the system either hands the task to the conceptual annotator, which annotates documents by ontology-based domain specifications, or hands it to the structural annotator, which annotates documents by page-layout specifications. The two annotators share a common annotation gene-rator. The ontology converter in the figure assures that our system is compatible to a semantic web standard, which is OWL (Web Ontology Language) [14].

The focus of this paper is to present five design issues we have studied when proposing this automated annotation system, which include: (1) the covering of web pages our system works for; (2) the general paradigm of semantic annotation process we have simplified to improve the degree of automation of our system; (3) three performance measurements (accuracy, speed, and resiliency) our system aims to achieve; (4) the compatibility to the semantic web standards our system holds; and (5) the resolution our system provides to handle annotating when there are no input ontologies. Through the discussion of these design issues, we not only describe the details of our proposed system, but also state the reasons that our improvements are effective to achieve practical semantic annotations.

In the rest of this paper, we start with the discussion of the web page coverage issue in Section 2. In Section 3, we present the reason that the adaptation of the ontology-based data recognizer increases the degree of automation of our annotation system by simplifying the paradigm of the typical annotation process. Section 4 presents the three performance measurements and the reason our two-layer annotation model improves their evaluations. Section 5 addresses the issue that our system is compatible to the semantic web standards. In Section 6, we present a mecha-nism in our system to semi-automatically assemble a domain ontology when no ontology inputs. In the end, we conclude in Section 7.

2. Web Pages Coverage

The first design issue we have addressed is the covering of web pages our automated annotation system works for, which is determined the data recognizers we are going to adapt. There are no automated IE tools that can effectively perform on all web pages [11]. Each individual IE approach has its favorite covering of web pages, from which it is effective to extract data. For example, it is favorable to use NLP (natural language processing) based IE tools to extract unstructured free-text documents, while HTML-aware IE tools are effective to extract data from fully structured HTML web pages [11].

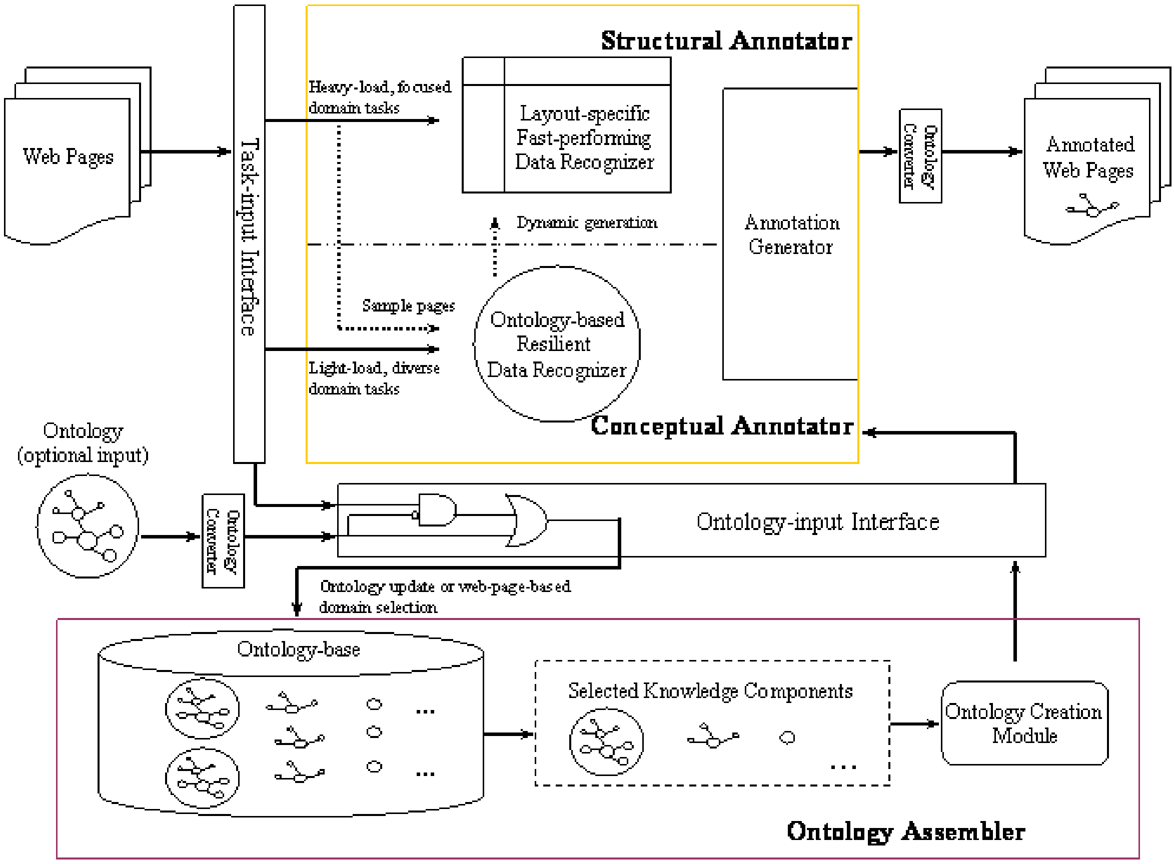

Structural Annotator
Web Pages
Task-input Interface
Heavy-load, focused domain tasks
Layout-specific Fast-performing Data Recognizer
Dynamic generation
Annotation Generator
Ontology Converter
Annotated Web Pages
Sample pages
Ontology-based Resilient Data Recognizer
Light-load, diverse domain tasks
Conceptual Annotator
Ontology (optional input)
Ontology Converter
Ontology-input Interface
Ontology update or web-page-based domain selection
Ontology-base
Selected Knowledge Components
Ontology Creation Module
Ontology Assembler

Figure 1: Framework of Automated Semantic Annotation System Using Extraction Ontologies

Moreover, the importance of this coverage issue is not only for the anxiety of determining applicable web pages, but also for the curiosity of knowing appropriate domains the annotation system can effectively manage. A fact in the web is that each domain usually has its typical web representing format. For example, news is usually written in free-text web pages; and shopping categories are often presented within complex HTML tables without many complete natural language sentences. Therefore, when an annotation system adapts an NLP-based IE tool, it can perform well on the news domain but with less accuracy on the shopping domain. Similarly, an annotation system with an HTML-aware data recognizer can effectively annotate the shopping domain but not so effective on annotating the news domain.

We plan to design a system that can effectively annotate semi-structured and fully structured data-rich web pages that each have a relatively narrow domain. The ontology-based IE tool matches this purpose [6, 7]. We must mention that this type of web page coverage is not unique for our annotation approach (see, for example, [12]). Moreover, this type of web page is common on the web (shopping, product portals, for example).

3. Annotation Process Paradigm

The second system design problem we encountered is to establish the entire annotation process. We have mentioned earlier that a significant problem in the typical annotation process is the requirement of the "post-processing and mapping of the IE results to an ontology." The reason causing this problem is the independency between ontologies and the non-ontology-based IE wrappers (to become the data recognizers). According to the survey written by Laender et. al., all the automated IE approaches except the ontology-based ones do not extract data with respect to ontologies [11]. Since ontology is a mandatory factor in the semantic annotation scenario,[3] this "post-processing and mapping" problem is unavoidable for the annotation systems with adapted non-ontology-based data recognizers.

To solve this problem, we have integrated domain ontologies with extraction engines so that the extraction process is executed directly with respect to the ontological declarations [5]. We declare extraction patterns to be the extensional semantics within domain ontologies, through which the ontology-based data recognizer extracts data instances. Hence these data instances are directly categorized into their corresponding ontology definitions without the needs of further "post-processing and mapping." Through this new paradigm, we do improve the system's degree of automation because the original "post-processing and mapping" often requires much human involvement. Our resolution actually fulfills what Kiryakov et. al. have suggested in [10].

4. Performance Measurements

After establishing the general semantic annotation process, we did a measurement study to ensure the overall performance of our system. We believe that three performance measurements, which are accuracy, speed, and resiliency, are equally important to a semantic annotation system. There is no problem that accuracy and speed measurements are crucial. For automated semantic annotation systems, high resiliency to web page layouts is also important because otherwise a system may have to regenerate data recognition patterns each time for a new page layout. This type of regeneration usually decreases the degree of automation for the system.

To achieve good performance on all the three measurements, we have proposed a two-layer annotation model that contains two annotators—a lower-layer conceptual annotator and an upper-layer structural annotator. The conceptual annotator employs an ontology-based data recognizer to perform resilient annotating on web pages with varied layouts. The structural anno-tator employs one or more layout-specific data recognizers, where each recognizer can annotate web pages with a common layout fast and accurate. Figure 1 shows that the two annotators are separated by the dot-dash line.

The resiliency property for the ontology-based data recognizer is inherited from the ontology-based IE approach [6]. Our ontology-based IE tool can continuously work on different web page layouts so long as the pages are for the same domain because its extraction process is based on the declarative domain-oriented extensional semantics without encoding of layout information. As a trade-off to be resilient, our ontology-based data recognizer requires a large number of computational cycles to enumerate all possible candidate instances and resolve ambiguities. Hence, its execution speed is relatively slow. Also, although the ontology-based data recognizer is designed to achieve good accuracy in general cases, it does not take into account local structural patterns, which can lead to higher extraction accuracy.

On the contrary, a layout-specific data recognizer performs very fast because it usually requires only a single pass through an entire document to do extraction. It also assures very high accuracy because it only processes web pages that match a known layout structure. Therefore, layout-specific data recognizers are not resilient. They usually fail to perform correctly when layouts are unknown or change. When there is a new layout pattern, the system needs a regeneration process to build a new layout-specific data recognizer.

[3] Ontologies are optional within the traditional IE paradigm. Due to the difficulty of ontology generation, many traditional IE researchers do not input ontologies to their automated IE systems.

Figure 1 illustrates how we integrate these two annotators. When the system needs to annotate large numbers of web pages, and especially if these web pages are for a focused domain and hold a common layout,[4] the system takes a small set of web pages out of the input task to be samples, which are sent to the conceptual annotator (as the long dotted arrow-head line in Figure 1). After the conceptual annotator annotates them, the system processes a structural analysis on the annotated sample pages, through which the system dynamically creates a layout-specific data recognizer according to the presented page layout (as the short dotted arrow-head line in Figure 1). This created layout-specific data recognizer and the annotation generator together compose the structural annotator. The system then sends the vast majority of the input web pages to the created structural annotator to take the benefit of fast speed and high accuracy annotation. On the contrary, when the number of input web pages is small and the layouts are varied, it is too expensive to process dynamic generation of multiple data recognizers, while each of them annotates only a very small number of web pages. The system therefore simply let the conceptual annotator annotates the whole task to take the benefit of resiliency.

We must further point out two characteristics about this dynamic generation of the structural annotator. First, it holds the property of resiliency. When the set of sample pages contains multiple layouts, the conceptual annotator can annotate them continuously due to its resiliency. Using the annotated web pages, the system can simultaneously create a layout-specific data recognizer for every input layout. Second, this dynamic generation process does not conflict to the elimination of the “post-processing and mapping” we have just discussed. Layout-specific data recognizers are generated from annotated web pages that already contain correct mappings between data instances and ontology concepts. Therefore, the mappings between extraction categories in the generated layout-specific data recognizers and ontology concepts are ensured.

5. Compatibility to Semantic Web Standards

When we design our semantic annotation system, we want it to be compatible to the semantic web standards so that it can be directly used by the rest of the semantic web society. However, the adapted ontology-based data recognizer requires OSMX (Object-oriented Systems Model in XML) ontologies [5, 6], which is not a semantic web standard. We thus need to do a conversion between OSMX and a semantic web standard. Since OWL is widely accepted to be a standard semantic web ontology language, we choose it to be the semantic web standard in our system.

Fortunately, the OSMX and OWL representations are quite similar and compatible to each other. Many conversions are straightforward. For example, an *object set* in OSMX is a *class* in OWL; a *relationship set* in OSMX is an *ObjectProperty* in OWL; a *participation constraint* in OSMX is a *Cardinality* restriction in OWL; an *isa* hierarchical relationship in OSMX is a *subClassOf* specialization in OWL. There are, however, some unique specifications in each language. For example, the data frames, which describe extensional semantics of ontology concepts in OSMX, are not well defined in OWL, while OSMX does not explicitly support the subproperty feature in OWL. Our ontology converter needs to address and solve these specialties.

As Figure 1 shows, we put the converter on both the input and output sides of our system. On the input side, when users input an OWL ontology, the converter transforms it to its OSMX representations. Otherwise, the system simply ignores the converter and takes the input OSMX

[4] This is quite common in the ordinary web. For example, the auto-generated web pages within many large commercial web sites, such as amazon.com or ebay.com, hold common layouts and domain.

ontology to the data recognizers. On the output side, when the original input ontology is in OWL, the system generates annotations directly with respect to the original OWL ontology, and thus, no conversion is needed. Otherwise, when there are no input OWL ontologies, the system converts the OSMX representations used by the data recognizers to their OWL representations, and outputs annotations with respect to the converted OWL ontology.

6. Resolution for Lack of Input Ontologies

Until now, we assume that there are input ontologies, which is also the assumption most of the current automated semantic annotation systems make [1, 3, 9, 10, 12, 13]. However, ontology creation is difficult and most of current ontologies are constructed manually. Many times users simply cannot find an existing ontology that is appropriate for their annotation task. Our system, therefore, propose the ontology assembler to help users build an ontology semi-automatically on the absence of input ontologies. The theme underlying our ontology assembler is to maximize the reuse of existing ontologies and minimize the work of constructing new ontologies.

The bottom part of Figure 1 shows our ontology-input interface and the ontology assembler. The logic circuit inside the ontology-input interface shows that the interface alternately inputs a set of descriptive web pages, which illustrate the domain of interest, to the ontology assembler when there are no input ontologies. The ontology assembler consists of two parts—an ontology-base and an ontology creation module. The ontology-base consists of pre-used and pre-constructed ontologies, snippets of ontology, and single concept recognizers. The ontology creation module in our system is an ontology editor that users can view and manually create or modify ontologies.

When there is an input ontology, the assembler simply updates the ontology-base with the input ontology and displays it by the ontology editor. With users' approval, the system sends the ontology to the annotators to accomplish the annotation task. Otherwise, there are no input ontologies but a set of descriptive web pages. The assembler performs a knowledge-selection process to look for relative ontology components within the ontology-base with respect to the descriptive web pages. These ontology components could be pre-existing ontologies, snippets of ontology, or single concept recognizers, as the dashed box inside the ontology assembler shows. The assembler thus sends these selected components to the ontology creation module, through which users can view these components, integrate the appropriate ones, and build missing parts, if necessary. Finally, users assemble the appropriate components to be a unified ontology, which is the name ontology assembler coming from.

7. Concluding Remarks

This paper presents a brief introduction of our automated semantic annotation system. Through the discussion, we present five design issues that we have addressed when proposing this system.

- The web page coverage issue affects both the web page types and potential applicable domains an annotation system can well-perform. It decides the theme of an annotation system. Our system focuses on annotating semi-structured and structured data-rich web pages, which are common on the ordinary web.
- With the focused theme, we need to figure out an effective process to accomplish annotation tasks. The ontology-based data recognizer helps to improve the degree of automation of our system by eliminating the requirement of the "post-processing and mapping of the IE results to an ontology."

- Having an effective annotation process, the success of our system is closely related to the performance measurements. The two-layer annotation model assures our system to obtain accurate annotations, to run fast, and to be resilient to page layouts.
- Nevertheless, the acceptance of our annotation system depends on whether our system is compatible to the semantic web standards. The ontology converter assures our system to accept OWL ontologies and to produce annotations with respect to OWL representations.
- The existence of ontologies is not promising, while the requirement of ontologies is demanding. When there are no input ontologies, our ontology assembler helps to build a task-oriented ontology through an interactive process by automatically choosing relevant ontology components according to an annotation task from an ontology knowledge base.

Upon to the time we submit this paper, this is an on-going project. Our papers [4] and [5] have described more details of our proposed system and what we have done. There is also an online demo of our annotator.[5] Through this study, we expect to deliver the vision that it is convincible to develop practical semantic annotation systems that can automatically accommodate the huge quantity of existing data-rich web pages on the ordinary web.

References

[1] L. Arlotta, V. Crescenzi, G. Mecca, and P. Merialdo, “Automatic annotation of data extracted from large web sites,” In *Proceedings of Sixth International Workshop on the Web and Database (WebDB 2003)*, pp. 7-12, San Diego, California, June 2003.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, “The Semantic Web,” *Scientific American*, 36(25):34-43, May 2001.

[3] S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. Guha, A. Jhingran, T. Kanungo, K.S. McCurley, S. Rajagopalan, A. Tomkins, J.A. Tomlin, and J.Y. Zien, “A Case for Automated Large Scale Semantic Annotations,” *Journal of Web Semantics*, 1(1):115-132, December 2003.

[4] Y. Ding, “Annotating Documents for The Semantic Web Using Data-Extraction Ontologies,” *PhD dissertation proposal*, Brigham Young University, September 2005.

[5] Y. Ding, D.W. Embley, S.W. Liddle, “Semantic Annotation Based On Extraction Ontologies,” 2005. (submitted to review)

[6] D.W. Embley, D.M. Campbell, Y.S. Jiang, S.W. Liddle, D.W. Lonsdale, Y.-K. Ng, and R.D. Smith, “Conceptual-model-based data extraction from multiple-record web pages,” *Data & Knowledge Engineering*, 31(3):227-251, November 1999.

[7] D.W. Embley, C. Tao, and S.W. Liddle, “Automating the extraction of data from HTML tables with unknown structure,” *Data & Knowledge Engineering*, 54(1):3-28, July 2005.

[8] T.R. Gruber, “A translation approach to portable ontology specifications,” *Knowledge Acquisition*, 5(2):199-220, 1993.

[9] S. Handschuh, S. Staab, and F. Ciravegna, “S-CREAM Semi-automatic CREAtion of Meta-data,” In *Proceedings of European Conference on Knowledge Acquisition and Management (EKAW-2002)*, pp. 358-372, Madrid, Spain, October, 2002.

[10] A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff, “Semantic Annotation, Indexing, and Retrieval,” *Journal of Web Semantics*, 2(1):49-79, December 2004.

[11] A.H.F. Laender, B.A. Ribeiro-Neto, A.S. da Silva, and J.S. Teixeira, “A brief survey of web data extraction tools,” *SIGMOD Record*, 31(2):84-93, June 2002.

[12] S. Mukherjee, G. Yang, and I.V. Ramakrishnan, “Automatic Annotation of Content-Rich HTML Documents: Structural and Semantic Analysis,” In *Proceedings of Second International Semantic Web Conference (ISWC 2003)*, pp. 533-549, Sanibel Island, Florida, October, 2003.

[13] M. Vargas-Vera, E. Motta, J. Domingue, M. Lanzoni, A. Stutt, and F. Ciravegna, “MnM: Ontology Driven Tool for Semantic Markup,” In *Proceedings of Workshop Semantic Authoring, Annotation & Knowledge Markup (SAAKM 2002)*, pp. 43-47, Lyon, France, July, 2002.

[14] W3C (World Wide Web Consortium) *OWL Web Ontology Language Reference*. URL: http://www.w3.org/TR/owl-ref/.

[5] http://www.deg.byu.edu/

Semantic Annotation for Semantic Social Networks Using Community Resources

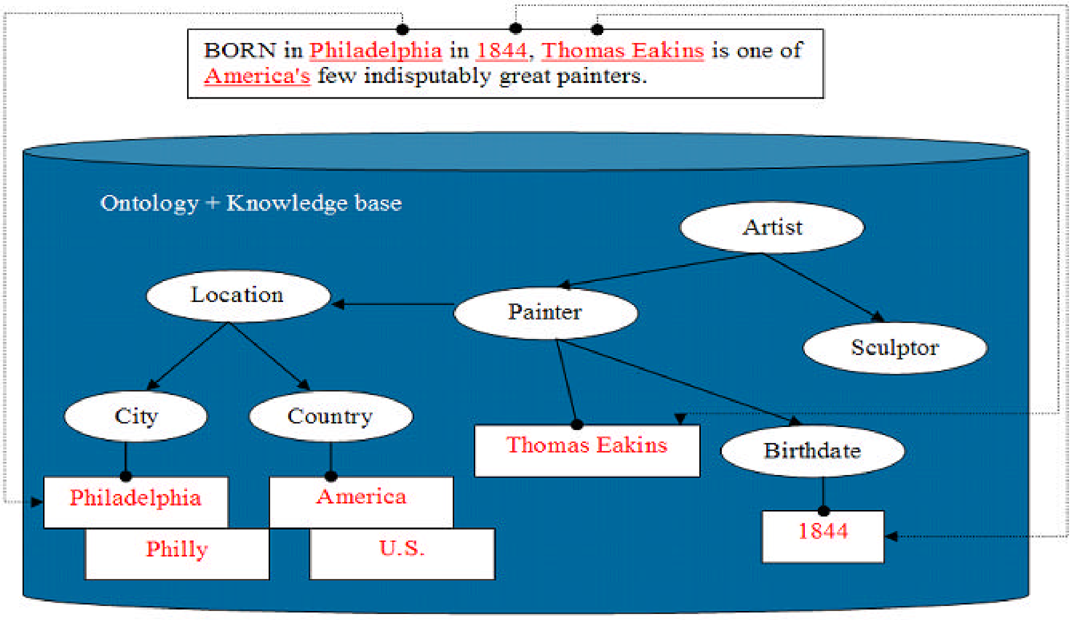
Lawrence Reeve and **Hyoil Han**
College of Information Science and Technology
Drexel University, Philadelphia, PA 19108
lhr24@drexel.edu
hhan@cis.drexel.edu

Reference to this article should be made as follows:
Reeve L., Han H., (2005), Semantic Annotation for Semantic Social Networks
Using Community Resources,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 52-56

Semantic Social Networks (SSN) merge the Semantic Web with social networking so that resources and people related to the resources are linked together (Downes, 2004). An advantage of social and Semantic Web merging is to facilitate information searches among people with related interests. SSN applications environments are to include, among other features, content creation facilities. An ongoing problem with any content creation on the Semantic Web is the semantic annotation (or semantic tagging) of information of the new content. While tagging of content using user profile information is addressed by SSN, semantic tagging of content is not. We propose to provide a semi-automatic semantic tagging facility for new (and existing) Web-based text content using community-based knowledge resources.

Automatic semantic annotation of information content is an open problem, but is crucial to the realization of the Semantic Web. Annotation systems require the initial definition of an ontology and as well as a knowledge base. Both of these resources work together to facilitate markup. The ontology identifies the important concepts in a domain, while the knowledge base provides additional information, such as term synonyms for concepts. For example, the concept {Lung Cancer} can be expressed using at least three different terms: {"Lung Cancer," "Cancer of the Lung," "Carcinoma of the Lung"}. Semi-automatic semantic annotation systems use the synonyms in the knowledge base to find instances in a text source, and then map the instance to an ontological concept (i.e., a concept in the ontology). Figure 1 shows an example of how entities in a text source are mapped into ontological concepts using a knowledge base. Ontological concepts are represented as ovals, while knowledge base entries are indicated by rectangles.

Figure 1: Semantic tagging using an ontology and knowledge base.


BORN in Philadelphia in 1844, Thomas Eakins is one of America's few indisputably great painters.
Ontology + Knowledge base
Artist
Location
Painter
Sculptor
City
Country
Thomas Eakins
Birthdate
Philadelphia
America
Philly
U.S.
1844

There are three classes of semantic annotation systems: manual, semi-automatic, and automatic. Manual annotation provides facilities within a content editing environment allowing a user to select concepts from a predefined knowledge source. An example of this style of annotation is Semantic Word (Tallis, 2003), which provides an environment for authoring as well as marking up documents from within a single interface. The most significant drawbacks to manual annotation are the expense and inconsistency of human annotators. Semi-automatic annotation is currently the most viable approach. Semi-automatic systems perform text analysis to identify instances and then label the text with their corresponding ontological concepts. These systems are not completely automatic, however, due to the problem of disambiguation. For example, in biomedicine applications, there are two concepts for the term {Mass}: a quantitative concept {"how much"} and a finding concept {"found a mass"}. If there are insufficient clues to disambiguate which concept is intended, the system must consult the user to disambiguate the term. If the disambiguation problem can be solved, then automatic systems will be possible. For more descriptions of semantic annotation systems, please see our review paper (Reeve & Han, 2005).

In order for semantic annotation systems to perform, the knowledge base and ontology must be defined. There is often a considerable amount of work associated with constructing and maintaining these knowledge sources. In addition, the result is usually domain specific. One attempt at large scale automatic semantic tagging is the Seeker platform (Dill et al., 2003). Seeker has tagged 264 million Web pages with 434 million semantic tags. The tagging is done using an application called SemTag. SemTag uses as its knowledge source the TAP knowledge base (TAP KB) from Stanford University (Guha, R., McCool Robert, 2003). TAP KB is shallow but broad, and covers 12 categories with approximately 72,000 tags: Authors, Autos, Baby products, Companies, Consumer electronics, Health, Home Appliances, Movies, Music, Places, Sports and Toys.

SemTag scans text sources, finds tag instances, disambiguates them, and finally annotates the text using TAP tags.

We propose a similar approach where the TAP KB component is replaced with the Wikipedia free encyclopedia (I. Wikimedia Foundation, 2005) as the knowledge source. There are several reasons for choosing Wikipdia as a knowledge source. Wikipedia is an online encyclopedia developed by volunteer authors. The content is subject to consensus, rather than authoritative, approval. In this way, the content reflects the views of the larger community rather than a particular viewpoint. It is this aspect that is useful for using Wikipedia as a knowledge source for tagging on the general Semantic Web. Typically, Semantic Web content is tagged using a domain-specific ontology. It is therefore possible to tag the same content with different ontologies to gain different views of the same content. The use of a community-based, consensus-built knowledge source is one way to bootstrap Semantic Web content. That is, it can provide an initial tagging of content that can later have additional ontology tagging performed to reflect different views. We also find Wikipedia useful because it has an active community, and current event topics are updated or added as they occur. This allows new Semantic Web content to be brought online quickly. Finally, Wikipedia content is licensed using GNU Free Documentation License (Free Software Foundation, Inc., 2002), and is freely downloadable in XML format for machine processing. Wikipedia is available in 200 languages, and has more than 50,000 article entries for each of the ten most active languages, making it a large, multilingual and actively-developed knowledge source.

In order to make use of Wikipedia as knowledge resource for semantic annotation, semantic labels must first be extracted. Since Wikipedia was not designed for semantic annotation, processing must be done to convert the article content into useful tags. We propose converting Wikipedia content into a metathesaurus format to store the concepts and their term representations, similar to the National Library of Medicine’s Unified Medical Language System (UMLS) Metathesaurus (United States National Library of Medicine, 2004). The UMLS Metathesaurus is composed of concepts and synonyms. The synonyms are based on medical vocabularies. UMLS also provides a semantic network to organize the concepts.

Each page in Wikipedia describes some topic. The topic name of each page becomes the concept name in the metathesaurus. The implication is that the community has determined that these topics are the most important. The name of the concept (topic page name) also becomes one of the lexical terms for identifying the concept. Figure 2 shows a fragment of markup for the opening text of the Wikipedia topic “John Roberts.” We define the opening text as the markup segment bounded by the start of the topic page markup to the first segment divider, which is indicated by the prefix “==” on its own line. The topic name (concept name) is indicated by triple quotes. In example 2, this is '''John Glover Roberts, Jr.'''. Links to other concepts are implemented using opening and closing brackets ([[and]]). These links are used as ‘related-to’ concepts and are helpful for disambiguating the topic page concept ({John Glover Roberts, Jr.} in this example).

Figure 2: Wikipedia opening text markup for the topic “John Roberts.”

'''John Glover Roberts, Jr.''' (born [[January 27]], [[1955]]) is the seventeenth [[Chief Justice of the United States]]. Roberts previously was a judge on the [[United States Court of Appeals for the District of Columbia Circuit]], spent 14 years in [[Law of the United States|private law practice]], and held positions in [[Republican Party (United States)|Republican]] administrations in the [[United States Department of Justice|U.S. Department of Justice]] and [[White House Counsel|Office of the White House Counsel]].

==Personal life, education, and memberships==

From the topic page opening text, “John Roberts” is identified as a concept and also as the preferred term for the concept. In addition, “John Roberts” becomes a lexical term. Additional synonyms are derived from the opening text. Synonyms are usually marked by including them within three single quotes (‘’’). The three-quote heuristic is one way of identifying synonyms. Synonyms can also be derived from Wikipedia redirect pages. Redirect pages redirect users using one term to the main topic page. For example, the topic page “John Glover Roberts, Jr.” redirects to the “John Roberts” topic page. “John Glover Roberts, Jr.” is then identified as a synonym for “John Roberts.”

A basic semantic network to classify related concepts can be generated from Wikipedia topic categories. A Wikipedia category is a special page provided within Wikipedia to organize topic pages. A Wikipedia topic page can belong to multiple categories. For example, the topic page (concept) “John Roberts” belongs to nine categories (“Chief Justices of the U.S.,” “Judges of the U.S. Court of Appeals for the DC Circuit,” “American lawyers,” “Harvard alumni,” “Harvard Law School graduates,” “Roman Catholic jurists,” “Ambidextrous people,” “People from New York,” “1955 births.”) Subcategories are also used within Wikipedia, forming a hierarchical tree of topics. For example, “Chief Justices of the U.S.” is a subcategory of “United States Supreme Court.” Building a semantic network from topic categories is somewhat problematic because multiple categorization schemes can exist at the same time (Wikimedia Foundation, 2005).

An important part of any semantic tagging application is the disambiguation step. It is possible for a term to map to multiple concepts (Wikipedia topics), and automated tagging is often unable to distinguish which concept is the intended concept. In these cases, the machine defers to a user to disambiguate among many candidate concepts. Wikipedia provides disambiguation pages to let users know a topic may also refer to other topic pages with similar topic page names. Using the example from above, the term “John Roberts” can map to “John Roberts” the Chief Justice of the United States, “John Roberts” the television journalist, or “John C. Roberts” the founder of an Australian construction company, among others. In our observation, the most important information is contained in the opening text. To perform automatic disambiguation, the concepts in the opening text are extracted as ‘related-to’ instances of the main concept. In Figure 2, these are indicated with the markup contained in opening and closing brackets ([[and]]). The related-to concepts are then used in the disambiguation step in the following manner. First, the source text is processed to identify all ambiguous concepts and unambiguous concepts. Each unambiguous concept has its corresponding unique concept and is annotated (or tagged) with the unique concept. Ambiguous concepts have more than one candidate concept. For each candidate concept of an ambiguous concept, the unambiguous concepts in the source text already tagged are used to find matching related-to concepts for each candidate concept. The candidate concept having the highest frequency count of related-to concepts becomes the disambiguated concept, and the source text is tagged with this concept. If all candidate concepts of an ambiguous concept do not have any related-to concepts, or a frequency count tie occurs, the machine is unable to successfully complete disambiguation and the user must be consulted for a final determination of the semantic label. This approach to disambiguation does not require any prior training, as is the case with other disambiguation approaches, such as SemTag (Dill et al., 2003). This approach also allows the re-use of concepts already manually tagged by the topic editors of Wikipedia.

Social semantic networks (SSN) integrate the existing technologies of social networks and the Semantic Web. We believe that by extending this integration to include community-based knowledge sources and applying these resources to semantic annotation, Semantic Web content for SSNs will appear more rapidly and be more valuable to their users.

References

Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., & Jhingran, A. et al. (2003). SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation. *Twelfth International World Wide Web Conference*, Budapest, Hungary, 178-186.

Downes, S. (2004). *The Semantic Social Network*. Retrieved October 9, 2005 from www.downes.ca/cgi-bin/website/view.cgi?dbs=Article&key=1076791198.

Free Software Foundation, Inc. (2002). *GNU Free Documentation License*. Retrieved October 9, 2002 from http://en.wikipedia.org/wiki/Wikipedia:Text_of_the_GNU_Free_Documentation_License.

Guha, R. and McCool, R. (2003). TAP: A Semantic Web Platform. *Computer Networks: The International Journal of Computer and Telecommunications Networking. Special Issue: The Semantic Web: An Evolution for a Revolution, 42*(5), 557-577.

Reeve, L., & Han, H. (2005). Survey of Semantic Annotation Platforms. *Proceedings of the 20th Annual ACM Symposium on Applied Computing, Web Technologies and Applications track,* Santa Fe, New Mexico.

Tallis, M. (2003). Semantic Word Processing for Content Authors. *Second International Conference on Knowledge Capture,* Sanibel, Florida, USA.

United States National Library of Medicine. (2004). *UMLS Metathesaurus Fact Sheet*. Retrieved July 31, 2005 from http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html.

Wikimedia Foundation. (2005). *Wikipedia Category*. Retrieved October 9, 2005 from http://en.wikipedia.org/wiki/Wikipedia:Category.

Wikimedia Foundation, I. (2005). *Wikipedia*. Retrieved October 9, 2005 from http://en.wikipedia.org/wiki/Main_Page.

Podcasting and Its Role in Semantic Social Networks, the Web 2.0, and the Semantic Web

G. Philip Rogers
School of Library and Information Science,
The University of North Carolina, Chapel Hill
gershomrogers@acm.org

Reference to this article should be made as follows:
Rogers P., (2005), Podcasting and Its Role in Semantic Social Networks, the Web 2.0, and the Semantic Web,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 57-61

Scholars who are interested in semantic social networks6 have focused much of their attention on behaviors related to the use of social networking sites such as Friendster and Orkut, as well as the use of content syndication technologies by members of the blogging community. Presented with an increasing volume of content, a significant number of information consumers are gravitating toward collaborative tagging applications7 that allow them to organize content in ways meaningful to them [1], as well as toward content aggregators8 that check for and download new syndicated content of a specified type. Meanwhile, the explosion in popularity of handheld devices, most notably the iPod, has made the content syndication model all the more appealing, not only for digital music downloads, but also for a wide variety of additional content that is now being “podcasted” to desktops and mobile devices worldwide.

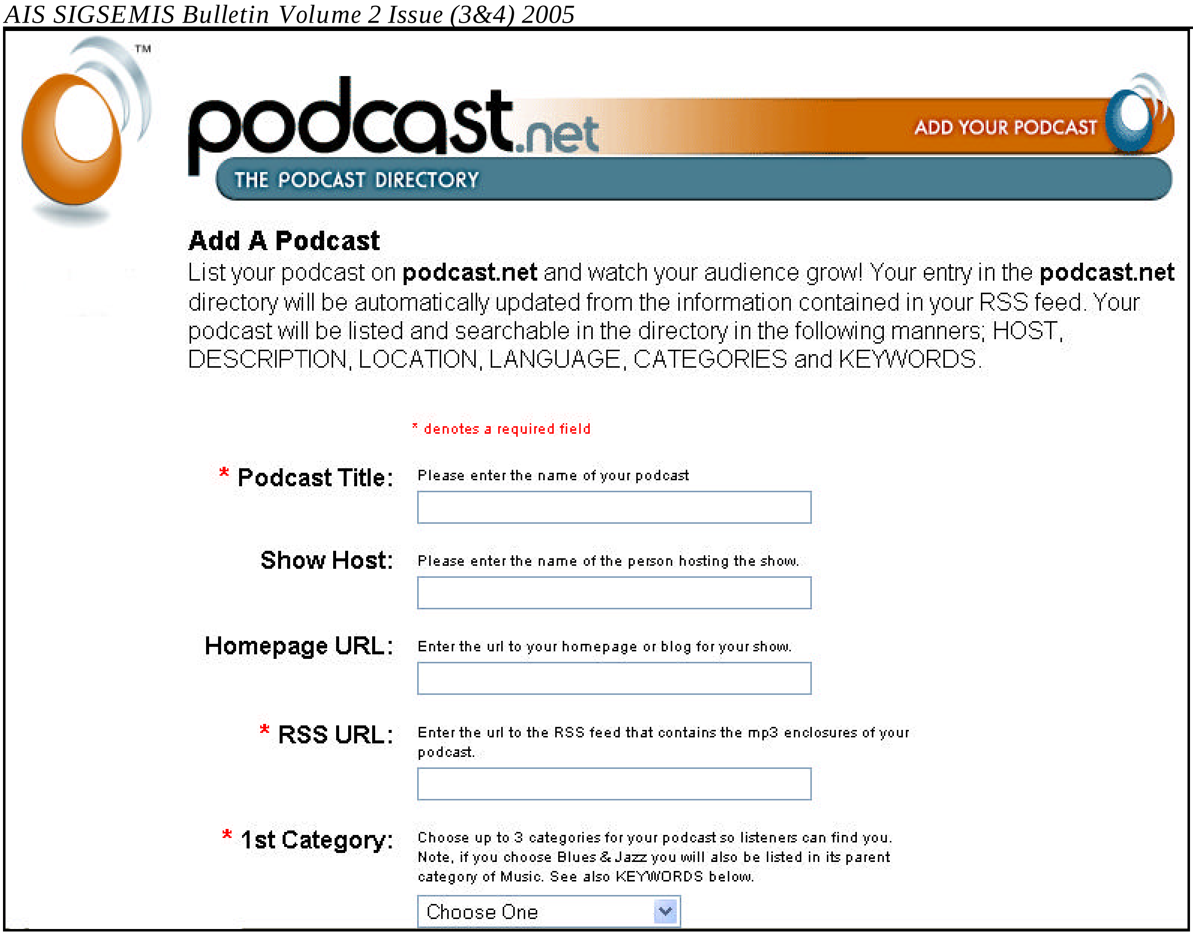
As the name implies, a podcast is content such as a radio show that is recorded in the ubiquitous MP3 format and broadcast (or more accurately, published) on a web site for download by anyone who cares to listen to it on a mobile device or a computer. Through the use of RSS (Really Simple Syndication), information about the web site and the podcasts (or other content) that is available on the web site is provided in a lightweight XML format. The RSS files, or “feeds,” can be harvested by content aggregators designed for podcasts, such as iPodder or iPodderX, or by other aggregators, such as iTunes, all of which can download “subscriptions” either on demand or at predetermined intervals.

Analysis of several podcast sites reveals a great deal about the nature of the producers and consumers of podcasts and the manner in which they are participating in semantic social networks. One way to gain insight into the world of podcasting is to look at the metadata that is available to anyone who wishes to publish a podcast, as shown in Figure 1, the Add Your Podcast page at podcast.net [2].

[6] The most commonly cited definition of a semantic social network is that of Stephen Downes, who suggests that content syndication and social networking technologies are converging, creating a “new type of internet, a network within a network, and in so doing reshap[ing] the internet as we know it.”
http://www.downes.ca/cgi-bin/website/view.cgi?dbs=Article&key=1076791198

[7] Examples of collaborative tagging applications include social bookmark managers such as del.icio.us and del.irio.us, and bookmark managers for scholars such as Connotea and CiteULike.

[8] Examples of content aggregators include Bloglines, FeedGator, and iPodder.



Add A Podcast

List your podcast on **podcast.net** and watch your audience grow! Your entry in the **podcast.net** directory will be automatically updated from the information contained in your RSS feed. Your podcast will be listed and searchable in the directory in the following manners; HOST, DESCRIPTION, LOCATION, LANGUAGE, CATEGORIES and KEYWORDS.

* denotes a required field

* Podcast Title: Please enter the name of your podcast

Show Host: Please enter the name of the person hosting the show.

Homepage URL: Enter the url to your homepage or blog for your show.

* RSS URL: Enter the url to the RSS feed that contains the mp3 enclosures of your podcast.

* 1st Category: Choose up to 3 categories for your podcast so listeners can find you. Note, if you choose Blues & Jazz you will also be listed in its parent category of Music. See also KEYWORDS below.

Choose One

Figure 1

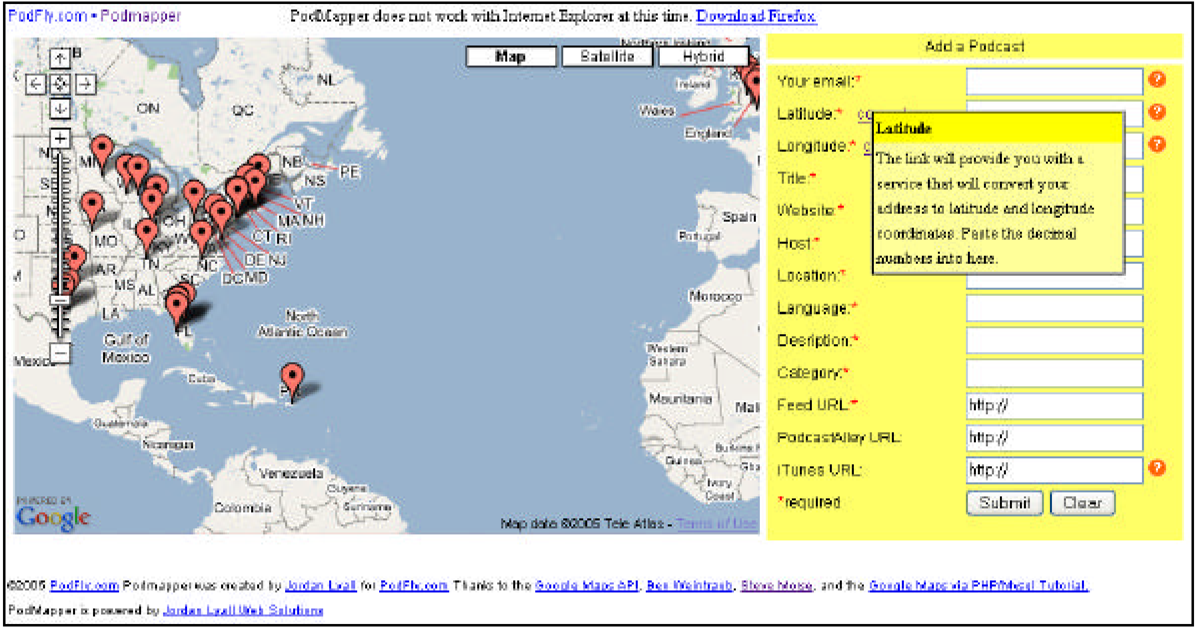
The complete list of metadata, including optional elements, is as follows: Podcast Title, Show Host (the person who recorded the podcast), Homepage URL, RSS URL, 1st Category (keywords), 2nd Category (additional keywords), 3rd Category (additional keywords), Description (what the podcast is about), Location (such as where the podcast was recorded), Language, and Adult Content (an optional rating system, where the choices are: Adult Content; Adult Language, and; Sexual Content).

One of the most intriguing features of podcast.net is the availability of a rich set of keywords.9 Not only do these keywords act as a means for listeners to search for podcasts that interest them, they also serve as tags to which users can subscribe. For example, by clicking the RSS link at the top of a particular page on podcast.net, a user can subscribe to a particular tag, such as "technology." As a result, when new podcasts are added that include the specified tag, the subscriber receives those new podcasts [3].

Another podcasting site, podfly.com, provides metadata fields not available to users on other podcast sites. The additional fields, Latitude and Longitude10, are employed on podfly.com's Podmapper (see Figure 2), an application written by Jordan Lyall that leverages Google Maps to display the geographic location of people who have added podcasts via the Podmapper [4].

[9] As of October 8, 2005, the ten most popular tags on podcast.net were as follows: 1. music (1036); 2. comedy (420); 3. news (381); 4. podcast (343); 5. radio (343); 6. rock (307); 7. politics (276); 8. technology (274); 9. talk (221); 10. humor (203). See http://www.podcast.net/browsetags/.

[10] Podmapper directs the user to Stephen Morse's Converting Addresses to Latitude/Longitude in One Step page, where the user enters their physical address and is then directed to paste the resulting Latitude and Longitude (in decimal format) into the Podmapper Latitude and Longitude fields, respectively.



PodFly.com • Podmapper

PodMapper does not work with Internet Explorer at this time. Download Firefox

Add a Podcast

Your email:*	
Latitude:*	
Longitude:*	
Title:*	
Website:*	
Host:*	
Location:*	
Language:*	
Description:*	
Category:*	
Feed URL:*	http://
PodcastAlley URL:	http://
iTunes URL:	http://
*required	Submit Clear

Latitude

The link will provide you with a service that will convert your address to latitude and longitude coordinates. Paste the decimal numbers into here.

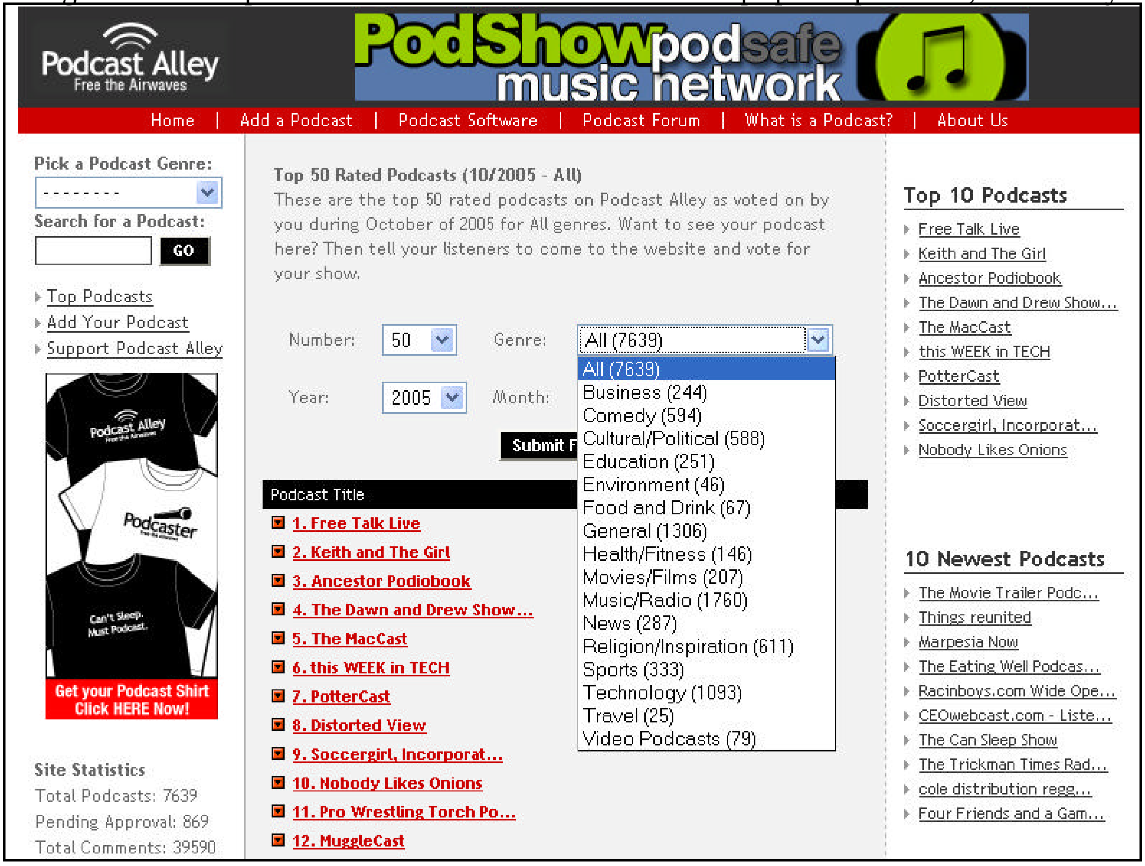
©2005 PodFly.com Podmapper was created by Jordan Lyall for PodFly.com Thanks to the Google Maps API, Ben Weintraub, Steve Moise, and the Google Maps via PHP/Mysql Tutorial.

PodMapper is powered by Jordan Lyall Web Solutions

Figure 2

Not only is the inclusion of a user's Latitude and Longitude a creative use of technology, the concept might have some appeal in social networking applications where users might be hesitant to enter some types of personal information such as the city where they live.

Yet another approach is demonstrated on Podcast Alley, where users who wish to add podcasts are asked to choose a Genre (such as Sports or Music/Radio), along with other metadata similar to what is available on other podcast sites. [5] The Genre element is particularly significant on Podcast Alley, because users have the opportunity to vote for podcasts. As shown in Figure 3, any visitor to Podcast Alley can navigate to the Top Podcasts screen and view the most popular podcasts, filtered by Genre. [6]



Podcast Alley
Free the Airwaves
PodShow podsafe music network
Podcast Alley
Free the Airwaves
Podcaster
Can't Sleep. Must Podcast.
Get your Podcast Shirt Click HERE Now!

Home | Add a Podcast | Podcast Software | Podcast Forum | What is a Podcast? | About Us

Pick a Podcast Genre: --------

Search for a Podcast: GO

- Top Podcasts
- Add Your Podcast
- Support Podcast Alley

Site Statistics
Total Podcasts: 7639
Pending Approval: 869
Total Comments: 39590

Top 50 Rated Podcasts (10/2005 - All)

These are the top 50 rated podcasts on Podcast Alley as voted on by you during October of 2005 for All genres. Want to see your podcast here? Then tell your listeners to come to the website and vote for your show.

Number: 50 Genre: All (7639)

Year: 2005 Month:

Submit F

All (7639)
Business (244)
Comedy (594)
Cultural/Political (588)
Education (251)
Environment (46)
Food and Drink (67)
General (1306)
Health/Fitness (146)
Movies/Films (207)
Music/Radio (1760)
News (287)
Religion/Inspiration (611)
Sports (333)
Technology (1093)
Travel (25)
Video Podcasts (79)

Podcast Title

1. Free Talk Live
2. Keith and The Girl
3. Ancestor Podiobook
4. The Dawn and Drew Show...
5. The MacCast
6. this WEEK in TECH
7. PotterCast
8. Distorted View
9. Soccergirl, Incorporat...
10. Nobody Likes Onions
11. Pro Wrestling Torch Po...
12. MuggleCast

Top 10 Podcasts

- Free Talk Live
- Keith and The Girl
- Ancestor Podiobook
- The Dawn and Drew Show...
- The MacCast
- this WEEK in TECH
- PotterCast
- Distorted View
- Soccergirl, Incorporat...
- Nobody Likes Onions

10 Newest Podcasts

- The Movie Trailer Podc...
- Things reunited
- Marpesia Now
- The Eating Well Podcas...
- Racinboys.com Wide Ope...
- CEOwebcast.com - Liste...
- The Can Sleep Show
- The Trickman Times Rad...
- cole distribution regg...
- Four Friends and a Gam...

Figure 3

The sites that have been mentioned in this paper are only a few of a growing number that syndicate podcasts, demonstrating how much interest in podcasting has grown since mid-2004 when Adam Curry wrote the first version of iPodder and notified the open source software development community of its existence [7]. According to a phone survey conducted by the Pew Internet and American Life Project between February and March of 2005, approximately 29 percent of the 22 million people who owned iPods or other MP3 players had downloaded podcasts (more than 6 million people). [8] There has been a corresponding increase in the number of podcasts that are available. On feedburner.com, for instance, the number of podcasts grew from about 200 in November 2004 to 13,782 in August 2005.11 [9]

It remains to be seen whether podcasts are a passing fad or will be part of the web landscape for some time to come. Even if podcasts are ultimately replaced by some other type of content, though, one wonders where podcasting might fit among the numerous applications and technologies that are seen as components of "Web 2.0." There is currently a very active debate on blogs and at conferences in regard to what exactly Web 2.0 means—one of the most concise definitions suggests that the Web 2.0 is where "Tim Berners-Lee's Semantic Web meets social software." [10] According to another definition, Web 2.0 includes five characteristics.12 Podcasting appears to fall well within the boundaries of at least two of these characteristics: 1) As a "Point of Presence" that exposes syndicated content, and; 2) As an entity that is based on openly accessible microcontent13 that leverages an infrastructure that is "open, decentralized, bottom-up and self-organizing." [11]

In closing, a brief overview of the history of RSS helps illustrate the nature of the difference between focusing on social software and content syndication (characteristics of Web 2.0), in contrast to focusing on building a broad foundation that can make possible the development of "semantically aware" applications (one of the characteristics of the Semantic Web). To briefly summarize, RSS developed along two main branches. The "Really Simple Syndication" branch, on which podcasting is based (RSS 0.91, 0.92, and now 2.0), features very simple and lightweight XML. By way of contrast, the "RDF Site Summary" branch (RSS 0.9 and 1.0) introduces greater complexity (such as XML namespaces), makes it possible to exchange structured metadata, and offers a modular extension mechanism. [12] In effect, by choosing to adopt RSS 2.0, syndicators aligned themselves with the Web 2.0 movement. It remains to be seen whether some syndicators will choose to make a richer set of metadata available to podcast producers by choosing to adopt RSS 1.0 and the rich set of extensions that it provides.

References:

1. Scott Golder and Bernardo Huberman, The Structure of Collaborative Tagging Systems, Information Dynamics Lab, HP Labs, p. 1. Also http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf
2. podcast.net. Add Your Podcast. http://www.podcast.net/addpodcast
3. podcast.net. What's New > Podcast.net…now with TAGS! http://www.podcast.net/news/00003
4. Podmapper. http://www.podfly.com/map/
5. Podcast Alley. Add a Podcast. http://podcastalley.com/add_a_podcast.php
6. Podcast Alley. Top 50 Rated Podcasts. http://podcastalley.com/top_podcasts.php?num=50
7. Wikipedia. Podcasting. http://en.wikipedia.org/wiki/Podcasting
8. Lee Rainie and Mary Madden, Data Memo RE: Podcasting, Pew Internet and American Life Project, April 2005, p. 2. Also http://www.pewinternet.org/pdfs/PIP_podcasting.pdf

[11] As of October 15, 2005, there were 21,285 podcasts available on feedburner.com. See http://www.feedburner.com/fb/a/home. Furthermore, In September 2005, Duke University hosted a Podcasting Symposium, where a wide range of topics related to podcasting were discussed. http://isis.duke.edu/events/podcasting/casts.html

[12] The five characteristics are The Web as Platform, Point of Presence, Microcontent-based, Second order content or metacontent, and Metaweb. http://phaidon.philo.at/martin/archives/000298.html

[13] According to Microcontent News, microcontent is User Generated Content that is written by people who are not professional journalists. http://www.microcontentnews.com/entries/20050420-17833.htm

9. Sheri Crofts, Jon Dilley, Mark Fox, Andrew Retsema, & Bob Williams, Podcasting: A new technology in search of viable business models, First Monday, 10 (9), http://www.firstmonday.org/issues/issue10_9/crofts/index.html

10. Kairosnews. A Weblog for discussing Rhetoric, Technology, and Pedagogy. Web 2.0: The New Buzzword in Internet Technology. http://kairosnews.org/node/4431

11. mediatope II. a cumulative Web 2.0 definition … http://phaidon.philo.at/martin/archives/000298.html

12. Tony Hammond, Timo Hannay, and Ben Lund. The Role of RSS in Science Publishing: Syndication and Annotation on the Web, D-Lib Magazine, 10 (12), December 2004.
http://dlib.org/dlib/december04/hammond/12hammond.html

PhD thesis title: AN APPROACH TO ONTOLOGY CONSTRUCTION AND ITS APPLICATION TO COMMUNITY PORTALS

Dissertation performed at the Leopold-Franzens University of Innsbruck

Author: Anna V. Zhdanova
email: anna.zhdanova@deri.org, homepage: http://homepage.uibk.ac.at/~c703261

Zhdanova A., (2005), An approach to ontology construction and its application to community portals, AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 61-62

Abstract:

The goal of the work reported in the thesis is to identify current limitations of community portals, introduce community-driven ontology management as a new approach to ontology construction and demonstrate the added value to community portals of being community-driven.

Three main parts of this work are (i) development of a framework allowing and motivating collaborative ontology construction and reuse for the final user (a person and a community), (ii) building a prototype on the basis of this specification, namely the People's portal, and (iii) application of the developed infrastructure to scenarios on Semantic Web community portals with involvement of real users.

The scope of work on the framework for community-driven ontology management is in enrichment with community-supporting features the established practices for ontology management in the areas of ontology development and population, storage, alignment and versioning. The objective of community-driven ontology management is to provide means and motivations for a large number of users to "weave" and adopt the Semantic Web.

The People's portal infrastructure allows end users to define the content structure (i.e., develop ontologies), populate ontologies and define the ways the content is managed on Semantic Web community portals where the People's portal infrastructure is applied. Content management features on the People's portal include ontology alignment support, personalization support (at the personal and community levels) and dynamic reaching of a consensus on the basis of heterogeneous ontologies.

The People's portal was deployed as a part of an intranet at DERI – Digital Enterprise Research Institute [1] and as an extension to the portal of a Semantic Web community (KnowledgeWeb network of excellence) [2]. Ontology matching part of the People's portal was deployed as a Web application open to everybody on the Web [3]. In all the empirical studies, the community's response and behavior were observed.

In conclusion, comparison to the functionalities of the existing (Semantic) Web community environments and the empirical results prove feasibility and the advantages of community-driven ontology management. Empirically, communities were capable to introduce on the community portals such ontology items as Classes, Subclasses, Properties, Instances, ontology mappings, and reuse these items afterwards.

[1] Zhdanova, A.V., Krummenacher, R., Henke, J., Fensel, D. "Community-Driven Ontology Management: DERI Case Study". In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 19-22 September 2005, Compiegne, France, IEEE Computer Society Press, pp. 73-79 (2005).

[2] knowledgeweb on the people's portal: http://people.semanticweb.org

[3] OWL Ontology Aligner: http://align.deri.org

The Development of the Semantic-based Intelligent Interactive Knowledge Architecture for the Health Education: Taking the Health Education for the Diabetics as an Example

An-Jim Long, Polun Chang
Institute of Public Health & Health Informatics and Decision Making
National Yang-Ming University, Taiwan,

Reference to this article should be made as follows:
Long A-J., Chang P.., (2005), The Development of the Semantic-based Intelligent Interactive Knowledge Architecture for the Health Education: Taking the Health Education for the Diabetics as an Example,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 62-64

Abstract

Semantic Web as an emergent technology has not been developed to become a killer application in healthcare. These researches created an integrated architecture with improved human interface, semantic web, web services and the website ranking services and expect this architecture can assist diabetic to obtain accurate knowledge and improve the self-care process.

Introduction

The World Wide Web is no doubt a significant resource to obtain patient education information. Most of the main healthcare providers in Taiwan have implemented their health information website. Nevertheless, the problem of difficulty for people to find the required health website, unfamiliar with the services provided by the website, low usability, and low credibility of the healthcare website are still the obstacles to promote the use of the web-based patient education.

Diabetes is a growing and massive silent epidemic that has the potential to cripple health services in all parts of the world. The latest World Health Organization estimate for the number of people with diabetes, world-wide, in 2000 is 177 million. This figure is likely to more than double, to reach 366 million, by 2030. Most of this increase will occur as a result of a 150% rise in developing countries. Because of its chronic nature, the severity of its complications and the means required to control them, diabetes is a costly disease, not only for the affected individual and his/her family, but also for the health authorities. Hense, diabetic patient education is needed for taking care of diabetic with diet, medicine, and exercise control to improve quality of life.

There are almost one million people in Taiwan are diabetic. According to the domestic research, 80% of complications can be avoided if the diabetic can take medical treatment in the early stage and take care of himself with appropriate medical knowledge. Hence this project took the health education for the diabetics as an example. This research developed an applicable architecture that integrates multiple information technologies. We aimed to unify the technology of user interface, semantic web [1], intelligence agent, web services with healthcare website ranking in a patient demand oriented aspect, to build semantic-based intelligent interactive knowledge architecture for the health education.

Materials and Methods

This research utilized Unified Medical Language System (UMLS) [2] as core ontology of Semantic-based Diabetes Content, we firstly collect frequent asked question and answer, and general information of diabetes from 181 websites cited from attended organizations to an official Department of Health (DOH) website competition in healthcare [2]. All the web pages have been extracted, truncated and compile statistics, then map all the collected medical terms to UMLS and translated into Chinese. The detailed architecture of this research is shown as Figure 1.

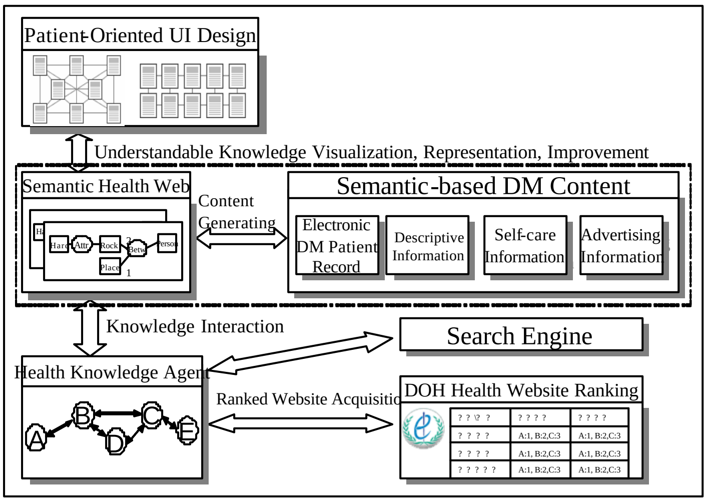


Figure 1 Interacted Semantic-based Health Knowledge Architecture for Diabetes

Result

Preliminary implementation was completed in order to test the feasibility of the architecture. In the first place, we collected frequent asked question and answer, and general information of diabetes Figure 2 and compiled into protégé and generated the ontology map as Figure 3. Then we implemented prototypal search interface for user to search by keywords of interest or by ontology map left side (see Figure 4). The system automatically matches keywords with all metathesaurus of UMLS related to the concept of the keyword from our predefined ontoloty. The system shows users the hit websites and their ranking and evaluated information from DOH so users can judge credibility by themselves. Also we implemented a prototypal registration system Figure 5 for websites or keywords not listed in our ontology map so the Diabetes knowledge can grow up by further maintainance. The perameters of the registration service includes the URL, RDF document, title, organization, general information of the website.

Figure 2 Classification of general asked information of Diabetes (in Chinese)

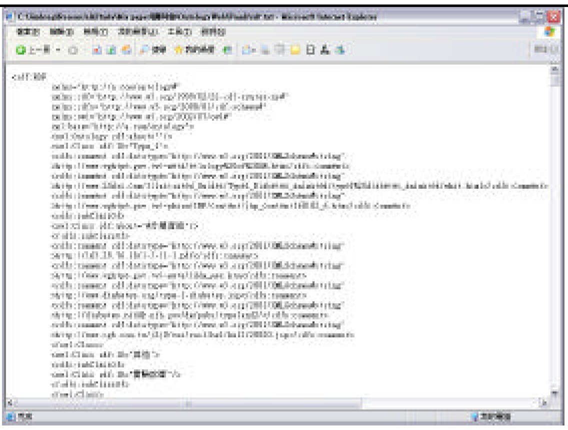


Figure 3 Generated RDF of Diabetes classification

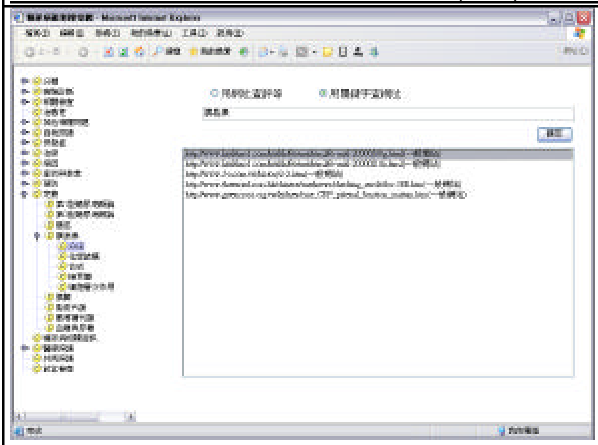


Figure 4 Prototypal search interface

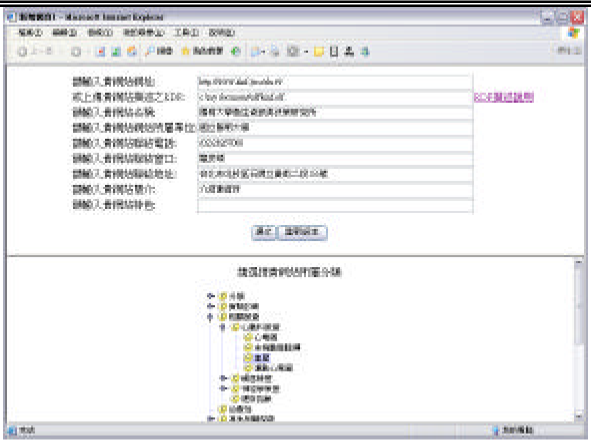


Figure 5 New Diabetes website registration

Discussion

The prototype shows feasibility for system integration of human interface, semantic web, web services and the website ranking services for the health education. The system is expected to assist diabetic to obtain accurate knowledge and improve the self-care process. Further implementation and evaluation should be conducted. We expect this architecture can solve the problem of usability and credibility for finding healthcare information on the web and provide a blue print for related reference.

Reference

[1]T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific Am., vol. 284, no. 5, May 2001, pp. 34–43.

[2] Healthcare Website Competition, Department of Health (in Chinese) http://awards.doh.gov.tw/

[3] National Library of Medicine. 1999. UMLS Knowledge Sources 10th Ed. http://umlsks.nlm.nih.gov

Protégé, tool for ontologies editing. http://protege.stanford.edu/

eQ: Better Personalized Adaptation on the Semantic Web and Grid

[1]Violeta Damjanovic
[1]GOOD OLD AI Research Group, FON, University of Belgrade,
POB 52, Jove Ilica 154, 11000 Belgrade, Serbia and Montenegro
vdamjanovic@gmail.com, http://vdamjanovic.bravehost.com

Reference to this article should be made as follows:
Damjanovic V., (2005), eQ: Better Personalized Adaptation on the Semantic Web and Grid,
AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 65-70

Abstract. In this paper, we have presented one approach to the personalized adaptation in the Semantic Web and Grid environment. We have focused on using multi-agent systems in the e-learning environments with the aim to facilitate the adaptation processes, as well as to enable better way of achieving the e-learning goals. According to the fact that we are dealing with the stereotypes of e-Learners, having in mind emotional intelligence concepts to help in adaptation to the e-Learners real needs and known preferences, we called this system eQ.

Introduction

Nowadays, the Adaptive Hypermedia (AH) systems have an important role in doing hypermedia exploration, as well as designing of user interfaces. Development of the AH systems can be roughly divided into three generation of research [1]:

- *the first generation* describes pre-Web hypertext and hypermedia (before 1996);
- *the second generation* is devoted to the Web-based AH systems (between 1996 and 2002);
- *the third generation* explores advanced developing technologies for “open corpus AH” and developing a component-based architecture for assembling adaptive Web-based educational systems (since 2002).

The specific category of the AH system represents the Adaptive Educational Hypermedia Systems (AEHS) that can be used in e-learning and training processes on the Web. In the same time, such kind of adaptive systems consider both *contextual information* about users, as well as educational *content information* about different learning and training materials. *Context management* includes user modeling, enabling reusability and sharing of the user model by various adaptive applications and user devices. In other words, context management can be used to collect, collate and process context information about users. The goal of this part is to design and implement a mechanism by which context information can be updated and distributed. Context manager must be able to detect modification and addition of user’s characteristics; it must have location awareness module, as well as a component that provides data about enterprise policy. *Content management* maintains the domain model (learning objects (LOs) with metadata, semantic concept networks/ontologies) and supports the authoring process (separation of content and layout, their reusability, semi-automatic annotation).

In this paper, we explore a component-based architecture for assembling both the AEHS and multi-agent systems (MAS) with the aim to improve the adapted e-learning processes in the Semantic Web and Grid environment. In addition, we are interested in how to manage teaching resources when the e-Learners have different emotions, perceptions, reactions. According to the fact that we are dealing with the user’s stereotypes, having in mind emotional intelligence (EQ) concepts to help in adaptation to the user’s real needs and known preferences, we called this system eQ. That stands for using EQ concepts on the Web (electronic EQ). We represent the way of measuring user’s EQ based on using MAS as a distributed test-sensor systems for observing and testing users, as well as for ontological representation of various user’s emotional intelligence facts. These results represent instances from the ontology for adaptation. In addition, as an example of professional training domain we represent the ontology ACCADEMI@VINCIANA with the aim to realize a new way of art education based on using the Semantic Web technology.

Adaptive Educational Hypermedia Systems and Multi-Agent Systems

The AEHS and MAS both can be useful in any application area where a hypermedia system, ontologies, LOs, and other learning materials are expected to be used by e-Learners with different goals, preferences and knowledge. They both can be used to support agent's ability to learn, based on experience, and to adapt their knowledge to make rational conclusions about e-Learners and their learning needs in different collaborative environments.

Role of AEHS

Adaptive education hypermedia system represents [2]: "… a quadruple

(DOCS, UM, OBS, AC) (1)

- DOCS (***DOC***ument *S*pace): a finite set of first-order logic (FOL) sentences with constant symbols for describing documents (knowledge concepts), and predicates for defining relations between these constant symbols;
- UM (***U***ser ***M***odel): a finite set of FOL sentences with constant symbols for describing users (user groups), and user characteristics, as well as predicates and formulas for expressing whether a characteristic applies to user;
- OBS (***OBS***ervation): a finite set of FOL sentences with constant symbols for describing observation, and predicates for relating users, documents / concepts, and observations;
- AC (***A***daptation ***C***omponent): a finite set of FOL sentences with formulas for describing adaptive functionality."

We use the component-based definition of AEHS [3], shown in (1), to implement the sets, functions, and operations of the proposed FOSP (Filter-Order-Select-Present) adaptation method.

Role of MAS

MAS are widely seen as the most promising technology for developing complex distributed software systems in the years to come. The most important reasons for using MAS when designing a system can be described as follow [4]:

- domains with different goals and information;
- a method for parallel computation by assigning different tasks (abilities) to different agents;
- full robustness of system and applications;
- an easy way to add new agents (scalability);
- the modularity of MAS and simpler programming;
- exploring of intelligence according to the need to deal with social interactions.

eQ Agent System – Conceptual Design

There are several key paradigms being used in conceptual design of the eQ agent system [5]:

- using BDI (Belief-Design-Intention) rational model for implementing MAS;
- introducing the notion of EQ with the aim to improve better personalized adaptation;
- introducing the FOSP method with the aim to specify adaptation strategy.

We give a brief explanation of each of its components in the following subsections.

BDI Paradigm

The BDI paradigm is based on the early philosophical work of Bratman about rational action theory [6]. Their primary contribution is in integrating the various aspects of BDI agent research, such as theoretical foundation from both a quantitative decision-theoretic perspective and a symbolic rational agency perspective, to the system implementation and building applications that are used a practical BDI architecture.

We have chosen to use the Jadex platform for implementing the eQ agent system. Jadex supports the development of rational agents on top of the FIPA-compliant JADE platform [7]. Jadex BDI model considers three types of attitudes of agent rational behaviors: belief (goals), desire and intention. Beliefs represent the information about agent's internal, as well as external states, and provide domain-dependent abstraction of entities. The motivational attitudes of agents are captured by goals, which represent a central concept of the Jadex BDI architecture. And, last but not least, plans are the means by which agents achieve their goals.

Emotional Intelligence on the Semantic Web

EQ represents an essential part of effective communication, adaptability, and personal satisfaction, especially in the field of education. The process of teaching and learning represents a highly social and emotional activity. Hence, we can conclude the cognitive progress depends on:

- e-Learner's psychological predispositions, such as their interest, confidence, sense of progress and achievement,
- e-Learner's social interactions with teachers who provide them with both cognitive and emotional support.

Our approach for achieving personalized adaptation is based on using EQ through the AEHS, in the Semantic Web and Grid environment. In order to explain using EQ for adaptation, we modeled stereotypical models of e-Learner's individual traits, such as [3]:

- personality factors (extrovert/introvert),
- cognitive factors,
- learning styles; and
- personality types (stereotypes): 1) Conventional personality, 2) Social personality, 3) Investigative personality, 4) Artistic personality, 5) Realistic personality, and 6) Enterprising personality.

FOSP – an Adaptive Educational Strategy

FOSP method, originally proposed in [8], stands for Filter, Order, Select, and Present operations of a novel method for specification of adaptation strategies in AEHS. The main idea is to separate the partial results produced by different authors in such a way that they can be reused. FOSP method consists of the following three levels [8]:

- *Level 1 - Operations:* filter, order, select, present;
- *Level 2 - Functions:* weight (the relevancy of the pedagogical role for the learning style), sequence (the presentation order of the role for the learning style), alternative (the relevancy of the media type for the learning style), threshold (the threshold for the object display based on the learning style), granularity (the max number of objects presented for the context);
- *Level 3 - Sets:* role, style, media, and context.

eQ Agent System – Knowledge Bases

It consists of two ontologies: ACCADEMI@VINCIANA ontology and ontology for adaptation.

Domain Ontology - ACCADEMIA@VINCIANA

ACCADEMI@VINCIANA ontology has several dimensions concerned professional training's intentions:

- it describes three fundamental painting components and theirs role in painting construction;
- it observes fundamental aspects for analyzing painting methods and techniques;
- it can be divided into the following two categories of trainings [9]:
  - trainings made by using physical methods (dermatoscope, microscopes, X-rays, UV exploring…);
  - trainings made by using chemical methods (microchemistry approach with pigments identification, emission spectral analysis, the iodine probe, DBA…).

In other words, ACCADEMIA@VINCIANA ontology contains of three main parts with the knowledge with the aim to support the following:

- learning about fine art painting methods and materials (education: painting, conservation treatments, preventive conservation strategies, restoration, reproduction);
- doing virtual experiments about painting methods and materials (education, classical painting technology analysis, painting damage diagnosis);
- doing online experiments (author identification, original expertise, fraud investigation).

Ontology for Adaptation

Ontology for adaptation includes context information about users (individual traits, such as: personality factors, cognitive factors (perceptual processing, phonological awareness, ability to maintain attentive focus), learning styles (moving, touching, doing, auditory, visual), personality types (conventional, social, investigative, artistic, realistic, and enterprising personality), as well as information about user's devices). This ontology is based on using the IEEE PAPI (Public And Private Information) Standard, which represents [10]: "a data interchange specification that describes learner information for communication among cooperating systems." The IEEE PAPI Standard is extended in the parts that are related to the learner *preference information* (IEEE 1484.2.24), as well as the learner *portfolio information* (IEEE 1484.2.26) (shown in Fig. 1, as well as Fig. 2). These extensions are made with the aim to enable using EQ concepts during the learning processes on the Web.

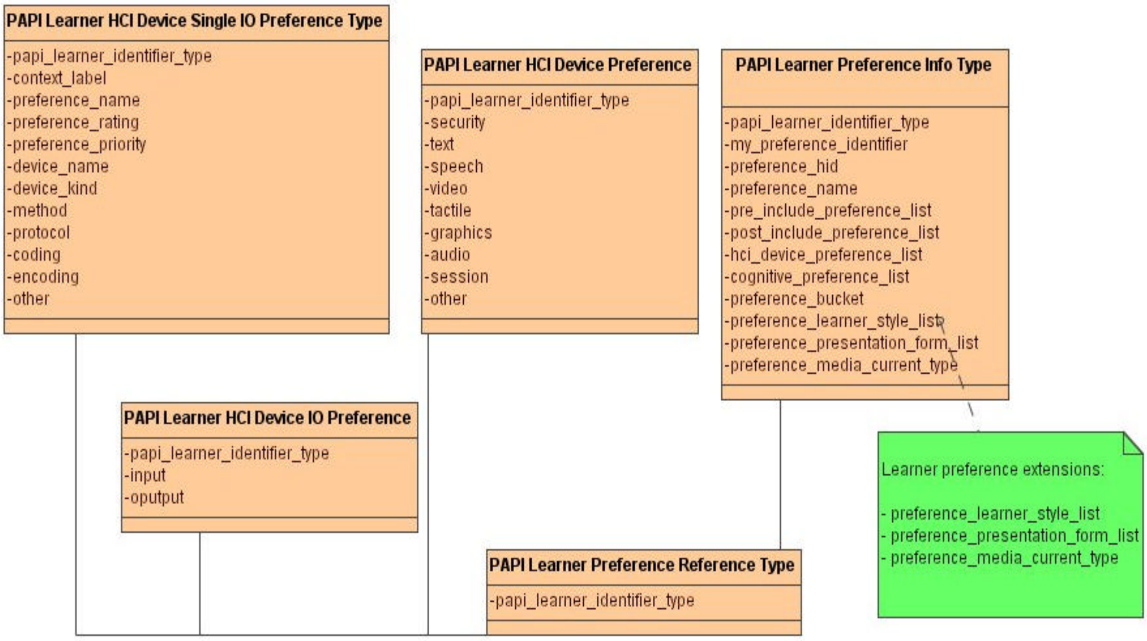

PAPI Learner HCI Device Single IO Preference Type
-papi_learner_identifier_type
-context_label
-preference_name
-preference_rating
-preference_priority
-device_name
-device_kind
-method
-protocol
-coding
-encoding
-other
PAPI Learner HCI Device Preference
-papi_learner_identifier_type
-security
-text
-speech
-video
-tactile
-graphics
-audio
-session
-other
PAPI Learner Preference Info Type
-papi_learner_identifier_type
-my_preference_identifier
-preference_hid
-preference_name
-pre_include_preference_list
-post_include_preference_list
-hci_device_preference_list
-cognitive_preference_list
-preference_bucket
-preference_learner_style_list
-preference_presentation_form_list
-preference_media_current_type
PAPI Learner HCI Device IO Preference
-papi_learner_identifier_type
-input
-oputput
PAPI Learner Preference Reference Type
-papi_learner_identifier_type
Learner preference extensions:
- preference_learner_style_list
- preference_presentation_form_list
- preference_media_current_type

Figure 1. An extension of the IEEE PAPI learner preference information (IEEE 1484.2.24)

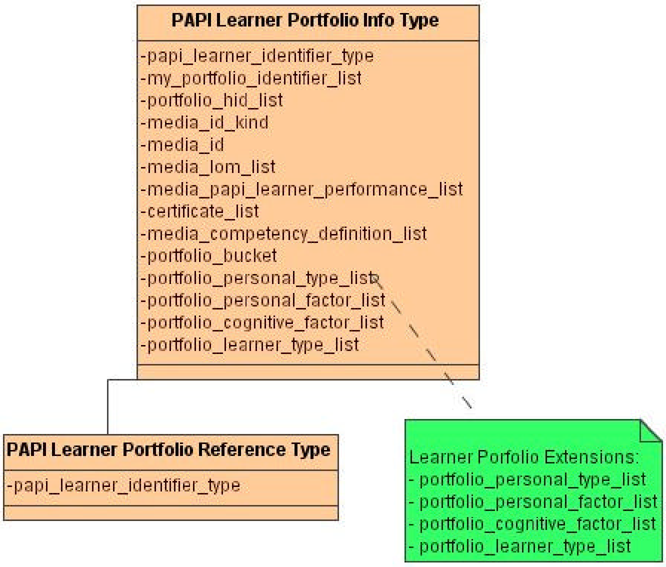

PAPI Learner Portfolio Info Type
-papi_learner_identifier_type
-my_portfolio_identifier_list
-portfolio_hid_list
-media_id_kind
-media_id
-media_lom_list
-media_papi_learner_performance_list
-certificate_list
-media_competency_definition_list
-portfolio_bucket
-portfolio_personal_type_list
-portfolio_personal_factor_list
-portfolio_cognitive_factor_list
-portfolio_learner_type_list
PAPI Learner Portfolio Reference Type
-papi_learner_identifier_type
Learner Porfolio Extensions:
- portfolio_personal_type_list
- portfolio_personal_factor_list
- portfolio_cognitive_factor_list
- portfolio_learner_type_list

Figure 2. An extension of the IEEE PAPI learner portfolio information (IEEE 1484.2.26)

Practical Results

When user starts application for fine art training and learning, this application automatically recognizes both user's individual traits and user devices on which the application is executed. All information about user's characteristics is contained within the ontology for adaptation (as context information). An eQ Context Manager Agent finds all context facts about observed user and sends these results to the eQ FOSP Manager Agent with the aim to perform personalized adaptation and present adapted content information to the user. eQ Context Manager Agent has a location awareness module, which role is to support changes in user device attribute values. For example, user starts using training application on laptop, and then migrates to a PDA. It means that the content information have to be additionally adapted and the eQ FOSP Manager Agent has to perform some kind of filtering which shrinks the images to a size that fits nicely on the screen of the PDA.

For example, the application for fine art professional training recognizes user with the "*artistic personality*" (personality type), "*introverted perception*" (personality factor), "*visual*" learning style, which user type is "*expert*" that explore "*art fraud*" and uses "*PDA*" (user device). Now, it should be done content adaptation for that user, what is the task of the eQ FOSP Manager Agent. eQ FOSP Manager Agent supervises four eQ agents that one after the other performs the main operations of the FOSP adaptive strategy, such as: Filter, Order, Select, and Present.

Fine art trainings based on using physical methods could be realized by using different optical tools (microscopes, dermatoscopes, and cameras). In the case of the above explained user, the eQ Present Agent brings some physical methods for fine art trainings as a result. Actually, it means that the eQ Present Agent offer trainings by using X-ray, UV exploring, as well as F-exploring, as training methods that could be used to achieve art fraud investigation. All points of the considered eQ agent system are shown in Fig. 3.

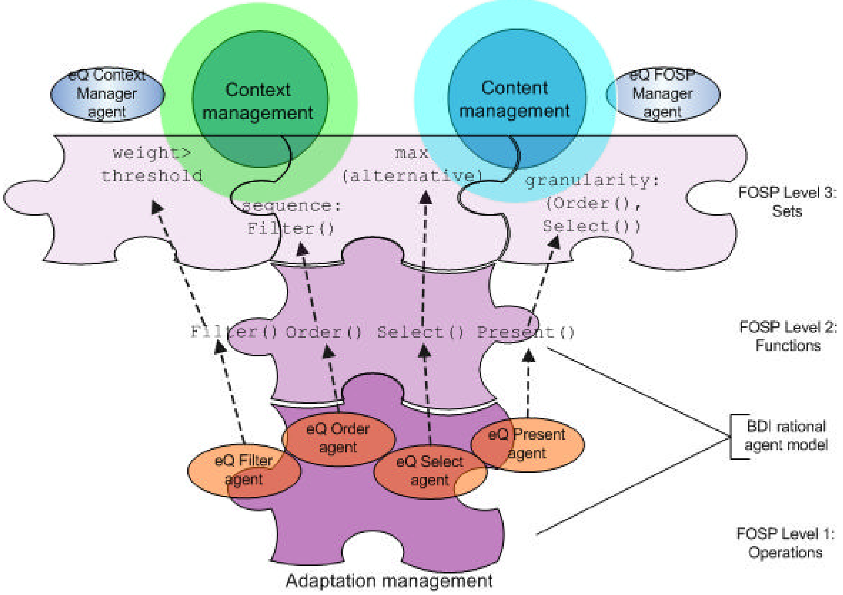


Figure 3. eQ agent system uses BDI agent's reasoning engine

Conclusion

The impacts of many technology trends in further development of the AH systems can be considered as developing comprehensive frameworks for adaptive web-based education, developing more intelligent educational material by using learning object metadata (LOM), exploring the ideas of the Semantic Web for content representation and resource discovery.

In this paper, an example of fine art professional training is used to illustrate the potential benefits of using personalized adaptation in training environment. As the potential benefits, we can mention the following:

- Adaptation by focusing on the main subjects from the domain of artistic training (painters, conservators, restorers, technologists, fraud investigators);
- Using all available resources (learning materials, training devices) wherever the user is physically located;
- Analyzing generated results and deciding about using preventive painting strategies;
- Collaboration with the aim to achieve the original expertise and art fraud investigation.

In addition, we explore using EQ concepts in the Semantic Web and Grid environment. The benefits of taking proposed approach are numerous, and can be characterized as follows:

- *collaboration* with other students, teachers, tutors, experts;
- *knowledge-based*: it includes domain knowledge representation in the form of ontologies, as well as knowledge about learner and his/her social and emotional context;
- *ubiquitous*: the capability to support multiple pedagogical models and to automatically adopt them.

Acknowledgements

This research is linked to the Network of Excellence (NoE) in Professional Learning – ProLearn, work package WP1: Adaptive Personalized Learning.

References

[1] Brusilovsky, P., 2004. Adaptive Educational Hypermedia: From Generation to Generation, In *M. Grigoriadou et al. (Eds): Proceedings of the 4th Hellenic Conference with International Participation in "Information and Communication Technologies in Education", New Technologies Publications*, pp. 19-33.

[2] Henze, N., and Nejdl, W., 2003. Logically Characterizing Adaptive Educational Hypermedia Systems, In *Proceeding of the International Workshop of Adaptive Hypermedia and Adaptive Web-based Systems (AH 2003)*, Budapest, Hungary, pp. 15-28.

[3] Damjanovic, V., Kravcik, M., Gasevic, D., 2005. eQ Through the FOSP Method, In *Proceeding of the ED-MEDIA 2005 World Conference on Educational Multimedia, Hypermedia and Telecommunications,* Montreal, Canada.

[4] Stone, P., 1997. Multiagent Systems. [Online] Available at:http://www-2.cs.cmu.edu/afs/cs/usr/pstone/public/ papers/97MAS-survey/node2.html

[5] Damjanovic, V., Kravcik, M., Devedzic, V., 2005. eQ: An Adaptive Educational Hypermedia-based BDI Agent System for the Semantic Web, In *Proceeding of the 5th IEEE International Conference on Advanced Learning Technologies (ICALT 2005),* Kaohsiung, Taiwan, pp. 421-423.

[6] Bratman, M.E., 1987. *Intentions, Plans, and Practical Reason,* Harvard University Press, Cambridge, MA.

[7] Braubach, L., Pokahr, A., Moldt, D., Lamersdorf, W., 2004. Goal Representation for BDI Agent Systems, In *Proceeding of the Second International Workshop on Programming Multiagent Systems (PROMAS-2004).*

[8] Kravcik, M., 2004. The Specification of Adaptation Strategy by FOSP Method, In *Proceedings of AH2004 Conference,* Eidhoven, Netherland.

[9] Damjanovic, V., Kravcik, M., Devedzic, V., 2005. An Approach to the Realization of Personalized Adaptation by Using eQ Agent System, In *Proceedings of UM'2005 Workshop on Personalized Adaptation on the Semantic Web (PerSWeb'05),* Edinburg, Scotland, UK.

[10] IEEE 1484.2.1, 2001.Standard for Learning Technology — Public and Private Information (PAPI) for Learners (PAPI Learner) — Core Features.

A General Approach to Metadata-based Web Content Filtering

Elisa Bertino[a], Elena Ferrari[b], Andrea Perego[c]
[a] *CS and ECE Departments, Purdue University, IN, USA*
bertino@cerias.purdue.edu
[b] *DSCPI, Università degli Studi dell'Insubria, Como, Italy*
elena.ferrari@uninsubria.it
[c] *DICo, Università degli Studi di Milano, Italy*
perego@dico.unimi.it

Reference to this article should be made as follows:
Bertino E., Ferrari E., Perego A., A General Approach to Metadata-based Web Content Filtering, AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 70-76

Abstract. Web content filtering is a means to make users aware of the characteristics of Web resources, in order to verify whether they satisfy given parameters concerning their 'appropriateness' with respect to their content, their authoritativeness /reliability, and so on. Currently, Web filtering is enforced for various purposes (among which minors' protection is one of the most relevant), by using diverse strategies and different software tools. One of the major research issues in this area is the definition of a unified filtering framework for all the possible application domains. This has the advantage of enforcing interoperability among the different filtering approaches and the systems based on them. In this paper, besides providing an overview of the main Web filtering strategies and application domains, we illustrate a general-purpose filtering model, addressing the drawbacks of the existing approaches, and its two implementations, carried out in the framework of the EU projects EUFORBIA and QUATRO.

1. Introduction

The Web makes publicly available a huge amount of multimedia data, which are heterogeneous not only in their content but also in their quality. This results in making very difficult for users, on one hand, to identify the resources they are looking for, and, on the other hand, to verify, e.g., the authoritativeness and/or reliability of the information provided. Search engines try to address these issues, but, despite the enforced indexing and ranking techniques grant the relevance of the returned search results, they can provide an assessment of resources' content and quality which is only probabilistic. This approach cannot be applied when, e.g., Web information is accessed by category of users' which should be protected (e.g., children) or when it is used for delicate purposes (e.g., imagine a medical Web site providing unreliable information concerning diseases and their therapy). As a result, as long as it is not possible to easily verify the content and the quality of resources, the Web will remain a potentially unsafe information space.

Since the Internet became accessible to everyone, such drawback has been addressed by enforcing filtering mechanisms in the Web. More precisely, Web resources' characteristics are evaluated with respect to given requirements, which may correspond either to the characteristics and / or preferences of end users, or to a set of predefined content / quality constraints. Depending on the results of such evaluation, different actions may be performed: for instance, the user may be notified of the in / appropriateness of the requested resource, or the access to a resource considered as inappropriate may be denied.

So far, Web content filtering has been applied to application domains concerning users' protection, privacy, and quality trustmarks, which make use of diverse strategies.

For instance, users' protection has been enforced by adopting two main approaches. In the former, rating services generate lists of 'good' and 'bad' Web sites, which are used by filtering systems to block the access to resources considered as inappropriate. Yet, this approach has the drawback of enforcing a too restrictive filtering of the Web. In fact, it is not possible to rate with the required accuracy all the online resources without the supervision of humans, and, as a result, only an insignificant part of the Web is accessible for anyone using such services. In order to overcome such drawbacks, another approach has been proposed by the W3C, based on the PICS standard [6], which defines a format for *content labels* (i.e., a set of metadata describing the characteristics of a resource) and how they can be distributed and retrieved. According to this strategy, filtering systems evaluate whether a resource is appropriate or not depending on the metadata contained in the label and on the filtering settings specified by the user or a supervisor. PICS has been the former attempt to provide machine understandable descriptions of Web resources: even though it did not obtain a great success and diffusion, the PICS approach has been considered the most suitable to the requirements of Web filtering and to the heterogeneity of Web users. For this reason, extensions have been proposed to the PICS standard, in order to allow the specification of filtering rules (i.e., PICSRules [8]), and to make PICS compliant with the Semantic Web technologies [9].

Another metadata-based approach is the one provided by the P3P W3C standard [11], which allows the enforcements of negotiation procedures between users and Web services. In this case, users are informed about which of their personal data are required by a Web service, and how they will be used. P3P defines not only the format, distribution, and retrieval of labels describing the 'privacy policies' of Web services, but also how users' preferences (referred to as *rulesets*) can be expressed and exploited by user agents [10]. Another interesting feature of P3P, formerly introduced by the PICSRules proposal, is the possibility of representing and distributing predefined profiles, in order to simplify to users the task of specifying their preferences.

Despite the advantages of labeling resources, application domains different from users' protection have not adopted such practice. This applies, for instance, to trustmark agencies, which certify the 'quality' of resources with respect to a particular use. In these cases, quality may concern the authoritativeness / reliability of the provided information, the security of a Web service, the conformance with the W3C guidelines for Web accessibility [12], and so on. Web sites satisfying the quality requirements stated by a trustmark agency can display an icon in their Web pages. Moreover, some trustmark agencies store the list of certified Web sites in their database, so that the user can have a feedback from the certification authority about the trustworthiness of the trustmark, but this is not a common practice. The main drawback of this approach is that the lack of a machine-understandable description of resources does not allow users to have an evaluation of resources' quality tailored to their preferences, according to an approach similar to PICSRules and P3P.

To summarize, independently from the application domain, Web content filtering is carried out according to either a list-based or a metadata-based approach. Filtering systems based on content labels, besides introducing strategies which are complaint with the Semantic Web technologies, allow the enforcement of filtering mechanisms more sophisticated and flexible than those supported by white / black lists. Nonetheless, the current metadata-based filtering systems have two main drawbacks. The former is that they adopt diverse annotation formats and evaluation strategies, which, as a consequence, require different software tools to perform filtering. Thus, what is lacking is a general and formal representation of Web filtering, which defines its main constructs independently from its possible application domains. The latter drawback is that users' characteristics are not taken into account in the specification of policies. In fact, filtering policies (even filtering policy templates, as those provided by PICSRules and P3P) are associated with a specific user or, at most, to set of users specified explicitly. Nonetheless, whether a resource is or is not appropriate depends on the characteristics of users, and not on their identity. Note also that users' characteristics may be useful for institutional users (such as school, libraries, universities) in order to optimize the task of policy specification and management. For instance, if a library decides to prevent the access to pornographic material to users whose age is less than 16, it is simpler to express this in the policy instead of listing explicitly the set of users to whom it applies; moreover, if a new 15-aged user is enrolled, we do not have to update the policy and inserting his/her identifier.

This paper presents a general model for Web content filtering, referred to as MFM, which supports policies taking into account the characteristics of both users and resources, described by using multiple metadata vocabularies. Such model has been implemented in two different application domains, in the framework of the EU projects EUFORBIA and QUATRO, concerning, respectively, users' protection and quality assurance.

The remainder of this paper is organized as follows: Section 2 discusses the main characteristics of the proposed model, whereas Section 3 illustrates how the model has been implemented in EUFORBIA and QUATRO. Finally, Section 4 concludes the paper.

2. A General Model for Web Content Filtering: MFM

Our filtering model, referred to as MFM (Multistrategy Filtering Model), provides a general framework for denoting interaction constraints between two sets of entities (referred to as *agents*) in a given domain. MFM is not designed for a particular context, and cannot be applied directly. Rather, its aim is to define a basic data structure which can be used to generate instances, which customize their characteristics according to the application domain.

MFM is the last version of a filtering model which has been formerly developed in the EUFORBIA project, and that has been extended and generalized in order to be applied to any Web filtering application domain. Due to space constraints, here we briefly illustrate the main characteristics of MFM. For a formal and thorough description of the previous and current versions of the model, we refer the reader to [2], [3], [5], and [4].

MFM consists of four main components:

- a set of constructs for denoting agent identity and characteristics;
- a constraint specification language for defining equivalence classes of agents sharing some given characteristics;
- a set of rules for policy propagation and conflict resolution.

In MFM, agents can be considered as entities which can perform and/or are subject to a given set of operations, denoted by the existing relationships between the agent sets they belong to. Depending on whether they have an active or passive role in the interaction process, they can be grouped into two subsets: the sets of *subjects* and *objects*, respectively. Examples of agents in the filtering domain are, for instance, users and Web pages, which perform the roles of subjects and objects, respectively, in an interaction process according to which users 'access' Web pages.

Policies are rules determining which kind of interactions between agent sets can or cannot be performed. They are specified by supervisors, possibly associated with different authority levels, and denoted by a pair of subject and object sets and by the type of interaction which exists between them. Agent sets can be denoted either explicitly, by listing the agents they contain (e.g., users u_1, u_2, and u_3 cannot access Web pages wp_1 and wp_2), or implicitly, by specifying properties which agents must satisfy (e.g., users who are students, and whose age is less than 16, cannot access Web pages regarding sexual content). The interaction types depend on the application domain. They may be similar to 'traditional' access permissions, like "read", "write", "execute", or they may be different. For instance, in case the application domain concerns users' protection, we may have just one interaction type, "access", which may be granted (prevented) depending on the appropriateness (inappropriateness) of the requested resource. Another interaction type may be "notify", according to which the user is notified of the characteristics of the requested resource, without being prevented from accessing it.

In order to describe agent properties, MFM adopts an object-oriented approach relying on the notion of *agent class*. An agent class specifies a set of *attributes*, denoting agent characteristics relevant in a given domain (e.g., the age of a user, the content of a Web page). Agents are then associated with *class instances*, which are sets of attribute-value pairs denoting agent properties.

Agent classes are organized into class hierarchies, which are exploited by a policy propagation principle according to which a policy applying to an agent class is inherited by all its children. Moreover, policies are associated with a sign, according to which they can be either positive or negative. For instance, consider a policy concerning a set of users and a set of Web sites: if it is positive, the users can access the Web sites; if it is negative, they cannot. Finally, since this feature allows the specification of conflicting policies (i.e., policies on the same pair of agent sets and of the same type, but with different sign), a conflict resolution mechanism is provided in order to verify which of them is prevailing. More precisely, the prevailing policy is determined by taking into account the different components of the policy:

Supervisor authority The prevailing policy is the one specified by the supervisor with the stronger authority.

Agent specification The prevailing policy is the one with the stronger subject / object specification, i.e., the one which is more specific with respect to the class hierarchy.

Policy sign The prevailing policy is the one with the stronger sign. Which sign is prevailing depends on the application domain.

As an example, consider two conflicting filtering policies fp_1 and fp_2, specified by supervisors sv_1 and sv_2, respectively. If supervisor sv_1 has more authority than supervisor sv_2, fp_1 is the prevailing policy. Otherwise, if both supervisors have the same authority level, the prevailing policy is determined by taking into account the agent specification, according to the following rules: given two agent specification $agSpec_1$ and $agSpec_2$, concerning, respectively, an agent class $agCls_1$, and a class $agCls_2$, child of $agCls_1$, we say that $agSpec_2$ is more specific than $agSpec_1$, and then the corresponding policy is stronger. Finally, if agent specifications cannot be used to solve the conflict, the prevailing policy is determined by the sign.

3. MFM Implementations

MFM was formerly implemented in the framework of EUFORBIA, the aim of which was to address the drawbacks of the existing rating and filtering approaches by enforcing two main features:

- supporting content labels based on metadata vocabularies which allow one to provide an accurate and as far as possible objective description of Web resources;
- supporting policies taking into account users' characteristics, and not only their identity.

The outcome of the project consists of an ontology for the specification of content labels, and two prototypes, WEBFILTER and MFILTER, designed in order to satisfy the requirements of, respectively, home and institutional users.

The EUFORBIA rating approach is based on content labels (namely, the *EUFORBIA labels*) which describe the main characteristics of both the content and of the structure of a Web site. More precisely, a EUFORBIA label consists of three main sections: 'aims' (the main objectives of the site), 'properties' (a description of the Web site relevant characteristics), and 'sub-sites' (the Web site's subsections and their main characteristics). The major difference between the EUFORBIA approach and the ones currently available is that resources are annotated by using a metadata vocabulary (i.e., the EUFORBIA ontology) which models domains not limited to those considered liable to be filtered (e.g., pornography, violence, racism). As a result, EUFORBIA labels provide descriptions which are more accurate and semantically richer than those adopted by PICS-based rating and filtering services (such as ICRA), which consist of plain sets of categories. Finally, EUFORBIA labels are expressed in NKRL [14], a formal language developed by the team of the CNRS of Paris involved in EUFORBIA, and which is compliant with the standard Semantic Web languages. The advantage of NKRL is that, thanks to its syntax, complex descriptions can be specified more compactly than in RDF / OWL.

MFM has been implemented in one of the two EUFORBIA filtering prototypes, currently referred to as MFILTER, the different versions of which are described in [2], [3], [5], and [4]. MFILTER is a proxy server consisting of three main components: a database, storing all the information needed by the system, a filtering module, which is in charge of evaluating the access requests submitted by users, and a Web interface which is used for system management and users authentication. Besides supporting all the filter of our model, MFILTER enforces strategies for optimizing the filtering task and to help the administrator to specify policies tailored to the users in the system. Optimization has been obtained by adopting precomputational mechanisms, which are used in order to avoid as far as possible to carry out at runtime the retrieval and evaluation of the policies concerning the requesting user and the requested object. Moreover, the system caches the logs of the access requests submitted by users, along with the result of the request evaluation. Note that the logs indicate also when resource evaluation could not be performed (i.e., when resources are not associated with content labels, and no policy has been explicitly specified for them). This information is used by the administrator in order to refine the effectiveness of the specified policies, by verifying whether the access to resources has been correctly granted or prevented. In such cases, the administrator can, on one hand, generate content labels, which are associated with a resource and stored locally, and, on the other hand, specify new policies or modify existing ones.

The application domain of the QUATRO project is more general with respect to EUFORBIA. QUATRO aims at defining a unified platform for ‘quality’ labels and trustmarks to be associated with Web resources. Quality labels do not describe the ‘quality’ of a resource, but rather they should be used in order to establish trust between content/service providers and end users, by advising the latter about the characteristics of the resource they are accessing. Such characteristics may concern the content of a resource, the authoritativeness and/or reliability of the information it provides, the privacy policies of a Web service, and so on. Consequently, the QUATRO platform is designed for any labeling vocabulary and for any application of Web content filtering, of which users’ protection (as in EUFORBIA) is only a particular case. Finally, QUATRO addresses the need of enforcing strategies for granting labels’ trustworthiness, an issue which has been neglected in the available rating and filtering approaches.

The QUATRO platform consists of two main components:

- an RDF schema for quality labels, which can be adopted by labeling services in order to provide a standard representation of their vocabularies and of the corresponding labels;
- a set of software tools, which collaborate in order to evaluate labels and to return the results of such evaluation to end users.

The definition of an RDF schema for labels aims at overcoming the drawbacks of the current situation, where labeling authorities provide labels and trustmarks which not only are stored in diverse formats, but, quite often, they are not even machine-understandable. Such standard format is the basis on which the QUATRO software tools are built. The first tool, the QUATRO proxy (referred to as QUAPRO), is in charge of evaluating the labels associated with Web resources, and to return the results to end users through the two other tools, the metadata visualizer (ViQ) and the search engine wrapper (LADI) [7].

The current version of the QUATRO RDF schema is described in [1], and it is the outcome of the collaboration of QUATRO with the W3C Semantic Web team and organizations interested in the diffusion of quality labels and trustmarks. The schema has been adopted by the Internet rating associations partners of QUATRO (ICRA, WMA, and IQUA), and a certification agency (Segala) external to the project.[14]

Besides providing RDF classes and properties for representing labeling vocabularies and content labels, the novelty of the QUATRO rating and filtering approach is the support for different scenarios concerning label generation and distribution, and the enforcement of strategies for granting labels’ trustworthiness. In QUATRO, we distinguish between *labeling services*, which provide labeling vocabularies, and *labeling authorities*, which are in charge of generating labels. The reason why we use such distinct notions is that, differently from PICS, labels can be generated and stored either by a labeling service, or be a third-part organization, or by the labeled site itself. Which of these three possibilities is adopted depends on the type and use of labeling vocabulary. For instance, labeling services (such as ICRA), providing vocabularies for

[14] ICRA (The Internet Content Rating Association: *http://www.icra.org*) provides labels for describing resources’ content. WMA (Web Mèdica Acreditada: *http://wma.comb.es*) and IQUA (The Internet Quality Agency: http://www.iqua.net) are trustmark agencies which certify the authoritativeness and reliability of medical Web sites. Finally, Segala (*http://www.segalamtest.com*) focuses on Web accessibility, in particular concerning mobile devices.

describing resource content, may allow content providers to generate and modify labels concerning their resources, whereas other labeling services (such as WMA), which rate the authoritativeness/reliability of the information and/or service provided by Web resources, may require a centralized control on labels, which must be generated and stored by the labeling service itself. Depending on the scenario decided by labeling services, different trust enforcement mechanisms can be adopted. More precisely, QUATRO provides three different controls which can be used to validate labels. The first is based on the verification of labels' integrity: a label is considered valid if it is the same that has been generated by the labeling authority. The second is based on the expiring time specified for the label: in this case, a label is considered valid until a given date. Finally, the third control is based on the use of content analyzers, specific for each labeling vocabulary, which are in charge of verifying whether the description provided in the label actually corresponds to the current characteristics of the labeled resource; in case of negative response, the label is considered invalid. Such controls can be used separately or in combination, depending on the requirements of the specific labeling vocabulary. For instance, ICRA may adopt only content control (since it allows labels to be modified by content providers), whereas WMA may use both integrity and expiring time controls.

In the QUATRO architecture (depicted in Figure 6), the three main components (QUAPRO, LADI, and ViQ) perform different tasks. QUAPRO is in charge of retrieving the labels associated with the resources requested by users and of evaluating their trustworthiness. QUAPRO relies on labeling services' databases for verifying the integrity of labels, and on content analyzers for performing content control. Such information is then returned to LADI and ViQ, which notify users of the presence/absence of labels, and whether they are trustworthy or not. More precisely, LADI provides an API to Google and Yahoo!, where each search result is associated with an icon in case a content label is available for the corresponding resource. By selecting the icon, the information concerning the label is displayed. By contrast, ViQ provides an interface, integrated in the Web browser, between end users and the labels possibly associated with the displayed resource.

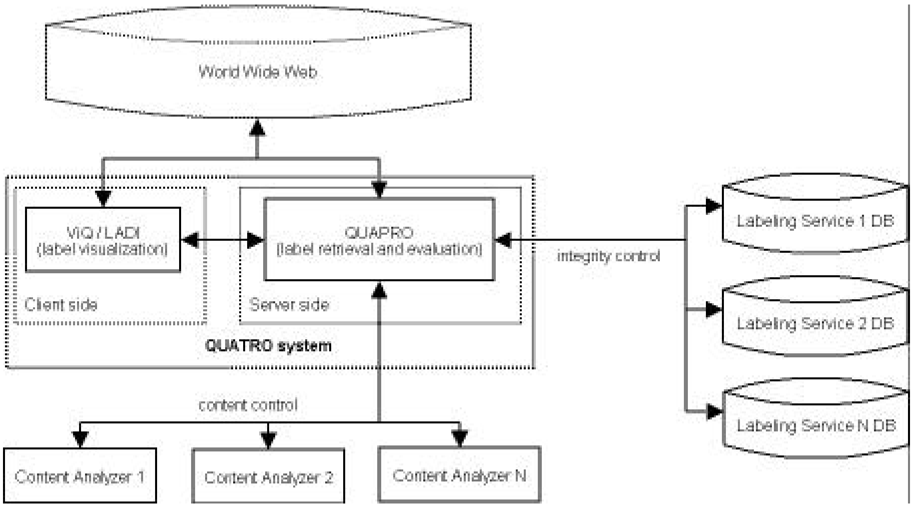


Figure 6: ***The QUATRO architecture***

The QUATRO system is now under development by the project consortium. More precisely, NCSR "Demokritos" (*http://iit.demokritos.gr*) and CoolWave (*http://www.coolwave.co.uk*) are responsible of QUAPRO and LADI, respectively, whereas we are in charge of building ViQ. Note that, at least in its former version, the QUATRO system does not evaluate labels with respect to users' preferences. The reason is that, differently from EUFORBIA, the QUATRO approach aims mainly at informing users about the characteristics of the requested resources, not at blocking inappropriate content. Thus, it is up to the end user to decide whether a resource satisfy his/her requirements, after being notified of the description provided by the possibly associated labels.

Nonetheless, we plan to investigate how such evaluation can be performed automatically, based on the profile and/or preferences of users, according to the MFM filtering approach. These features can be enforced in ViQ, since both QUAPRO and LADI are designed in order to evaluate and notify the presence/absence of labels independently from which user submits an access request. Currently, ViQ allows end users to specify preferences concerning the labeling vocabularies to be taken into account; thus, end users may decide that they are not interested in labels of a given vocabulary, which, as a consequence, will not be retrieved and evaluated by the system. Our purpose is to extend ViQ in order to support the possibility of specifying user profiles and policies, and to enforce filtering mechanism to be used for deciding whether a resource is appropriate or not depending on the specified policies.

4. Conclusions

In this paper we illustrated MFM, a general model for Web content filtering, and the two prototypes implementing it, developed in the framework of the EU projects EUFORBIA and QUATRO, and designed for two different application domains: users' protection and Web quality assurance. Besides modeling the characteristics of the filtering domain, MFM improves the approaches currently available by supporting policies taking into account both users' and resources' characteristics, described by using multiple metadata vocabularies. As a result, MFM, on one hand, allows the enforcement of interoperability among filtering systems based on different approaches, and, on the other hand, it can be easily tailored to users' requirements and preferences. Starting from these results, we now plan to investigate how MFM can be extended in order to integrate it in the architecture of Web services proposed by the W3C [13], where filtering, by granting the possibility to be aware of the in/appropriateness of Web resources, could play a relevant role in establishing trust between users and service providers.

References

1. P. Archer, N. Shimuzu, K Ahmed, D. Brickley, D. Appelquist, and K. Chandrinos. RDF content labels: Schema description. QUATRO Technical Specification, July 2005. Available at *http://www.w3.org/2004/12/q/doc/content-labels-schema.htm*.
2. E. Bertino, E. Ferrari, and A. Perego. MaX: An access control system for digital libraries and the Web. In *Proc. of the 26th International Computer Software and Applications Conference* (COMPSAC 2002), pages 945–950. IEEE CS Press, 2002.
3. E. Bertino, E. Ferrari, and A. Perego. Content-based filtering of Web documents: The MaX system and the EUFORBIA project. *International Journal of Information Security*, 2(1):45–58, Nov. 2003.
4. E. Bertino, E. Ferrari, and A. Perego. Web content filtering. In E. Ferrari and B. Thuraisingham, editors, *Web and Information Security*, chapter 6, pages 112–132. IDEA Group Publishing, Hershey, PA, 2006.
5. E. Bertino, E. Ferrari, A. Perego, and G. P. Zarri. A multi-strategy approach to rating and filtering online resources. In *Proc. of the 16th International Workshop on Database and Expert Systems Applications*, pages 519–523. IEEE CS Press, 2005.
6. P. Resnick and J. Miller. PICS: Internet access controls without censorship. *Communications of the ACM*, 39(10):87–93, Oct. 1996.
7. K. Stamatakis, V. Karkaletsis, A. Perego, D. Rose, and P. Archer. Specifications of the QUATRO software tools. QUATRO Technical Specification, Nov. 2005.
8. World Wide Web Consortium. PICSRules 1.1. W3C Recommendation, Dec. 1997. Available at *http://www.w3.org/TR/REC-PICSRules*.
9. World Wide Web Consortium. PICS rating vocabularies in XML/RDF. W3C Note, Mar. 2000. Available at *http://www.w3.org/TR/rdf-pics*.
10. World Wide Web Consortium. A P3P Preference Exchange Language 1.0 (APPEL1.0). W3C Working Draft, Apr. 2002. Available at *http://www.w3.org/TR/P3P-preferences*.
11. World Wide Web Consortium. The Platform for Privacy Preferences 1.0 (P3P1.0) specification. W3C Recommendation, Apr. 2002. Available at *http://www.w3.org/TR/P3P*.
12. World Wide Web Consortium. Web Content Accessibility guidelines 2.0. W3C Working Draft, June 2005. Available at *http://www.w3.org/TR/WCAG20*.
13. World Wide Web Consortium. Web Services Architecture. W3C Working Group Note, Feb. 2004. Available at *http://www.w3.org/TR/2004/NOTE-ws-arch-20040211*.
14. G. P. Zarri. NKRL, a knowledge representation tool for encoding the 'meaning' of complex narrative texts. *Natural Language Engineering — Special Issue on Knowledge Representation for Natural Language Processing in Implemented Systems*, 3:231–253, 1997.

Discovering the Wealth of Public Knowledge: An Approach to Early Threat Detection

Protima Banerjee, Xiaohua Hu, Illhoi Yoo
College of Information Science, Drexel University, Philadelphia, PA 19104

Reference to this article should be made as follows:
Banerjee P., Hu X., Yoo I., (2005), Discovering the Wealth of Public Knowledge:
An Approach to Early Threat Detection, AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), 2005, pp: 77-85

Abstract

Don Swanson first coined the concept of Undiscovered Public Knowledge almost twenty years ago – that is, the idea that existing knowledge bases can be mined to generate new and novel associations between bodies of information that are similar but distinct. Since then, he and his colleagues have applied his ABC model to medical literature, successfully postulating that fish oils could be used to offset the effects of Raynaud's disease – a hypothesis later verified by experimental research. Later researchers have expanded on Swanson's work, attempting to fully automate his process and, in particular, incorporate semantic information to streamline hypothesis generation.

Undiscovered Public Knowledge has special significance to Homeland Security since so much information relating to terrorist threats is already a part of the public domain. Using Swanson's model and NLP techniques, the potential exists to generate a culled list of "intelligent" hypotheses of potential threat situations which could then be evaluated by an analyst. As time continues to be of the essence in threat detection, any reduction in human effort that might result from this process could prove to be a significant advantage in the war on terror.

This paper will describe the potential adaptability of the Biomedical Semantic-based Knowledge Discovery System (Bio-SbKDS) developed at Drexel University to mine news items in conjunction with GIS information (for proximity searches) and other publicly available knowledge bases for threat hypotheses.

1. Introduction

Don Swanson first coined the concept of Undiscovered Public Knowledge almost twenty years ago – that is, the idea that existing knowledge bases can be mined to generate new and novel associations between bodies of information that are similar but distinct. Swanson postulated that the hidden within the body of knowledge which is readily available are nuggets of information which can be extrapolated and/or inferred – in other words, that we know more than we think we know.

One can immediately see both the value and the relevance of this Undiscovered Public Knowledge (UPK) to the Homeland Security domain. Information pertinent to Homeland Security encompasses a wide range of topics and comes from a wide variety of sources, many of which are in the public domain. News reports, public FBI watch lists, the Office of Foreign Assets Control (OFAC) list of blocked individuals, GIS information, and even weather reports might be seen as relevant to threat detection. In addition, many other pieces of less structured information which are available in the public domain such as emergency response procedures, policy documents and government organization charts might also be valuable.

The construction of a knowledge base that comprises these elements (in addition to proprietary and potentially classified data) is a significant portion of the job of an intelligence analyst. But the manual process of sifting through the mountains of data to find non-obvious connections that might result in a potential threat situation is laborious and time consuming. Any steps that might be taken to automate this process could potentially have huge benefits to both the Homeland Security and intelligence community.

Swanson's first papers dealing with UPK [Swanson, 1986a], [Swanson, 1986b], [Swanson, 1987] proposed the ABC model that is described in greater detail in a subsequent section of this text. This model sets up a systematic way in which new hypotheses can be formed from a knowledge base. Furthermore Swanson's method sets up procedures for culling the list of hypotheses produced so that those that are trivial or already known can be discarded.

The primary drawback to Swanson's early work, however, was the large amount of manual intervention required to make the ABC model successful. His subsequent work on the Arrowsmith project [Swanson & Smallheiser, 1999] attempts to automate the process, but there are still many manual steps required and the entire system is dependent on the availability of domain expertise. Several other approaches to automating the ABC model [Srinivasan, 2004] have been developed and have successfully replicated Swanson's pivotal Raynaud's disease/fish-oil, migraine/magnesium discovery. Recent work at Drexel University [Hu & Yoo, 2005] on the prototype *Biomedical Semantic-based Knowledge Discovery System* (Bio-SbKDS)has incorporated semantic information into the process to reduce the dependency on domain experts.

It is the goal of the paper to propose a plan to adapt the Bio-SbKDS system to the Homeland Security domain, using a prototype ontology of threat detection. The subsequent sections of this paper will discuss Swanson's ABC model in greater detail, present the architecture of the Bio-SbKDS system, and outline a roadmap for its application to threat detection. This paper should be seen as the first in a series which will, it is hoped, eventually demonstrate hypothesis generation capability for terrorist threat detection and/or prevention.

2. Swanson's ABC Model

Swanson's ABC model can best be described as the process to induce "*A* implies *C*", which is derived from "*A* implies *B*" and "*B* implies *C*"; the derived knowledge or relationship "*A* implies *C*" is not conclusive but hypothetical. The *B* concepts are the bridge between *C* and *A* concepts. The following steps summarize the procedure [Swanson & Smalheiser, 1997] .

1. Specify the user's goal (called starting concept *C* such as a disease, or symptom etc)
2. Search the relevant documents *LC* from the biomedical literature for *C*;
3. Generate a set of selected words ("*B*" list) from *LC* using predefined "stop-list" filter; "*B*" concepts are chosen from only the titles of the documents.
4. Search literatures to get those documents *AL* related to "*B*" concepts.
5. Generate a set of words ("*A*" candidates) from the *AL*, "*A*" concepts are also from only the title of the documents.
6. Check whether *A* and *C* cocited together in the literature, if not, Keep *A*
7. Rank "*A*" terms based on how many linkages are made with *B* terms.

One of the drawbacks of Swanson's method is the large amount of manual intervention required. Though Swanson's later work on the Arrowsmith system automates some of the steps in the above process [Swanson & Smalheiser, 1999], much manual intervention is still required in the procedure such as the choice of proper lists of stop words, to filter through the large number of connections to identify the real novel connections/hypotheses.

2.1 The Application of the ABC Model to Threat Detection

Swanson originally applied his ABC model to the biomedical domain and defined a "logic of suggestibility" [Swanson, 1988], which in the most general sense indicates a plausible conjecture between two pieces of disparate information. This mechanism is then used to construct a set of relationships between a starting concept, intermediary concepts and terminating concepts.

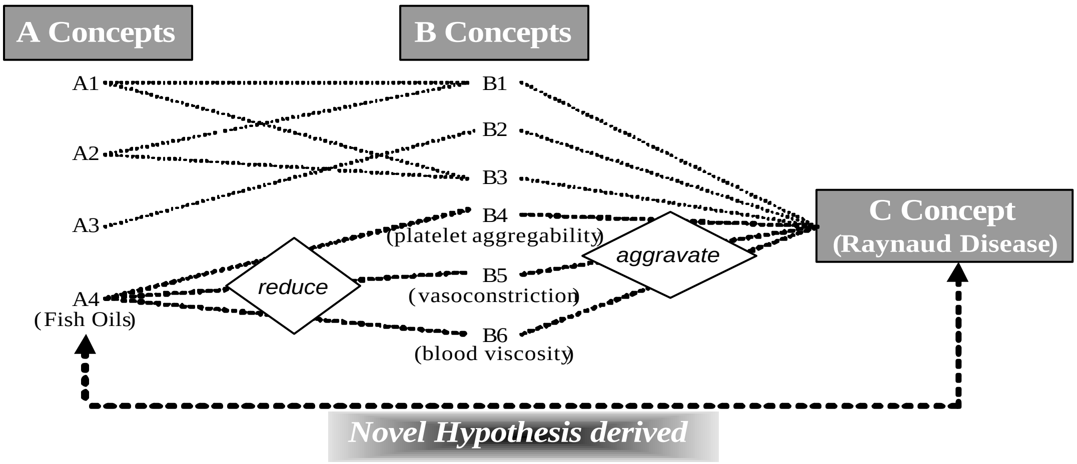


Figure 1: The connection of fish oils to Raynaud's disease

Swanson's discovery of the relationship between fish oils and Raynaud's disease can be used as an example to illustrate the effectiveness of the ABC model (Figure 1). In that case, Swanson started with a target concept, Raynaud's disease, about which he was looking to develop hypotheses.In one set of medical literatures, Swanson was able to find a strong relationship between Raynaud's disease and blood viscosity, showing that increased blood viscosity leads to an aggravation of the symptoms of Raynaud's disease. In an independent set of medical literatures, Swanson was able to locate connections between blood viscosity and fish oils – that is, that the consumption of fish oils has been found to reduce blood viscosity. Based on this information, Swanson plausibly came to the conclusion that the consumption of fish oils is likely to alleviate the effects of Raynaud's disease. Or, that:

- A (The consumption of fish oils) implies reduction in B (blood viscosity)
- B (High blood viscosity) implies aggravation of C (Raynaud's disease)

Indicating that:

- A (The consumption of fish oils) might affect C (Raynaud's disease)

Swanson was able to strengthen this hypothesis by finding two other intermediary B concepts (vasoconstriction and platelet aggregabilty) that tied fish oils and Raynaud's disease to one another. His final hypothesis, postulating a connection between the consumption of fish oils and the effects of Raynaud's syndrome was eventually validated by experimental research. (As a side note, Swanson himself suffered from Raynaud's disease which most likely spurred his initial interest in this particular problem.)

In order for Swanson's ABC model to be applied effectively for threat detection, the C concept that the search centers around would, by its very nature, need to take the form of one or more defended assets. The mining process could then search for appropriate B concepts that indicate that could link a threat (A concept) to those defended assets. One example of the ABC model applied to threat detection might be the following:

- A (Low flying aircraft) can ignite B (Nuclear waste)
- B (Nuclear waste) is dumped in the vicinity of C (Nuclear power plants)

Indicating that:

- A (Low flying aircraft) can threaten C (Nuclear power plants)

It should be noted, however, that an important part of the application of plausible inference to threat detection includes proximity. That is, geospatial information indicating the relative positions of defended assets and threats play an important role in determining whether or not a threat condition actually exists.

3. Biomedical Semantic-based Knowledge Discovery System (Bio-SbKDS)

As might be surmised from the preceding sections, Swanson's approach and model are highly dependent on domain expertise. Without the appropriate domain knowledge used in the determination of the list of stop words and to filter the linking B concepts, the hypotheses produced by the ABC model would be much more numerous and require a significant level of effort from the human in the loop to sift through. Incorporating semantic knowledge into this framework serves to alleviate these burdens from the human user of the system, as well as providing a systematic way of ensuring that domain knowledge is consistently applied to all situations – reducing the likelihood, as it were, of operator error.

3.1 Background and Roots

The Biomedical Semantic-based Knowledge Discovery System (Bio-SbKDS) prototype developed at Drexel University [Hu & Yoo, 2005] is rooted in several algorithms which have been developed to overcome the limitations of Swanson's approach. Based on approaches which emphasize the use of association rules to find connections among medical concepts [Joshi, *et al*, 2004] [Srinivasan, 2004] [Pratt and Yetisgen-Yildiz, 2003] [Hristovski, *et al*, 2001] [Hristovski, *et al*, 2003], the system adapts existing text mining methods to the problem of undiscovered public knowledge. The foundation of the system is the creation of a large association rule base upon which the algorithm for discovering new relations between concepts can operate. However, Bio-SbKDS is unique in its incorporation of semantic knowledge into the data mining process, which gives it a significant advantage over the earlier approaches mentioned above; those approaches used only the importance measure for words or terms to select high informative words or terms without considering the semantic relations among concepts. However, the tackling the association problem by not only the importance measure but also the semantic information among the concepts provides a significant advantage. [Hu & Yoo, 2005] focus on developing fully automated approaches to the UPK problem based on the semantic knowledge about the medical concepts and their relationships. They use semantic information to prune irrelevant medical concepts and bogus or non-interesting relationships among the medical concepts. This approach replaces manual ad-hoc pruning by using existing biomedical ontologies, and the use of an intermediate set of automated identified semantic types helps to manage the sizable branching factor.

In order to create an automated approach to identify those interesting and meaningful terms for the *B* and *A* concepts in the ABC model, [Hu & Yoo, 2005] propose reliance on semantic types (e.g., medical condition or disease and a potential treatment) that were plausible for terms that could be correlated. At that point, the system can filter out any concepts that do not match the semantic-type criteria which defines the search criteria. By way of contrast, Swanson addressed this problem by manually creating a query-customized list of stop words to filter out the uninteresting concepts, but such a level of word-based customization could be difficult to scale to new kinds of query concepts and connections. A further advantage of the Bio-SbKDS method over other approaches is the minimum human intervention that is required.

3.2 System Architecture & Algorithm

The Bio-SbKDS algorithm takes full advantage of the semantic knowledge in the Unified Medical Language System (UMLS), which is a mechanism for integrating all major biomedical vocabularies started by National Library of Medicine as a long-term R&D project in 1986. In addition, the system makes use of Medical Subject Headings (MeSH), also published by the National Library of Medicine, which consist of a controlled vocabulary and the thesaurus for cataloging medical documents. UMLS and MeSH are key components in selecting the appropriate semantic types for *B* and *A* concepts through mutual qualifications and to identify relevant *B* and *A* concepts. The advantage of the algorithm is that, using only initial relations (possible relationships between *C* concept and *A* concepts), all the semantic types for both *B* concepts and *A* concepts are automatically derived using a biomedical ontology (UMLS).

The following is a summary of the algorithm used by the system:

INPUT: A starting concept *C* plus a date range, the initial semantic relations *(ISR)* between the starting concept and the to-be–discovered target concept, the role of keyword for possible relations (subject or object)

OUTPUT: Target Concept List (*A* concepts)

1. Find the semantic types of the starting concept C from the ontology
2. Find all the possible semantic types of the to-be-discovered concepts B related to C. These will be the list of B concept candidate concept semantic types.
3. Extract all semantic types related to ISR, which are the candidate semantic types for the to-be-discovered target concepts A.
4. Extend the candidate semantic types obtained in (3) by following through the IS-A relations in the ontology. The result of this step should be an extended list of semantic types which are then passed on to the next step in the algorithm.
5. Check if there are relations between the B concept semantic type candidates and the extended list of A concept semantic types produced in (5), and also if the two semantic type sets pass the relation filter. If not, such semantic types are dropped from their semantic type list. Remove the irrelevant semantic types.
6. Search the biomedical literature to get all the documents related to concept C. Then, extract all the keywords (MeSH terms in the biomedical context) from these documents. These keywords now become the candidate list of B concepts.
7. Apply B concept category restriction to the list produced in (6); selecting only those terms belong to at least one semantic type from the list produced in (5). In addition, Bi-Decision Maker further qualifies the B concepts.
8. Search all the top ranked B terms to get all the documents that contain A concepts. Then, extract keywords (MeSH terms) to get an initial list of A concept candidates..
9. Apply A concept category restriction to the list in (8). In addition, Bi-Decision Maker further qualifies the list of A concept candidates.
10. From the list of A concept candidates that are not co-occurred with C concept in the existing document base (Medline for medical documents), the top ranked A concepts are selected.The steps of

the algorithm are numbered in the data flow diagram depicted in Figure 2 below.

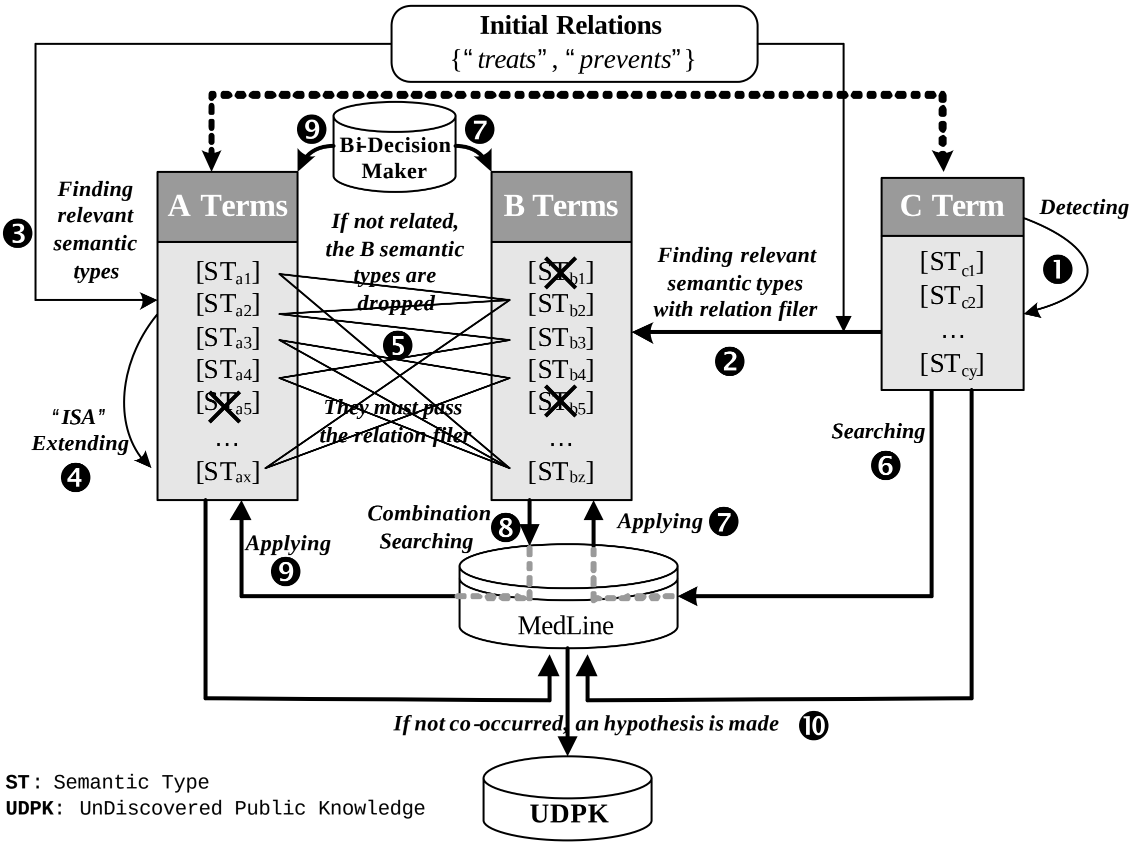


Figure 2: Data Flow of the Bio-SbKDS system

4. Application of the Bio-SBKDS Algorithm to Threat Detection

4.1 Threat Detection Ontology

The most obvious alteration to the Bio-SbKDS prototype in order for it to be relevant to a threat detection scenario is the incorporation of a threat detection ontology. However, unlike the biomedical domain, a validated ontology for threats scenarios is not currently available in the public domain. While several private groups such as Semagix and MITRE advertise their proprietary homeland security ontologies, no government or military group of subject matter experts has published and/or validated such a knowledge base for the public domain. Though the development of a fully developed threat ontology is beyond the scope of this project, for proof of concept purposes, we propose the creation of a limited ontology for our effort. Top level concepts of this scaled-down ontology are shown in Figure 3 below.

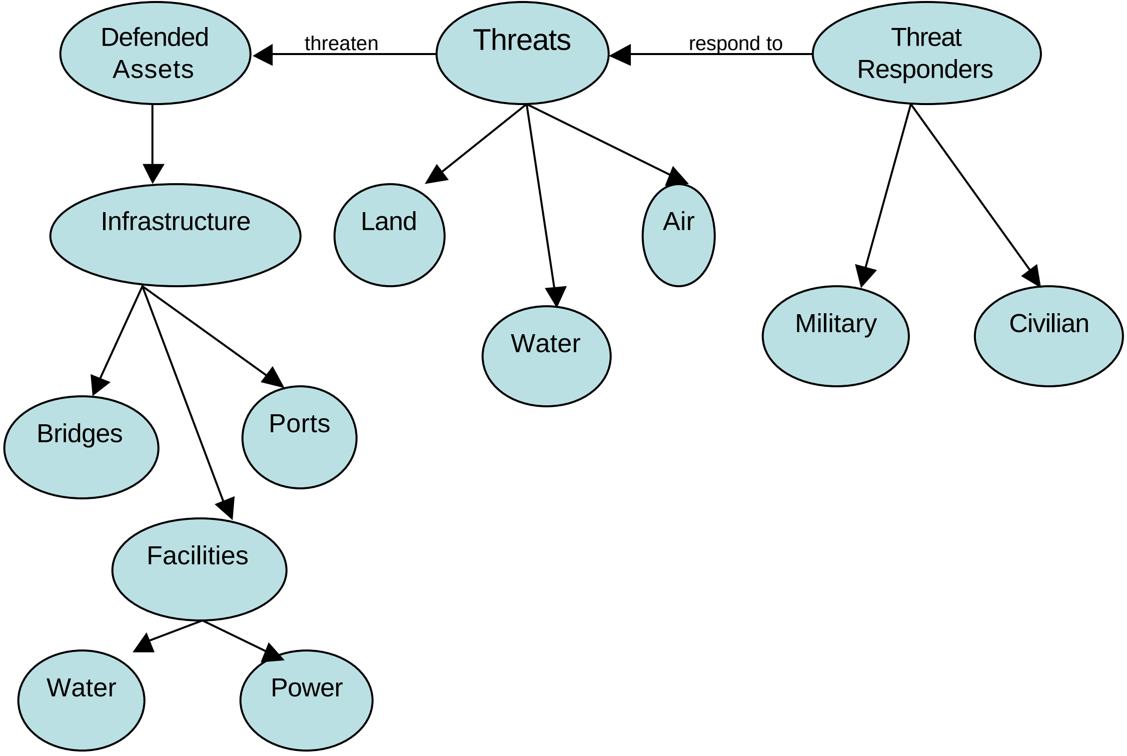


Figure 3: Top level concepts for a sample threat ontology

In the future, it is hoped that the Department of Homeland Defense or similar government agency will release a validated ontology.

4.2 Geospatial Information

As mentioned in a previous section, it is critical to the threat detection function to include geospatial information into the threat detection process. Proximity relationships between entities are a major factor in determining whether or not a threat condition exists. Furthermore, the ability to use a GIS (Geographic Information System) with semantic knowledge would provide a powerful and novel capability. There is currently no placeholder for GIS information within the structure of Bio-SbKDS, and the architecture of the system would have to be modified to take this information into account.

4.3 Natural Language Processing and Concept/Entity Extraction

Medical information systems, such as Medline, use a consolidated list of terms to identify keywords or concepts within a document. [Swanson, 1988] discusses at length how these keywords are often subjective, rather than objective, and that no information is presented within the keyword list about the relationships among keywords. For example, a medical document may be about migraine headaches and magnesium, and contain a causality relationship between the two terms, an inverse causality relationship, or show that no relationship exists at all. Thus, Swanson showed that it was difficult, based on keywords or title alone, to determine the concepts and entities contained in a medical document.

For texts associated with Homeland Security, which may be news releases, structured data such as OFAC and MISLE, weather reports, web services and standard policies documents, which do not have any such meta-information associated with them, this problem is magnified ten-fold. Entity and information extraction algorithms will first have to be run on the various document sets to be mined for UPK before any list of concepts can be extracted from them.

4.4 Data Sources

Finally, the success of the system for threat detection is influenced most heavily by the choice of data sources. However, data sources should be chosen based on knowledge of the defended asset C, about which hypotheses should be generated. For the purposes of this proof of concept, it is our aim to search for threats to the Port of Philadelphia and to limit the data sources to news archives in the local region that deal with the port. As well, we will consider web sources such as the website of the Port of Philadelphia itself and the Coast Guard and Homeland Defense websites.

In addition, there are several structured data sources, in the form of public web services that should be incorporated into the mining process. These include watch lists published by the FBI and OFAC (the Office of Foreign Assets Control), as well as the public MISLE (Marine Information for Safety and Law Enforcement) data source maintained by the Coast Guard. The process by which this structured information can be incorporated into either the mining or the decision making process will need to be investigated, and perhaps become a modification to the algorithm.

5. Conclusion and Next Steps

The intent of this paper was to provide high level guidelines and outline a plan for the creation of a proof of concept that would enable mining Undiscovered Public Knowledge for the purposes of threat detection. Using the framework of the already developed Bio-SbKDS prototype system, we can take advantage of semantic data mining methods that are already being successfully used for a similar purpose in the biomedical domain. Much work, however, still needs to be accomplished and this is intended to be the first in a series of papers that detail the full UPK mining process for homeland security applications.

The next steps in this work are, in order:

- Investigating text processing and entity extraction methods by which news (or other) documents can be classified in terms of their component concepts.
- Finding and integrating an appropriate Geographic Information System into the decision making process of the mining algorithm
- Investigating the inclusion of structured data sources into the mining algorithm
- The creation of a sample ontology (using semi-automated methods) for threat detection for the Port of Philadelphia

It is our expectation that over the course of the next few months one or more papers detailing the progress of these efforts will be written and the full scope of the work concluded by year end.

Acknowledgement: This research work is supported in part from the NSF Career grant (NSF IIS 0448023). NSF CCF 0514679 and the PA Dept of Health Tobacco Settlement Formula Grant (#240205, 240196,).

6. Reference

[1] Hristovski D, Peterlin B, Mitchell JA, Humphrey SM., Improving literature based discovery support by genetic knowledge integration, Stud. Health Technol. Inform. 2003, Vol. 95, pp. 68-73.

[2] Hristovski D, Stare J, Peterlin B, Dzeroski S. Supporting discovery in medicine by association rule mining in Medline and UMLS. Medinfo. 2001, 10(Pt 2), 1344-8.

[3] Hu X, *Mining Novel Connections from Large Online Digital Library Using Biomedical Ontologie*s, Library Management Journal, special issue in Libraries in the Knowledge Era: Exploiting the knowledge wealth for Semantic Web Technology, Vol. 26, No 4/5, 2005, pp261-270

[4] R. Joshi, X.L. Li, S. Ramachandaran, T.Y. Leong, Automatic Model Structuring from Text using BioMedical Ontology, In American Association for Artificial Intelligence (AAAI) Workshop, pp. 74-79, San Jose, California, July 2004

[5] Lindsay, R.K, & Gordon, M.D. (1999). Literature-based discovery by lexical statistics. Journal of the American Society for Information Science, 50(7), 574-587.

[6] National Library of Medicine (NLM), 2004AC UMLS Documentation, http://www.nlm.nih.gov/research/umls/documentation.html, 2004.

[7] Padmini Srinivasan, Text mining: Generating hypotheses from MEDLINE, Journal of the American Society for Information Science, 2004, Vol. 55, No. 4, pp. 396-413

[8] Pratt, Wanda and Yetisgen-Yildiz, Meliha, LitLinker: capturing connections across the biomedical literature, K-CAP'03, pp. 105-112, Sanibel Island, FL, Oct. 23-25, 2003

[9] Swanson, DR., 1986a, Undiscovered public knowledge. Libr. Q. 56(2), pp. 103-118.

[10] Swanson, DR., 1986b, Fish-oil, Raynaud's Syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine 30(1), 7-18.

[11] Swanson, DR., 1987, Two medical literatures that are logically but not bibliographically connected. JASIS, Vol. 38, No. 4, pp. 228-233.

[12] Swanson, DR., 1988, Migraine and magnesium: eleven neglected connections. Perspectives in Biology and Medicine, 31(4), 526-557.

[13] Swanson, DR. & Smalheiser, NR., 1997, An interactive system for finding complementary literatures: A stimulus to scientific discovery. Artificial Intelligence 91(2), 183-203.

[14] Swanson, DR. & Smalheiser, NR., 1999, Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. Library Trends 48(1), 48-59

SemNetMan

Project announcement

Semantic Based Network Management: SemNetMan

The project **SemNetMan** (semantic based network/cluster management) combines two methods relevant for the knowledge society: the social network analyses (SNA) and techniques of the semantic web.

The resulting method should enable efficient network management within the scope of project environments and related requirements.

The design of the methodology is embedded in three use cases: the network of the Semantic Web School, the Platform Wissensmanagement and the Austrian Biogasnetzwerk.

The project is realized by seven Austrian organisations from research to industry containing specialists from network management, semantic technologies and social network analyses.

SemNetMan is supported by the protecNET*plus* programm of FFG.

Core partners: **NIWA-WEB** (www.niwa.at), **punkt.netServices**(www.punkt.at), **M2N consulting and development** (www.m2n.at), **DERI Innsbruck** (www.deri.at), **Semantic Web School** (www.semantic-web.at)

Project description provided by **Anna V. Zhdanova** (anna.zhdanova@deri.org)

Semantic WS-Agreement Partner Selection (SWAPS)

Nicole Oldham, Kunal Verma and Amit Sheth
LSDIS Lab, Computer Science Department, University of Georgia, USA
http://lsdis.cs.uga.edu/projects/meteor-s/

In a service oriented environment it is advantageous for service consumers and providers to obtain guarantees, usually pertaining to quality of service (QoS) aspects, regarding the services that they both require and offer. WSDL does not provide a means to express these guarantees; therefore such standards as WS-Policy[15] and Web Service Level Agreements (WSLA[16]) exist to allow for the expression of additional nonfunctional attributes. However, these standards are not expressive enough to represent the truly complex nature of the relationship between a service consumer and provider. The WS-Agreement specification[17] defines a language and protocol for capturing this intricate relationship with agreements between two parties. An agreement specifies one or more service level objectives (SLO) which state the requirements and capabilities of each party on the availability of resources and service qualities. WS-Agreement is more expressive than the previous policy standards because in addition to SLOs, an agreement contains scopes for which the guarantee holds, conditions which must exist in order for the guarantee on the SLO to be valid, and business values, such as penalties and rewards, which incur if the SLO is not satisfied. In addition, these agreements often contain alternative sets of guarantees and are symmetric such that each provider does not only state guarantees regarding capabilities but likely also has requirements.

As each consumer seeking a suitable provider has many options to choose from, the manual partnering of these parties is time consuming, tedious and error prone. With the increasing acceptance and popularity of WS-Agreement coupled with the ever present need to protect the quality of service with guarantees, the development of an approach to facilitate the automatic matching of these agreements is imperative. To date, there has been very little work done to advance the tools available for WS-Agreements. Existing tools are solely for the creation and monitoring of agreements. The SWAPS project, developed as part of the METEOR-S project on Semantic Web Services and Processes at the LSDIS Lab at the University of Georgia, presents the framework and implementation of an innovative tool for the dynamic partnering of WS-Agreements. The approach utilizes Semantic Web technologies in combination with ARL rules to achieve rich and accurate matches. A key feature is the novel and flexible approach for achieving user personalized matches.

We have accomplished the following as the main contributions of our work:

- Adding to the small body of work for WS-Agreement by defining a framework and implementation for the matching of agreements which eliminates tedious and error prone manual matching.
- Developing a strategy for reasoning over the complex and multifaceted WS-Agreements.
- Presenting a flexible approach for specifying and reasoning over user defined preferences which allows the matching to be modified and customized without changing code or possessing programming knowledge.

The framework and implementation of SWAPS are given in detail at http://lsdis.cs.uga.edu/projects/meteor-s/SWAPS/

[15] http://www-128.ibm.com/developerworks/library/specification/ws-polfram/
[16] http://www.research.ibm.com/wsla/
[17] https://forge.gridforum.org/projects/graap-wg/document/WS-AgreementSpecificationDraft.doc/en/10

BRAHMS – a high-performance RDF/S storage

Maciej Janik, LSDIS Lab, Computer Science Department, University of Georgia

BRAHMS is designed and implemented as a high-performance main-memory RDF/S storage, capable of storing, accessing and querying large ontologies. To maximize performance, all data from ontology model are stored in main memory. Disk is used only to store a snapshot of previously parsed RDF/S data. This allows faster startup of the system, as parsing of even large RDF/S file is done only once.

BRAHMS was created as a framework for testing semantic association discovery algorithms for SemDis[1] project in LSDIS lab at University of Georgia. Discovery of semantic associations in Semantic Web ontologies is an important task in various analytical activities. Several query languages and storage systems have been designed and implemented for storage and retrieval of information in RDF ontologies. However, they are not always adequate for semantic association discovery of longer length in large ontologies.

Algorithms for semantic association discovery that were tested, may access entity neighborhood and iterate over it millions of times during execution. They also need other simple graph access methods. BRAHMS offers a wide range of methods in API to fulfill these needs. For high-performance together with strict control over used memory, it is implemented in C++. This gives full control over destroying previously created objects and iterators to smoothly free allocated resources. For convenience of Java programmers, there are provided bindings to API SemDis[2] in Java.

Design of BRAHMS was driven by the needs of ontology query algorithms. As a result systems focuses on providing high performance methods for accessing ontology data, but it does not allow to modify data in parsed ontology. Each time ontology changes, a new snapshot must be generated from the beginning. This is the design trade-off that was made for offered speed.

Tests performed with in-memory ontology model using Jena, Sesame, and Redland, which are three of the well-known RDF storage systems, proved that for the same semantic association discovery algorithms BRAHMS achieved much better performance. Due to compact representation of ontology in memory, algorithms that used BRAHMS were able to successfully run on larger ontologies, where other systems reached machine memory limits.

For further information about BRAHMS, visit http://lsdis.cs.uga.edu/projects/semdis/brahms

This work supported by the National Science Foundation under Grant No. IIS-0325464 titled "SemDis: Discovering Complex Relationships in Semantic Web".

[1] SemDis project – http://lsdis.cs.uga.edu/projects/semdis
[2] API SemDis – http://lsdis.cs.uga.edu/~aleman/research/apisemdis

REGULAR COLUMNS

In this Issue:

RDF Technologies – Foundations, Applications and Developments

Columnist: Heiner Stuckenschmidt, Vrije Universiteit Amsterdam

There is a wide agreement that the Semantic Web will largely be built on top of RDF. Therefore, a flexible and scalable infrastructure for storing, managing and retrieving RDF-based information is essential. An increasing number of software tools is available supporting the complete life-cycle of RDF models. Editors and converters are available for the generation RDF schema representations from scratch or for extracting such descriptions from database schemas or software design documents. Storage and retrieval systems have been developed that can deal with RDF models containing millions of statements, and provide query engines for a number of RDF query languages. Annotation tools support the user in the task of attaching RDF descriptions to web pages and other information sources either manually or semi-automatically using techniques from natural language processing. Finally, special purpose tools support the maintenance of RDF models in terms of change detection and validation of models. Further, an increasing number of applications that use RDF for representing, integrating and reasoning about information are available. Example for such applications can be found at http://challenge.semanticweb.org. Most of the existing applications of RDF are in the area of information systems. In this area, the benefits of RDF in terms of conceptual representations, interoperability and reasoning support are directly visible.

This column will discuss RDF as a key technology for intelligent information systems on the web. The discussion be centered around there aspects of RDF technology:

Foundations

We will review the principles of RDF and relate them to other well established and emerging technologies such as graph theory, relational databases, topic maps and XML. The discussion will focus on identifying commonalities and differences and point to insights from other areas that can be used to improve RDF technologies.

Applications

We will review existing and potential applications of RDF technologies and discuss the benefits and problems of RDF in areas such as information integration and thesaurus-based information retrieval. Besides surveying existing approaches and their features, we will try to summarize lessons learned and open problems.

Developments

We will discuss recent research questions that have been raised in connection with RDF technologies. Examples are topics like query language standards for RDF, the notion of views or provenance in RDF representations. We will introduce the topic and its relevance, present the current status of the discussion and review existing proposals for a solution.

The column will be targeted at members of the semantic web as well as the information systems community. We aim to provide researchers and practitioners in information systems with a better understanding of the benefits and trade-offs of using RDF for building information systems. Further, we want to point semantic web researchers and practitioners engaged in the development of RDF technologies to the area of information systems as a fruitful application area and provide more insight in the special needs and problems of that area. In summary this column tries to strengthen the link between information systems and RDF technologies by discussing topics that are at the border of the two disciplines.

RDF and Thesauri – A Perfect Match

Heiner Stuckenschmidt, University of Mannheim, Germany
heiner@informatik.uni-mannheim.de

The Semantic Web relies on the availability of terminological knowledge in terms of ontologies and semantic schemas for describing application data. RDF has been designed to support the representation of such terminological information by means of the RDF Schema and OWL vocabulary. A well-known problem with respect to the development of the semantic web is the fact that building these terminological models is very time consuming and guaranteeing the quality of newly built models is a major problem. The best way out of this problem is to re-use existing models of terminological knowledge that have been developed outside the semantic web community and make these models accessible by semantic web systems by encoding them in RDF.

Thesauri are used in the Digital Library community for quite some time to improve information access by providing explicit models of the terminology used in a domain. Thesauri are often developed and maintained over a long time and with serious effort by both commercial and non-commercial organizations. This makes thesauri attractive candidates for being reused as terminological knowledge on the semantic web. In this part of the Column we will discuss the relation between thesauri and RDF as well as existing approaches for making thesauri available in semantic web applications by converting them into RDF.

Thesauri and RDF

Before we can start discussing ways of making thesauri available in RDF format, we first have to clarify the conceptual and terminological differences between thesauri and RDF and point out to commonalities and differences that might hamper a conversion.

Thesaurus Terminology

- **Facets** are broad topic areas that divide the Thesaurus into a number of independent hierarchies. Facet names are not themselves preferred terms (i.e. they cannot be used as index terms). However, they might as well be characterized as top-level preferred terms.
- Each Facet consists of a hierarchy of **preferred terms.** These terms are used as index terms to describe the content of information in a resource. A term can occur in more than one facet.
- Preferred terms are enriched by a set of **synonyms**. Synonyms represent alternative terms that can be used to refer to the corresponding preferred term. Synonyms can be used by a person to index or query information, but will be normalized to the preferred term internally. The distinction between synonyms and preferred terms is not fixed. In an annual update a synonym can be promoted to preferred term status and a preferred term can be demoted to synonym status.
- **Qualifiers** define legal combinations of index terms. Link terms are used as subheadings for another index term. They denote a context or aspect for the main term to which they are linked. Qualifiers do not have a fixed link term role. They can also be indexed as single main index terms.
- **Checktags** describe the intended interpretation of a term. They are used as a control mechanism for the indexer. Check tags, too, are a subclass of the preferred terms. They have intended interpretations in special scope notes (note that most link terms also have special scope notes). Check tags are on a separate list, checked for each article by the indexer. Qua format they do not differ from other single index terms in EMBASE records. They differ only in the semantic background.

RDF Terminology

As the terms in a thesaurus are rather at a meta-level (they encode information *about* a specific piece of information in terms of an index). The resource description format RDF has been proposed as a data model for representing such meta-data using an XML syntax. The basic model underlying RDF is very simple. Every type of information about a resource, which may be a web page or an XML element, is expressed in terms of a triple, e.g.:

(document1 has-index aspirin)

Thereby, the property is a two-placed relation that connects a resource to a certain value of that property. A value can be a simple data type or a resource. Additionally, the value can be replaced by a variable representing a resource that is further described by linking triples making assertions about the properties of the resource that is represented by the variable:

(abstract1 summarizes document1)
(document1 has-index c123)
(c123 synonym-term aspirin)

Further, RDF allows multiple values for single properties. For this purpose, the model contains three built-in data-types called collections, namely unordered lists (bag) ordered lists (seq) and sets of alternatives (alt) providing some kind of an aggregation mechanism

(document1 has-index [aspirin headache])

Another feature of RDF is its reification mechanism that makes it possible to use an RDF-triple as value for the property of a resource. Using the reification mechanism we can make statements about facts. Reification is expressed by nesting triples:

(document1 has-index (aspirin treats headache))

One of the main features of the semantic web is the idea to enrich metadata descriptions with explicit models of their intended meaning. These models range from simple schema definitions to complex ontologies that define concepts and describe them by necessary and sufficient conditions thus enabling intelligent applications to reason about their members and their relation to each other. A thesaurus can be seen as a special type of such a semantic model. In general, languages for describing these models provide a special vocabulary for defining additional information of names. RDF schema for example provides a language for encoding structural background information about the vocabulary used in an RDF model. This structural information provides insights into the relations between the different elements of the model and helps to draw conclusions that could not be found from the plain model.

For this purpose, the languages define a special set of semantic primitives that can be used to clarify the relation between names. A typical example is the introduction of a class hierarchy and the possibility to assign objects to classes. In the following example, the semantic relations are underlined.

(document1 type Article)
(Article subClassOf Document)

Another typical example for semantic relations is the possibility to assign types to relations that are used to describe information. For example, we can claim that the relation 'has-index' connects documents with index terms and nothing else.

(has-index domain Document)
(has-index range Index-term)

The semantic relations in the examples above are taken from RDF schema. Other, more expressive languages such as the web ontology language provide a much richer set of semantic relations that can be used to define background for a domain.

Comparison

There are some differences between the terminology that is usually used in order to refer to elements in a thesaurus and in RDF. In the following we briefly compare these notions.

Terms vs. Classes: The central element used to specify background knowledge is the notion of a 'concept'. In a thesaurus concepts are specified by a set of terms that are normally used to refer to the concept. In RDF schema, concept are directly represented in terms of classes with a unique identifier and possibly a set of characteristic properties. In the following we will talk about concepts and terms.

Hierarchy vs. Taxonomy: In a thesaurus, terms are organized in one or more term hierarchies using the 'Broader Term' and the 'Narrower Term' relation. The broadest term in a thesaurus hierarchy is sometimes explicitly marked. In an RDF schema, classes are also organized in a taxonomy using the subclass relation which is equivalent to the broader term relation. Normally there is no explicit inverse to this relation. In the following we will use the term hierarchical relationship in order to refer to the relations mentioned above.

Equivalence vs. Synonymy In a thesaurus, a single concept can be described by more than one term. Normally one of the defining terms is chosen to be the preferred term and all other terms are said to be synonyms of the preferred term. Thus, synonymy holds between terms that specify the same concept. In semantic models we often find the notion of equivalence. The equivalence relation, however, is defined between concepts and indicates that two concept represent the same set of real-world entities.

Converting Thesauri into RDF

Recently, there has been some interest in guidelines for translating existing thesauri into RDF. The rational for this is two-fold: On the one hand, thesauri are important sources of terminological knowledge. As the task of building high quality ontologies is one of the main bottlenecks of the semantic web, reusing thesauri as they are or as a basis for ontology development is a valuable aid for making the semantic web work. On the other hand, despite the existence of standards for encoding thesauri, most existing thesauri are rather isolated entities and linking them to information sources or other thesauri is a time consuming and tedious endeavour. As we will see in the next section, an RDF encoding of thesauri eases this task significantly and has advantages in terms of development time and flexibility of usage.

A major effort for defining the translation of thesauri into RDF is undertaken in the SWAD-E Activity of the W3C (http://www.w3.org/2001/sw/Europe/). Work Package 8 of this activity is entirely concerned with the relation between thesauri and RDF. Reports available from this project cover an overview of existing work on the topic, an RDF schema and an ontology for thesauri, a demonstration prototype as well as guidelines for migrating thesauri to the Semantic Web using RDF. The corresponding reports are available at http://www.w3.org/2001/sw/Europe/reports/thes/reports.html.

Another approach for converting Thesauri to RDF (and OWL) is proposed by van Assem and others (see http://thesauri.cs.vu.nl/). The authors propose a three step process and state a number of guidelines for ensuring the quality of the translation that we will briefly discuss in the following.

Step 1: Syntactic Conversion

The first step of the proposed conversion method is a syntactic translation from the data format of the thesaurus into an RDF encoding. This translation is syntactic in the sense that the structure and the conceptual model of the thesaurus are preserved. This means that the names of relations and types in the thesaurus format are directly translated into RDF types and relations. The resulting RDF model of the thesaurus is a direct copy of the original encoding where the only difference is the use of RDF syntax for the corresponding modelling elements. If necessary, structures that are only implicitly represented in the thesaurus are made explicit for example by introducing new types for modelling elements that are only distinguished by syntactic variations (e.g. angle brackets) in the original format. The syntactic conversion step makes sure that all the information of the original model is still present in the RDF encoding. This step if necessary, because a conversion that tries to directly map a thesaurus onto RDF schema elements is in danger of leaving out information that cannot directly be encoded using RDF schema.

Step 2: Semantic Conversion

As an essential part of the functionality of semantic web systems is based on the semantics of RDF Schema and OWL modelling elements, a purely syntactic conversion will fail to provide the expected benefits. For this reason, the second step in the proposed conversion method is a semantic conversion. In fact the term conversion is misleading as in this step semantic information is added on top of the result of the syntactic conversion. This semantic enrichment is done using RDF schema and OWL constructs and their semantics. There are two aspects of this semantic enrichment. The first is to make the semantics of the original thesaurus format explicit. A typical example is the transitivity or functionality of certain thesaurus relations that can explicitly be encoded by defining the corresponding properties to be instances of owl:transitiveProperty and owl:functionalProperty respectively. This step is still independent from the envisioned use of the thesaurus in semantic web applications. Therefore the other aspect of the semantic enrichment is to link thesaurus relations to elements of the RDF schema representation, for instance by making the broader term relation a subproperty of the rdfs:subClassOf property. This linking to RDF schema elements allows semantic web applications to use the content of the thesaurus in the same way any other RDF schema is used.

Step 3: Standardization

Depending on the original format of the thesaurus, the RDF translation created so far can look very differently. In order to ease the use of the resulting model, it is useful to also link it to existing standards for encoding thesauri in RDF like the proposal of the SWAD-E project mentioned above. Although there is not yet a commonly agreed standard way of representing thesauri in RDF, being compatible with existing proposals for standards ensures a wider usability of the result model and should therefore be done before publishing the result.

The authors tested their approach on two well known examples of thesauri frequently used in digital libraries and semantic web applications, namely MeSH and Wordnet. The result of the corresponding conversion can also be found on the web page mentioned above.

The DOPE System: An example

The DOPE (Drug Ontology Project for Elsevier) project (http://www.aduna.biz/dope/) sponsored by Elsevier is a good example of the reuse of an existing thesaurus in order to build a semantic web application by translating it to RDF and using existing RDF tools to implement the required functionality. In the following, we will briefly describe the application that has been created on the basis of a converted thesaurus without going into details about the actual translation.

The aim of the DOPE project is to investigate the possibility of providing access to multiple information sources in the area of life science through a single interface. This approach is sketched in figure 1 (the following letters refer to the figure):

A. Elsevier's main life science thesaurus, EMTREE 2003©, has been converted to an RDF-Schema format.
B. Using EMTREE 2003, several large data collections (5 million abstracts from the MEDLINE database, and about 500,000 full text articles from Elsevier's ScienceDirect have been indexed using Collexis Fingerprinting technology. In addition to the fingerprint (a list of weighted keywords assigned to a document) metadata about the document such as the authors and the document location are posted on the Collexis server.
C. The Collexis metadata have been dynamically mapped to an RDF model in two steps: the first transformation creates an RDF model, which is an exact copy of the data structure provided by the fingerprint server. The final model is a conceptual document model used for querying the system.
D. An RDF database, using the SOAP protocol, communicates with both the fingerprint server and the RDF version of EMTREE.
E. A client application UI allows the user to interact with the document sets indexed by the thesaurus keywords, using SeRQL queries sent by HTTP.
F. The system is designed in a way can be extended by adding new data sources, which are mapped to their own RDF data source models and communicate with Sesame.
G. New ontologies or thesauri can be added, which can be converted into RDF-Schema, and which also communicate with the Sesame RDF server.

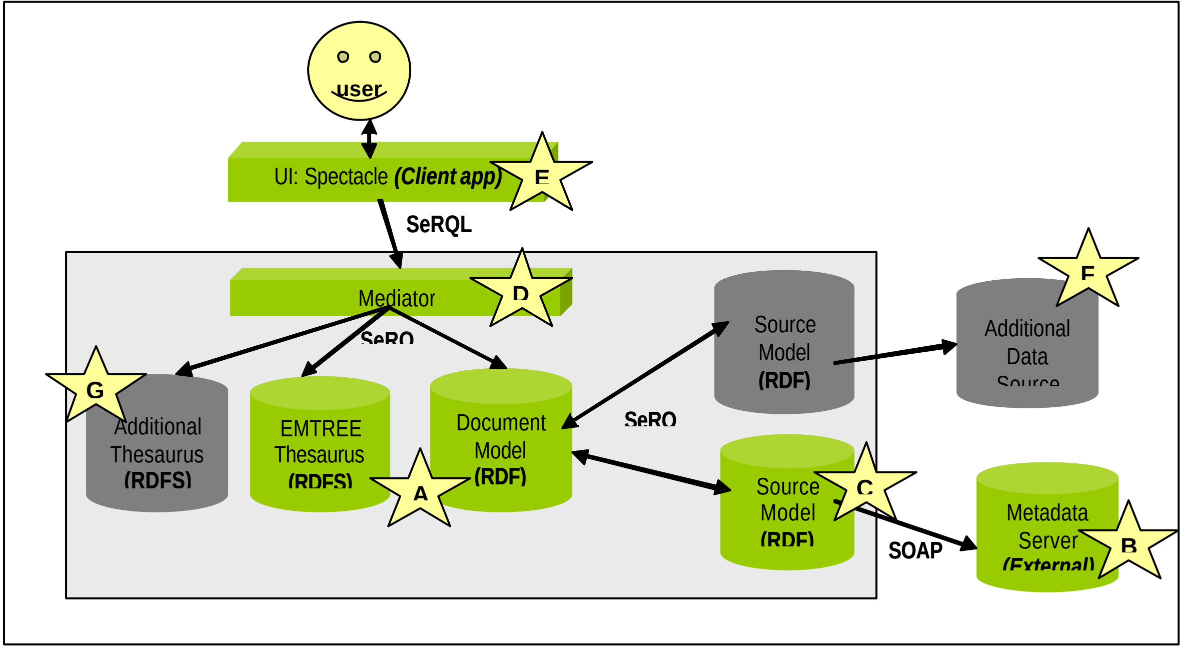
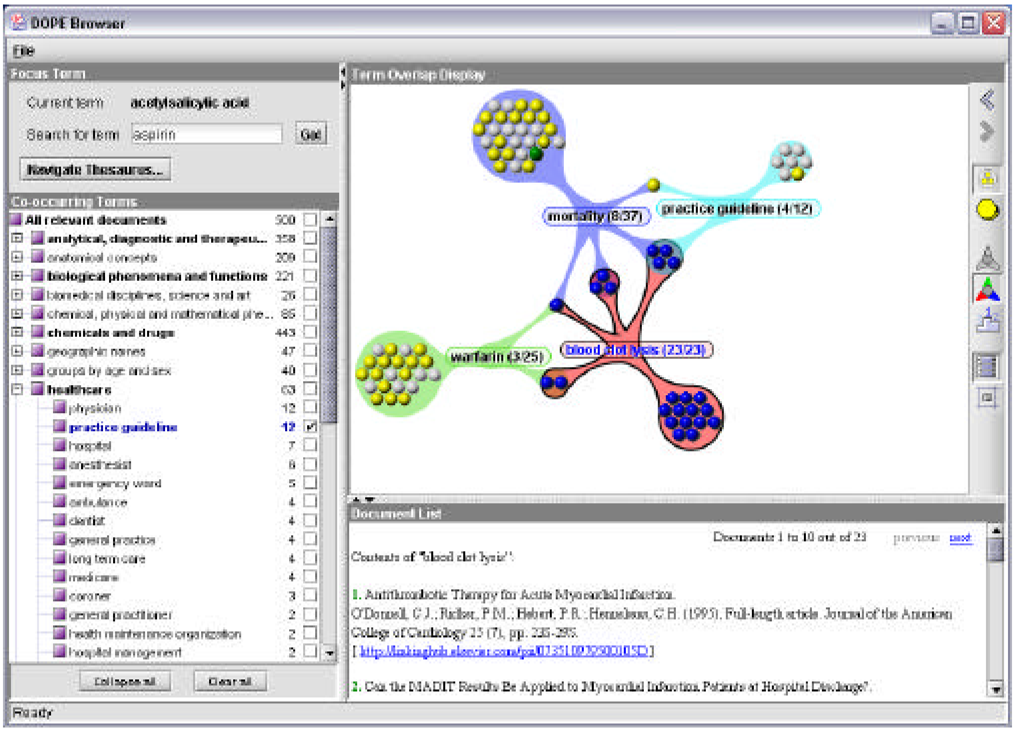


Figure 7: Architecture of the DOPE Prototype

A prototype of a user interface client called the "DOPE Browser" has been designed and created (compare figure 2). It provides querying and navigation of a collection of documents using thesaurus-based techniques, while hiding much of the complexity of the back-end, such as the existence of multiple data sources, any thesaurus or ontology mapping that may take place, etc. In this system, the user sees a single virtual document collection made navigable using a single thesaurus (EMTREE).



Current term acetylsalicylic acid

Search for term aspirin

All relevant documents	300	□
analytical, diagnostic and therapeu...	358	□
anatomical concepts	209	□
biological phenomena and functions	221	□
biomedical disciplines, science and art	26	□
chemical, physical and mathematical phe...	85	□
chemicals and drugs	443	□
geographic names	47	□
groups by age and sex	40	□
healthcare	63	□
physician	12	□
practice guideline	12	☑
hospital	7	□
anesthesist	6	□
emergency ward	5	□
ambulance	4	□
dentist	4	□
general practice	4	□
long term care	4	□
medicare	4	□
coroner	3	□
general practitioner	2	□
health maintenance organization	2	□
hospital management	2	□

Documents 1 to 10 out of 23 previous next

Contents of "blood clot lysis":

1. Antithrombotic Therapy for Acute Myocardial Infarction.
O'Donnell, C.J.; Ridker, P.M.; Hebert, P.R.; Hennekens, C.H. (1995). Full-length article. Journal of the American College of Cardiology 25 (7), pp. 23S-29S.
[http://linkinghub.elsevier.com/pii/0735109795001050]

2. Can the MADIT Results Be Applied to Myocardial Infarction Patients at Hospital Discharge?.

Figure 8: Screenshot of the DOPE Prototype

A new query can be started by typing in a search string. This will empty the rest of the interface and load a new set of documents and co-occurring keywords. The Thesaurus Browser provides an alternate starting point for a next query. When a focus keyword has been selected, the user can click the "Navigate Thesaurus..." button at the upper left. He is then confronted with a dialog that lets him select a new focus keyword, by browsing through the thesaurus, starting from the focus keyword. The user can iteratively select a broader, narrower or alternative keyword until a keyword has been selected as new focus keyword.

The visualization conveys several types of information. First, the user obviously sees document characteristics such as index terms and article types. Visualizing a set of keywords shows all Boolean combinations, without the need to express them all separately. Furthermore, the graph also shows within the scope of the selected set of documents how these keywords relate, i.e. if they have some overlap and if so, which documents constitute that overlap. Consequently, the geometric distance between keywords or between documents is an indication of their semantic distance: keywords that share documents are located near one another, and so are documents with the same or similar keyword memberships.

Conclusions

In this edition of the column we tried to argue that there is a great potential in combining thesauri with RDF technologies. Thesauri offer terminological knowledge for various domains which have often been carefully designed and maintained and are therefore an invaluable asset for intelligent applications. RDF on the other hand is becoming the major format for encoding semantic information in machine readable format and a wide range of tools for RDF have been developed that ease the development of complex applications significantly. Combining the content of existing thesauri with an RDF encoding is a win-win situation because thesauri provide the semantic information needed for many applications and RDF provides the infrastructure for exchanging, linking and using this information in applications. The relevance of these advantages can be seen by the increasing number of commercial institutions being involved in related activities. Examples are the two projects SWAD-E and DOPE mentioned above. Despite this increasing interest, there is still a lack of a commonly agreed standard for representing thesauri in RDF. Coming up with such a standard would be a major step towards a wider adoption of the ideas laid out in this article. Further, more experiences with actual conversion is needed to make the encoding of thesauri in RDF a routine process and enable a wider range of applications to benefit from this interesting combination

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.

Ontology Matching for Search Engines

by
H. Peter Alesso
www.web-iq.com

Overview

Since the Semantic Web is distributive, numerous resource descriptions are introduced where two concepts are equivalent, but are described using different terms. The resolution of such terminology conflicts requires ontology engineering. The word ontology refers to a hierarchical data structure containing the relevant entities and their relationships.

In the field of Artificial Intelligence (AI), ontology applications have been developed for knowledge management, e-Commerce, education, and for new emerging directions like the Semantic Web.

The realization of the Semantic Web will require the construction of ontologies for the various representation languages, query languages, and inference technology that are needed. A semantic search engine would require ontology matching between various sources. It would involve both "one-to-one" mappings of concepts plus complex mappings such as "many-to-one" mappings and attribute mappings.

In this article, we discuss ontology construction, matching, and mapping which support semantic search engines. We present three types of mapping and an example mapping tool for each. As new tools and algorithms are developed and tested, the best algorithms can be incorporated into search engine for comparison.

Searching for Content

Today, the huge amount of information on the Web inhibits accurate search. Search results can often produce volumes of irrelevant references while failing to find the most desirable relationships. Because Web search engines use keywords they 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 characteristics fundamental to search engines performance. It is important to consider useful searches as distinct from fruitless ones. There are three necessary criteria to evaluate searches as useful:

- 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.

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 that they could fine-tune any search based on our past interests.

Since the Semantic Web is distributive, there are a lot of resource descriptions where two concepts are equivalent, but they are described using different terms. The Semantic Web should enable information to be exchanged between applications that are built using different terminology for similar concepts (ontologies). Resolving polysemy and synonymy terms is central to information processing across ontologies and requires mapping equivalent terms. A semantic search engine should be able to process ontology mapping automatically.

Semantic Web data expressed in OWL or RDF documents can use ontology matching for search engines. When the user inputs a query, the program transfers it to a machine processing agent which evaluates the similarity between different ontologies and returns the matched item to the user. The semantic search engine will list the most similar concepts to the ontology that is entered as a keyword input. The system should consider both lexical similarities and textual similarities. In order to achieve this goal, it is necessary to compute the lexical similarities of the structures of concepts among ontologies.

For example, suppose we want to relate a postal code and an address. The Web is unable to represent this relationship since the names of the postal code system may be different in different countries and as a result we may not get what we expect.

By contrast, the Semantic Web could map the equivalence of a zip code to an appropriate postal code in another country. However, it is difficult to utilize such data on a large scale. To make the Semantic Web work, well-structured data and rules are necessary for agents to roam the Web.

Semantic Search Algorithm

A semantic search engine uses ontology matching to produce a remarkably useful application. It can discover if two documents are similar even if they do not have any specific words in common and it can reject documents that share only uninteresting words in common.

In broad terms what is involved is an algorithm that forms a Web of documents and words - connecting all documents to all words. Given such a model of words and documents one can then establish values based on the distance of documents from each other. The 'value' of any document to any other document might be designated as a function of the number of connections that must be traversed to establish a connection between documents. If two documents are connected by multiple routes then those documents might have a high degree of correlation.

Some of the preparatory work needed to get documents ready for comparison is very language-specific, such as, stemming. For English documents, we could use an algorithm called the Porter stemmer to remove common endings from words, leaving behind an invariant root form.

The implementation algorithm for semantic search could follow as:

For each document:

1. Stem all of the words and throw away any common 'noise' words:
   a) Make a complete list of all the words that appear in the collection
   b) Discard articles, prepositions, and conjunctions
   c) Discard common verbs (know, see, do, be)
   d) Discard pronouns
   e) Discard common adjectives (big, late, high)
   f) Discard frilly words (therefore, thus, however, albeit, etc.)
   g) Discard any words that appear in every document
   h) Discard any words that appear in only one document

2. For each of the remaining words perform a ontology matching evaluation:
   a) Visit and remember each document that has a direct relationship to this word.
   b) Score each document based on a distance function from the original document and the relative scarcity of the word in common.

3. For each of the as-of-yet-unvisited new related documents now being tracked

Recursively perform the same operation as above.

One possible weighting algorithm could work like this: For each increase in distance, divide a baseline score by two. Then the score of each document is equal to the baseline divided by the square root of the popularity of the word.

Overall this algorithm delivers a cheap semantic lookup based on walking a document and word graph. There are many other scoring algorithms that could be used.

The idea is then that the weighting algorithm feeds input into the semantic algorithm which first stems the words appropriately, scores them according to the semantic algorithm and sorts the results into the new rank order reflecting the semantic analysis.

Ontology

The word ontology was originally applied in the field of Philosophy to the concept of existence. It is the theory of objects and their interrelationships. As used in information science, the term ontology frequently refers to a hierarchical data structure containing the relevant entities and their relationships and rules within that domain. An ontology which is not tied to a particular problem domain but attempts to describe general entities is known as a foundation ontology or upper ontology.

The realization of the Semantic Web will require the construction of ontologies for the various representation languages, query languages, and inference technology that are needed.

In the following example, we will describe an ontology for a 'spot.' This ontology consists of three owl:Classes (spot, ellipse, and point) and six rdf:Properties (shape, center, x-position,y-position, x-radius, y-radius). Together, these vocabularies can be used to describe a spot. Figure 1 organizes the relationships for these elements.

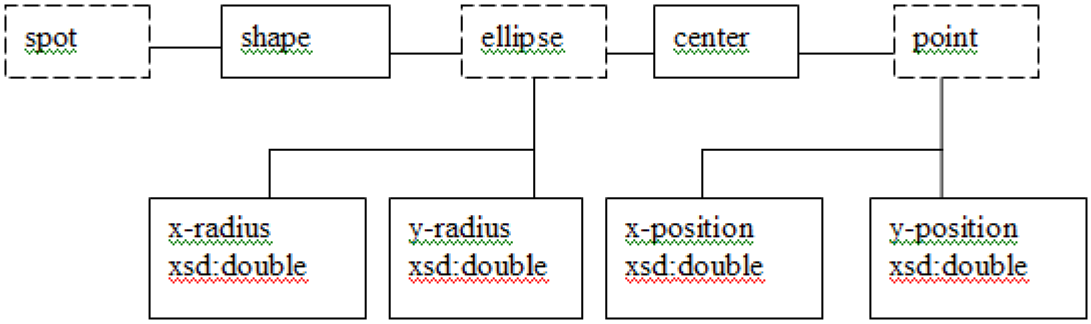


Figure 1 Example Ontology O1 for 'spot'

Classes

The three OWL classes are:

Spot - Conceptually, a "spot" of a Two Dimensional defined as a closed region on the plane of 2D.

Point – A point is defined as a location on a Cartesian plane. A Point has two attributes; its x-position and y-position on an implicit coordinate system of the plane.

Ellipse - Ellipse here is defined as a circle stretched along either x- or y-axis of a coordinate system. The major and minor axis of an Ellipse shall be parallel to the coordinates of the implicit coordinate system (See Figure 2).

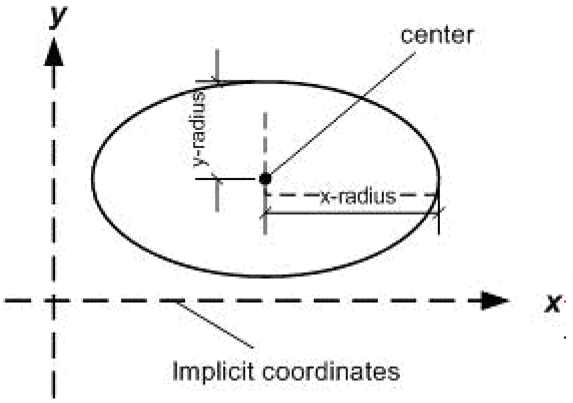


Figure 2 Definition of an Ellipse.

Properties

The six RDF properties are:

Shape – A Spot is defined to be a closed region of a 2D plane, a Spot therefore must assume a shape of certain kind and we require that the shape be an Ellipse. Therefore the domain of shape is Spot and the range of Spot is Ellipse.

Center - The center refers to the center point of the Ellipse. Therefore, it has an rdfs:domain of Ellipse and an rdfs:range of Point.

x-position - An x-position is an owl:Datatype property that has a domain of Point. Its value (of type xsd:double) indicates its distance from the origin on the x-axis of the coordinate system.

y-position - A y-position is a owl:Datatype property that has a domain of Point. Its value (of type xsd:double) indicates its distance from the origin on the y-axis of the coordinate system.

x-radius- A x-radius is a owl:Datatype property that has a rdfs:domain of Ellipse. It refers to the radius that is parallel to the x-axis of the coordinate system (see Figure 2).

y-radius - A y-radius is a owl:Datatype property that has a rdfs:domain of Ellipse. It refers to the radius that is parallel to the y-axis of the coordinate system (see Figure 2)

The OWL file for this ontology example is as follows:

```
     <?xml version="1.0" encoding="iso-8859-1" ?>
      <!DOCTYPE rdf:RDF (...)>
    <rdf:RDF xmlns="http:/ / example#"
        xmlns:example="http:/ /        example#"        xmlns:rdf="http:/ / www.w3.org / 1999 / 02 / 22-rdf-syntax-ns#"
        xmlns:rdfs="http:/ / www.w3.org / 2000 / 01 / rdf-schema#"
        xmlns:xsd="http:/ / www.w3.org / 2001 / XMLSchema#"        xmlns:owl="http:/ / www.w3.org / 2002 / 07 / owl#"
        xmlns:dc="http:/ / purl.org / dc / elements / 1.1 / "
        xml:base="http:/ /  / example">
   <owl:Ontology rdf:about="">
  <rdfs:isDefinedBy rdf:resource="http:// example/" />
  <dc:author>Smith</ dc:author>
  <dc:title>Example Ontology</ dc:title>
  <rdfs:comment>This file defines a partial ontology in OWL</ rdfs:comment>
  <owl:versionInfo> 2005</ owl:versionInfo>
      </ owl:Ontology>
  <owl:Class rdf:ID="Spot" />
  <owl:Class rdf:ID="Ellipse" />
  <owl:Class rdf:ID="Point" />
  <owl:ObjectProperty rdf:ID="shape">
  <rdfs:domain rdf:resource="#Spot" />
  <rdfs:range rdf:resource="#Ellipse" />
      </ owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="center">
  <rdfs:domain rdf:resource="#Ellipse" />
  <rdfs:range rdf:resource="#Point" />
      </ owl:ObjectProperty>
  <owl:DatatypeProperty rdf:ID="x-radius">
  <rdfs:domain rdf:resource="#Ellipse" />
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double" />
      </ owl:DatatypeProperty>
   <owl:DatatypeProperty rdf:ID="y-radius">
  <rdfs:domain rdf:resource="#Ellipse" />
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double" />
      </ owl:DatatypeProperty>
   <owl:DatatypeProperty rdf:ID="x-position">
  <rdfs:domain rdf:resource="#Point" />
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double" />
      </ owl:DatatypeProperty>
   <owl:DatatypeProperty rdf:ID="y-position">
```

```
        <rdfs:domain rdf:resource="#Point" />
        <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double" />
           </owl:DatatypeProperty>
           </rdf:RDF>
```

Ontology describes concepts and relationships with a set of representational vocabulary. The aim of building ontologies is to share and reuse knowledge. Since the Semantic Web is a distributed network, there are different ontologies that describe semantically equivalent things. As a result, it is necessary to map elements of these ontologies if we want to process information on the scale of the Web.

Ontology Matching

An ontology typically provides a vocabulary that describes a domain of interest and a specification of the meaning of terms used in the vocabulary. Depending on the precision of this specification, the notion of ontology encompasses several conceptual models, for example, classifications, database schemas, fully axiomatized theories. However, in open or evolving systems, such as the Semantic Web, different parties could adopt different ontologies.

Ontology matching is a promising solution to the semantic heterogeneity problem. It finds correspondences between semantically related entities of the ontologies. These correspondences can be used for various tasks, such as ontology merging, query answering, data translation, or for navigation on the Semantic Web. Thus, matching ontologies enables the knowledge and data expressed in the matched ontologies to interoperate.

The problem of finding the semantic mappings between two given ontologies is called ontology matching. This problem lies at the heart of numerous information processing applications. Virtually any application that involves multiple ontologies must establish semantic mappings to ensure interoperability.

Despite its pervasiveness, today ontology matching is still largely conducted by hand, in a labor-intensive and error-prone process. The manual matching has now become a key bottleneck in building large-scale information management systems.

String Matching

String matching is widely used in many fields such as text processing, image and signal processing, information retrieval, pattern recognition, and pattern matching in large databases. There are many string matching methods, among which edit distance is a well-known method for measuring the similarities of two strings. Suppose we have two strings S1 and S2, if we use limited steps of edit operations (insertions, deletions, and substitutions of characters), S1 can be transformed to S2 . Such a sequence is called edit sequence. The edit distance is defined as the weight of edit sequence.

String matching can help in ontology matching. The existing ontology files on the Web (see http://www.daml.org/ontologies) show that people usually use similar elements to build ontologies, although the complexity of each ontology may be different, and the terms may vary. This is because there are some properties necessary to describe a concept and people have already established a name.

The value of String Matching lies in that it can estimate the lexical similarity. However, we also need to consider the real meaning of the words and the context. In addition, there are some words that are similar in alphabet form while they have really different meaning such as "too" and "to". Hence, it is not enough if we only use string matching. Other approaches to improve the performance include normalization. Since edit distance does not consider the lengths of the strings compared, which may produce a side effect — a pair of long strings that differ only in one character may get the same edit distance as that of a pair of short strings. For example, suppose two words whose lengths are both 1000 only differ in one character, suppose further another pair of words whose lengths are both 3 also differ in one character. If we use traditional method to compute their edit distance, we will get exactly the same value. However, this is an unfair result since the long strings should get higher value. In this case, we should use an efficient uniform-cost normalized edit distance.

Comparing Ontologies

An ontology can be represented in a taxonomy tree where each node represents a concept with its attributes. Figure 3 shows a different spot (see Figure 1) ontologies in tree form. The concept spot on the left of Figure 1 should be recognizable as an equivalent match to the spot on the left of Figure 3. We can readily see that the only difference between these two figures (ontologies) is the term 'point' has been replaced by 'origin.'

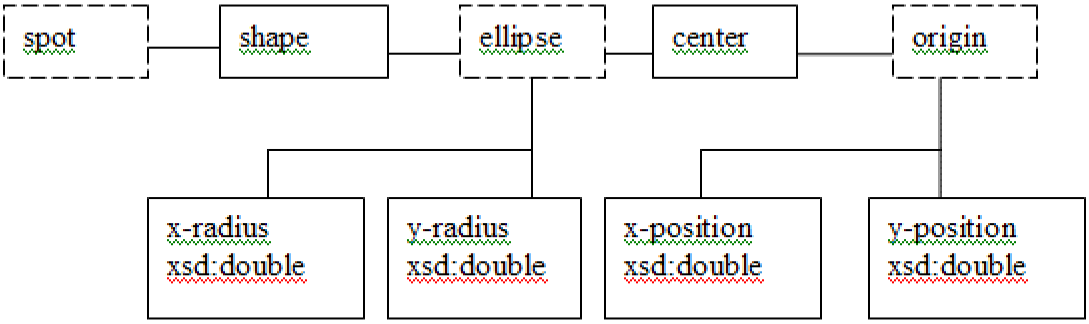


Figure 3 Ontology O2 for 'spot'

The aim of ontology matching is to map the semantically equivalent elements. This is a one-to-one mapping of the simplest type. We can also map the different types of elements: e.g., a particular relation maps to a particular attribute. Mapping can be more complex if we want to map the combination of some elements to a specific element.

An approach for semantic search can be based on text categorization for ontology mapping than compares each element of an ontology with each element of the other ontology, and then determines a similarity metric on a per pair basis. Matched items are those whose similarity values are greater than a certain threshold.

An ontology can be represented in taxonomy tree form where each node represents a concept with its attributes. Figures 1 and 3, show two different ontologies. For example, the concept spot on the left of Figure 1 has three classes: spot, ellipse, and point. The aim of ontology matching is to map the semantically equivalent elements. For example, "point" maps to "origin" in Figure 3. This is a one-to-one mapping, the simplest type. We can also map the different types of elements, e.g. a particular relation maps to a particular attribute.

Similarity measures play a very significant role in ontology matching. A set of similarity measures which measures similarity between ontologies at two semiotic levels — the lexical level and the conceptual level is different from previous approaches that use the formal structures of ontologies and match the concept nodes. However, all ontologies in real-world not only specify the conceptualization by logical structures, but also refer to terms restricted by human natural languages use. For the two ontologiesO1 and O2, we could compute a similarity measure and use that measure to decide whether the onologies match. We could apply the notion of the joint probability distribution between any two concepts as a similarity measure.

It is important to give a definition of similarity between two concepts because this can make aims clear and facilitate the evaluation process. There are many measures solely based on the joint probability distribution.

Ontology Mapping

Ontology mapping enables interoperability among different sources in the Semantic Web. It is required for combing distributed and heterogeneous ontologies. Ontology mapping transforms the source ontology into the target ontology based on semantic relations at conceptual level. There are three approaches for combing distributed and heterogeneous ontologies:

1. mapping between local ontologies,
2. mapping between integrated global ontology and local ontologies, and
3. mapping on ontology merging, integration, or alignment.

Ontology merge, integration, and alignment can be considered as ontology reuse process.

Ontology merge is the process of generating a single, coherent ontology from two or more existing and different ontologies on the same subject. The original ontologies have similar or overlapping domains but they are different and not revisions of the same ones.

Ontology integration is the process of generating a single ontology from two or more differing ontologies in different subjects. The subjects of the different ontologies may be related.

Ontology alignment creates links between two original ontologies. Ontology alignment is made if the sources become consistent with each other but are kept separate.

Ontology Mapping Tools

In this section, we review three types of ontology mapping tools and provide an example of each.

1/. Ontology mapping between local ontologies.

Example tool: GLUE

GLUE is a system which semi-automatically creates ontology mapping using machine learning techniques. Given two ontologies, GLUE finds the most similar concept in the other ontology. For similarity measurement between two concepts, GLUE calculates the joint probability distribution of the concepts. GLUE uses a multi-strategy learning approach for finding joint probability distribution.

GLUE has a Content Learner and Name Learner, and a Meta Learner.. The Content Learner exploits the frequencies of words in the textual content of an instance in order to make predictions and uses the Naïve Bayes's theorem. The Name Learner uses full name of the input instance. The Meta-learner combines the predictions of base learners and assigns weight to base learners based on how much it trusts that learner's predictions.

2/. Ontology mappings between source ontology and integrated global ontology.

Example Tool LSD:

In LSD (Learning Source Description), Schema can be viewed as ontologies with restricted relationship types. Therefore, the mediated schema can be considered as a global ontology. The LSD system uses a multi-strategy learning two phase approach: training and matching. In the matching phase, prediction combiner combines meta-learner's prediction and match the schema of new input source to the mediated schema. This process can be considered as ontology mapping between information sources and a global ontology.

3/. Ontology mapping in ontology merging, alignment, and integration.

Example Tool OntoMorph:

OntoMorph provides a powerful rule language for specifying mappings, and facilitates ontology merging and the rapid generation of knowledge base translators. It combines two syntactic rewriting and semantic rewriting. Syntactic rewriting is done through pattern-directed rewrite rules for sentence-level transformation based on pattern matching. Semantic rewriting is done through semantic models and logical inference.

Conclusion

The development of the Semantic Web and semantic search requires much more development in ontology engineering. Areas such as ontology mapping are still in there early stages of development. As new tools and algorithms are developed and tested, the best algorithms can be incorporated into search engine for comparison.

Reference

1/. Alesso, H. P., *Developing Semantic Web Services*, ISBN: 1568812124, A. K. Peters, Ltd., 2004.

2/. H. Chalupsky. “Ontomorph: A Translation system for symbolic knowledge” In Principles of Knowledge Rep-resentation and Reasoning, 2000.

3/. Doan, A., Madhavan, J., Domingos, P., Halevy, A., "Learning to Map Between Ontologies on the Semantic Web, May 2002, http://www2002.org/CDROM/refereed/232/

4/. Doan, A., Domingos, P., Halevy, A., “Learning to Match the Schemas of Data Sources: A Multistrategy Approach”, Machine Learning 50 (3): 279-301, March 2003,

5/. Wang, P., “A Search Engine Based on the Semantic Web,” M.S. Thesis, University of Bristol, September 2003.

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.

The Semantic Web Technologies Column by Dr. Jessica Chen Burger

Unleashing Business Processes through Semantic Web – a Fact or Myth?

Yun-Heh Chen-Burger
AIAI, CISA, Informatics, the University of Edinburgh, UK

Business process modelling (BPM) is nowadays a common practice that is regarded to have tremendous value when businesses wish to analyse, design and redesign their operations in order to optimise their work performance. This concept of being able to capture business processes and describe them in a more formal and reusable format, however, was not initially recognised. Business processes are often thought of as informal. They typically vary in different organisations and may change in circumstances. They may also be carried out differently depending on the person who implements it. In addition, these processes are often renewable that they must evolve through time and response to changes within the environment that they operate. As a result, in the past there is often only limited amount of processes within an organisation that are more clearly described, typically they are those directly related to IT automations.

This view was largely altered when a MIT lead project conducted their research in search of generic and reusable business and process components across different business companies and sectors. This enables them to place these commonly shared components in structured hierarchies. This work involved over forty university and industrial partners. Their results are published in the MIP Process Handbook [6]. A related attempt was lead by NIST's PIF [9] and PSL [10] initiatives for manufacturing processes; some 250 industrial and university partners are involved to create a common ontology and interchange language to promote the sharing between process concepts and model primitives in different applications.

In addition to sharing business practice, process modelling technologies clearly provide great technical advantages towards implementing a correct and appropriate workflow system. For instance, they allow critical analysis and simulation of possible scenarios before one commits cost to build the actual system. However, according to Delphi's survey in 2002 that before the year of 2000 only very few workflow systems are built based on business process modelling methods. Instead, this approach was only recognised and taken up after the year of 2000. Since then, the concept of linking BPM to workflow systems has been adopted rapidly. Today, BPM plays a dominant role in Business Process Re-Engineering, Business Process Change, Business Analysis, and Workflow Management. More recently, it has been recognised to be a vital and integrated part within the field of Knowledge Management.

Through the popularity and strong recognition of the importance of Semantic Web (SW). BPM and process modelling communities have been trying to bridge themselves and ride the SW wave. They are interested in several fronts. They want to be able to describe their processes in a semantically meaningfully way – and more importantly, consistently over the web. They also want to describe their processes in standardised fashion so that they facility communication between applications. Example previous efforts of these types are OASIS's standard work [7], BPML [2] and later on OWL-S [8]. More over, these communities want not just to describe their processes - they want to enact them and particularly through web services. One very good example is BPEL4WS [1]. However, despite their best attempts none of them has provided sufficient support for enabling semantically rich services via the Web. As a result, a newer breed of the process language WSML and WSMO [11] together try to tackle this problem.

Although WSML and WSMO are too new to make a definite judgement, it is relatively clear to see that current other existing efforts are not proficient to provide support that will unleash the full potential of business processes through the Web. Upon translating and mapping a typical BPM language onto OWL-S, it was found that various typical business operations could not be expressed easily in OWL-S [4]. This is partly because OWL-S provides mainly high-level concepts that lack constructs to enable more sophisticated co-ordinations, such as temporal synchronisation, between processes. Similar attempts were also conducted to map to BPEL4WS and BPML [5]. The latter two methods, however, suffer from the inability of describing their data model independently when representing their process models. OWL-S has advantage regarding this issue, as it may be used in conjunction with OWL that is an ontological language and is native for describing data.

Ideally, to fully support and enact a business process system over the web, a few properties should be considered. For example:

- To truly integrate data and process models in their languages;
- To enable a direct link to organisational structure, function and rationale, so that business decisions may have a direct influence on web based business operations;
- To allow explicit use of existing and prominent standard ontologies, including data as well as business process ontologies;
- Provision for sophisticated agent coordination and temporal synchronisation between separately run processes;
- To facilitate explicit representation of process life cycle, i.e. to treat a process as an independent entity, so that it may be tracked, monitored, suspended and repaired according to business rationale; the same point is also applicable for representing data life cycle;
- To provide a framework for explicit representation of error and recovery mechanism, in particular for a distributed and peer to peer environment;
- Finally yet importantly is to simplify the too many layers of the "semantic-layered cake for web-based workflow systems" [3].

There is definitely a great deal of potential for enacting BPM based workflow systems over the Web and there are plenty of advantages to be gained. For instance, it provides for a much more open and interactive environment that allows the participation of any arbitrary or purposely-selected agents. Agents are also able to carry out more agile, speedy and autonomous decision-making and business operations based on (own) goals. Such tasks may be standardised; or can be dynamically adapted at run time. They may be related to business communication, deals making, contract negotiation and secure business transactions over the Web. When and if all of these facilities are realised, this new breed of web and knowledge based autonomous business systems should revolutionise the conventional ways of carrying out business and thinking about doing business.

Reference:

[1] BPEL4WS: http://www-128.ibm.com/developerworks/library/specification/ws-bpel/.
[2] BPML: http://www.bpmi.org/.
[3] Yun-Heh Chen-Burger. e-Science Workflow Services Workshop, December 3-5, 2003, National e-Science Center, Edinburgh, UK.
[4] Li Guo, Yun-Heh Chen-Burger, Dave Robertson (2004) Mapping a business process model to a semantic web services model. In proceedings of 2004 IEEE International Conference on Web Services, July 6-9, 2004, San Diego, California, USA.
[5] Li Guo, Dave Robertson, Yun-Heh Chen-Burger (2005) Enacting the Distributed Business Workflows Using BPEL4WS on the Multi-Agent Platform. Third German Conference on Multiagent System Technologies. University of Koblenz-Laudau, Koblenz, Germany. September 11-13, 2005.
[6] Organizing Business Knowledge. The MIT Process Handbook. Ed. Thomas W. Malone, Kevin Crowston and George A. Herman. 2003.
[7] OASIS standard work: http://www.oasis-open.org/specs/index.php.
[8] OWL-S: http://www.daml.org/services/owl-s/1.0/.
[9] PIF: http://ccs.mit.edu/pif/.
[10] PSL: http://www.mel.nist.gov/psl/.
[11] WSMO, WSMO and WSMX: http://www.wsmo.org/TR/.

The Semantic Desktop Grapevine by Leo Sauermann

Leo Sauermann is researcher at the *German Research Center for Artificial Intelligence DFKI* in the *Knowledge Management Lab* and member of the *Competence Center Semantic Web CCSW* at the DFKI. He studied information science at the Vienna University of Technology and graduated 2004. Under the project name "gnowsis" he merged *personal information management* with *semantic web* technologies, resulting in a master thesis about „Using Semantic Web technologies to build a Semantic Desktop“. Working as a researcher at the DFKI since 2004, he continued and now maintains the associated open-source project *gnowsis.org*. His research focus is on practical semantic web and its use in knowledge management, his vision is to bring the semantic web alive by the paradigm of the *semantic desktop*. In autumn 2003 he started to give talks about his work - before his master's examination was nominated as best thesis, it was already a success. His work was published on several conferences, he founded the *Semantic Web Lounge Austria*, is lecturer at the *Semantic Web School* in Vienna and co-chair of the *Semantic Desktop Workshop* at the ISWC. He is co-initiator of the EU-IST integrated project *Nepomuk Social Semantic Desktop*.
From 1998 to 2002 he has been working in several small software companies, including the position of lead architect at *Impact Business Computing* developing mobile CRM solutions. He is an expert programmer in Delphi and Java and C#.

About the Column
In the following years you will find information about and around the *Semantic Desktop* in this column. The idea of the topic is to bring semantic web technologies to desktop computers, allowing users to benefit from the semantic web. Three main directions are at hand:

- authoring semantic web content,
- publishing this content
- and benefiting from content published by others.

As the majority of information today is entered using applications running on desktop computers, they are the key to the semantic web, being it a browser, a peer-to-peer application or an email client: the content will come from everyday applications and from everyday users doing everyday tasks.

First the column is focused at developers, researchers and journalists. In the future, when we have a better and more developed semantic web, we will provide information about success stories of semantic desktop end users. You will find:

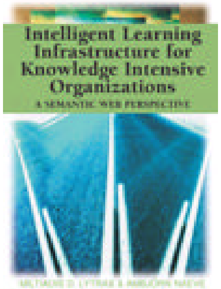
- Visionary papers about the idea of the semantic desktop itself, possible architectures, ontologies, data exchange formats and application scenarios.
- Product descriptions of existing parts that are building blocks of the semantic desktop
- Success stories and best practice reports from ongoing semantic desktop projects and participants.
- News, rants, ideas, dispute, discussion, comments, interviews and provoking articles about the current state of affairs and the participants.

The content of this column will consist of short articles by the columnist and work of others involved in the topic. It will not be limited to scientific papers but open to reports from open source programmers and the industry.

Semantic Web Books Column

INTELLIGENT LEARNING INFRASTRUCTURES FOR KNOWLEDGE INTENSIVE ORGANIZATIONS: A SEMANTIC WEB PERSPECTIVE,
Miltiadis Lytras and Ambjorn Naeve (eds),
IDEA Group Publishing

http://www.idea-group.com/books/details.asp?id=4925



A best selling book:
http://www.amazon.co.uk/exec/obidos/tg/browse/-/14166431/202-9850453-7335042

The only one with a teaching orientation fro Semantic Knowledge and Learning Management!!!

At the beginning of each chapter we provide the ***Editor's Note*** section, where we give our basic understanding for the chapter. Authors also provide a relevant section entitled ***Inside Chapter, which*** is an abstract like short synopsis of their chapter.

At the end of each chapter there are some very interesting sections, where reader can spend many creative hours. More specifically the relevant sections are entitled:

- **Internet Session:** In this section we present one more web sites, relevant to the discussed theme in each chapter. The short presentation of each chapter is accompanied by the description of an ***Interaction*** where the reader (student) is asked to make a guided tour in the web site and to complete an assignment.
- **Case Study:** For each chapter, we provide "realistic" descriptions for one or more case studies that reader must consider in order give strategic advice. The questions exploit the key concepts and technologies presented in the chapters. Of course as the reader reaches the last chapters these case studies can be analyzed in more detail and answers have to combine directions given in several chapters.
- **Useful links**: They refer to web sites, with content capable of exploiting the knowledge communicated in each chapter. We decided to provide these links in every chapter, even though we know that several of them will be broken in a time horizon, since their synergy with the content of the chapter can support the final learning outcome.
- **Further Readings:** These refer to high quality articles available both in web and electronic libraries. We have evaluated these resources as of significant value and conditionally can initiate criticism and creative ideas. For sure the reader can spend many hours with these resources.
- **Essays:** Under this section a number of titles for assignments are given. In the best case essays could be Working Research Papers. The general rule is that we provide 3 to 6 essays' titles in each chapter and in their abstract title the reader can find an excellent context of questioning. The ultimate objective is the knowledge delivered in each chapter to be exploited towards a scientific document that will provide a thesis of the reader.

Table Of Contents

Call for Chapters

Open Source for Knowledge and Learning Management: Strategies beyond Tools

An edited book for publication by Idea Group Inc.

Editors:

Miltiadis Lytras, Research Academic Computer Technology Institute, ELTRUN, Athens University of Economics and Business, Greece and AIS SIGSEMIS,
Email: Lytras@ceid.upatras.gr

Ambjorn Naeve, Head of the Knowledge Management Research group, Computer Science and Communication (NADA), Royal Institute of Technology (KTH), Sweden, Email: amb@nada.kth.se

THERE ARE STILL 2-3 OPEN SLOTS for CHAPTERS!!! Pls send a mail to Miltiadis Lytras

Call for Chapters

Our vision for Knowledge and Learning Management has an ultimate objective: The promotion of a world of peace and prosperity through knowledge and learning for everyone. In this context, two significant shifts in recent years are evident:

- On the one hand, Knowledge and Learning Management is considered as a key milestone towards a Humanistic Approach to Information Systems and New Technologies. The social responsibility through knowledge and learning management is promoted further with worldwide initiatives that provide unforeseen knowledge and learning experiences.
- On the other hand, a kind of new philosophical paradigm, the Open Source Paradigm, shows in practical ways the benefits of trusted and value adding knowledge sharing and semantic social integration. The communities of Open Source, the Tools of Open Source and the People who share this vision cultivate jointly a key context for an emerging type of a new whole Research and Practice Era: OPEN RESEARCH is getting continuously more and more space in the global knowledge-driven economy. This book brings together these two critical aspects and sets the full range of the relevant research agenda. Furthermore, it has a practical orientation. A number of Open Source Knowledge and Learning Management Systems are yet available, and the discussion is focused on the strategies that they support.

Additionally, a transparent editing strategy is applied in the context described. Knowledge and Learning Management is the Context, Open Source Paradigm is the Tool and Human Computing is the Philosophy that brings together our key understanding for the next generation of Information Systems: Human Information Systems is the key word!

Recommended topics fall within the following four sections:

SECTION A: OPEN SOURCE and OPEN RESEARCH: From a paradigm to a vision

- The Open Source Paradigm: Philosophical Routes, Demonstration of Various Communities, Success Stories, Lessons Learned
- Open Source and Knowledge Sharing: Effective Strategies, Collaboration Models and Types, Integration Issues, Extensibility, Interoperability

- Open Research: A vision for dissemination of knowledge beyond limits

SECTION B: Open Source Insights to KNOWLEDGE AND LEARNING MANAGEMENT

- Challenges for Knowledge and Learning Management
- Open Source for Knowledge Management Systems: Approaches/ Demonstrations/ Projects/ Cases/ Lessons Learned
- Open Source for Learning Management Systems: Approaches/ Demonstrations/ Projects/ Cases/ Lessons Learned

SECTION C: Business and Societal Applications of Open Source Approaches to KNOWLEDGE AND LEARNING MANAGEMENT

- Business Challenges for the deployment of Knowledge and Learning Management Systems
- Knowledge and Learning Systems for specific Business/Societal objectives (e.g. Internal Performance, Customer Support, etc.) in specific sectors (e.g. Health, Pharmaceutical, Academia, Learning Industry, People with Disabilities, Knowledge Society, etc).

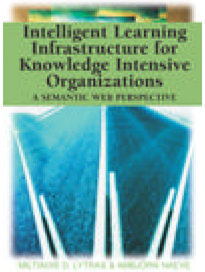
SECTION D: Government Policies towards the promotion of Open Source to Knowledge and Learning Management

- Summary of current initiatives at Governmental level
- Challenges for the future / Specification of Government Policies for the Promotion of Open Source for Knowledge and Learning Management.
- Roadmaps for the future

Important Dates

Researchers and practitioners are invited to submit on or before July 15th, 2005, a 2-page manuscript proposal clearly explaining the mission and concerns of the proposed chapter. Authors of accepted proposals will be notified by July 30, 2005 about the status of their proposals and sent chapter organizational guidelines. Full chapters are expected to be submitted by October 15, 2005. All submitted chapters will be reviewed on a blind peer review basis. The book is scheduled to be published by Idea Group Inc., www.idea-group.com, publisher of the Idea Group Publishing, Information Science Publishing, IRM Press, CyberTech Publishing and Idea Group Reference imprints. Chapter proposal submissions (Word document) can be forwarded via email to: **Lytras@ceid.upatras.gr and cc: amb@nada.kth.se**

15th July 2005	Chapter Proposals Due
30th July 2005	Notification of accepted chapter proposals
15th October 2005	Full Chapters Due
15th December 2005	Revised Chapters Due
30th January 2006	Camera Ready Chapters
Late 2006	Publication


Intelligent Learning Infrastructure for Knowledge Intensive Organizations
A SEMANTIC WEB PERSPECTIVE
MILTIADIS D. LYTRAS & AMBJÖRN NAEVE

Our Just Published Book: Don't Miss it!

Intelligent Learning Infrastructure for Knowledge Intensive Organizations: A Semantic Web Perspective

Miltiadis D. Lytras , Ambjörn Naeve; Research Academic Computer Technology Institute, Greece and AIS SIGSEMIS; Royal Institute of Technology, Sweden

http://www.idea-group.com/books/details.asp?id=4925

Ubiquitous and Pervasive Knowledge and Learning Management:

Semantics, Social Networking and New Media to their full potential

Miltiadis Lytras, Research Academic Computer Technology Institute, RU5,
ELTRUN, Athens University of Economics and Business, Greece and AIS SIGSEMIS, Email: Lytras@ceid.upatras.gr
Ambjorn Naeve, Head of the Knowledge Management Research group, Computer Science and Communication (NADA), Royal Institute of Technology (KTH), Sweden, Email: amb@nada.kth.se

THERE ARE STILL 2-3 OPEN SLOTS for CHAPTERS!!! Pls send a mail to Miltiadis Lytras

Call for Chapters

The concepts of Ubiquity and Pervasiveness contribute to our vision for a world of Peace and Prosperity through knowledge and learning for everyone. Last Years the advent of New Media (Mobile and wireless technologies as well as the fascinating evolution in broadcasting) have set the context for Ubiquitous and Pervasive Information Systems.

In a way our ultimate objective for the emerging Digital Divide is to provide value in new contexts. Context Modelling as well as Context Exploitation requires an integrative approach to Semantics, Social Networking and New Media (Mobile and Wireless Networks, Broadcasting and Cross (New) Media).

This Edited Book goes beyond the traditional research agenda and relevant discussion of M-Learning or Ubiquitous Learning. Two novel characteristics provide the motivation of the book as well as the editing strategy:

- On the one hand the discussion is organized around CONTEXTS. Ubiquitous and Pervasive Knowledge Management is analysed in Business Contexts, Academia, Government (wide scale initiatives for Social Responsibility), in Health, Research, in Culture, in Tourism, etc. This unified approach is due to our deep belief that Knowledge and Learning Management is applied in EVERY context and Ubiquitous and Pervasive Computing enable new contexts critical for the promotion of a better WORLD.
- On the other hand, our “understanding” and our “value proposition” for Ubiquitous and Pervasive Knowledge and Learning Management integrate Semantics, Social Networks and New Media (Mobile and Wireless, Broadcasting). This mix has a critical justification: Semantics enable Meaning and describe Context, Social Networks promote Human Computing and Collaborative Life vision and New Media expand boundaries of services and value delivery.

For this edited book we pursue an ambitious goal: Ubiquitous and Pervasive Knowledge and Learning Management sets a new CONTEXT for the perceived borders of Knowledge and Learning Management. In this CONTEXT we bring new insights and we DO believe that fresh ideas and multidisciplinary approaches contribute to a new era of fascinating HUMAN COMPUTING.
Recommended topics fall within the following four sections:

<u>SECTION A: UBIQUITOUS and PERVASIVE COMPUTING:</u>

- Ubiquitous and Pervasive Computing: Practices, Methods, Demonstrations, Success Stories and Cases
- Enabling Ubiquitous and Pervasive CONTEXTS
- Context Modelling and Exploitation

SECTION B: Ubiquitous and Pervasive Computing insights to KNOWLEDGE AND LEARNING MANAGEMENT

- Challenges for Knowledge and Learning Management
- Context Aware Knowledge and Learning Management Systems
- Ubiquitous and Pervasive Knowledge Management Systems: Approaches / Demonstrations / Projects / Cases / Lessons Learned
- Ubiquitous and Pervasive Learning: Approaches / Demonstrations / Projects / Cases / Lessons Learned
- Architectures, Infrastructures, Designs, Implementations, Agents
- Smart / Ambient (Handheld) Devices, Intelligent Interfaces
- New Media and Cross Media Delivery
- Multimode Annotation of Knowledge and Learning Objects
- Semantics of Ubiquitous and Pervasive Knowledge and Learning Management
- Semantic Social Networking and Pervasive Knowledge and Learning Management

SECTION C: Business and Societal Applications of Ubiquitous and Pervasive Approaches to KNOWLEDGE AND LEARNING MANAGEMENT

- Business Challenges for the deployment of Knowledge and Learning Management Systems
- Knowledge and Learning Systems for specific Business / Societal objectives in specific sectors (Business Contexts, Academia, Government (wide scale initiatives for Social Responsibility), Health, Research, Culture, Tourism, People with Disabilities, Knowledge Society etc).

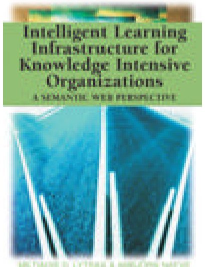
SECTION D: Government Policies towards the promotion of Ubiquitous and Pervasive Knowledge and Learning Management

- Summary of current initiatives at Governmental level
- Challenges for the future / Specification of Government Policies for the Promotion of Ubiquitous and Pervasive Knowledge and Learning Management.
- Roadmaps for the future

Important Dates

Researchers and practitioners are invited to submit on or before July 25th, 2005, a 2 page manuscript proposal clearly explaining the mission and concerns of the proposed chapter. Authors of accepted proposals will be notified by August 5th 30, 2005 about the status of their proposals and sent chapter organizational guidelines. Full chapters are expected to be submitted by October 30th, 2005. All submitted chapters will be reviewed on a blind peer review basis. The book is scheduled to be published by Idea Group, Inc., www.idea-group.com, publisher of the Idea Group Publishing, Information Science Publishing, IRM Press, CyberTech Publishing and Idea Group Reference imprints. Chapter Proposal Submissions (Word Document) can be forwarded via email to: Lytras@ceid.upatras.gr and cc: amb@nada.kth.se

25th July 2005	Chapter Proposals Due
5th August 2005	Notification of accepted chapter proposals
30th October 2005	Full Chapters Due
15th January 2006	Revised Chapters Due
30th March 2006	Camera Ready Chapters
2007	Publication


Intelligent Learning Infrastructure for Knowledge Intensive Organizations
A SEMANTIC WEB PERSPECTIVE
MILTIADIS D. LYTRAS & AMBJÖRN NAEVE

Our Just Published Edited Book: Don't Miss it!!!
Intelligent Learning Infrastructure for Knowledge Intensive Organizations:
A Semantic Web Perspective
Miltiadis D. Lytras , **Ambjörn Naeve**; Research Academic Computer Technology Institute, Greece and AIS SIGSEMIS; Royal Institute of Technology, Sweden
http://www.idea-group.com/books/details.asp?id=4925

I would like to thank Morten for his kindess to provide us his interview with Prof. Winograd!!

An Interview with Terry A. Winograd

Morten Thanning Vendel[1]
Copenhagen Business School
Department of Informatics
Howitzvej 60, 4.
DK-2000 Frederiksberg
Denmark
Phone: +45 38 15 24 08
Fax: +45 38 15 24 01
e-mail: **mtv@cbs.dk**
http://www.cbs.dk/staff/mtv/

[1] My traveling related to this interview was generously funded by an always supportive colleague of mine, Associate Professor Janni Nielsen, Department of Informatics.

An Interview with Terry A. Winograd

I. Introductory Note

Terry A. Winograd (born 1946) began his academic career within the field of artificial intelligence. His early research on natural language understanding by computers was a milestone in artificial intelligence. Later he moved to the field of human computer interaction, and within this field he has done extensive research and writing on design of human-computer interaction. Foremost, focusing on the theoretical background and conceptual models for human-computer interaction design.

Terry Winograd received his B.A. in Mathematics from The Colorado College in 1966. He studied Linguistics at University College, London in 1966-1967, and earned his Ph.D. in Applied Mathematics at Massachusetts Institute of Technology (MIT) in 1970. From 1970 he was an instructor and assistant professor of Electrical Engineering at MIT, before coming to Stanford University in 1973, where he is now a professor of Computer Science. At Stanford, he directs the Project on People, Computers, and Design, and the teaching and research program on Human-Computer Interaction Design. He is one of the principal investigators in the Stanford Digital Libraries project, and the Interactive Workspaces Project. He is also a consultant to Interval Research Corporation, and serves on a number of journal editorial boards, including the Journal of Human Computer Interaction, Computer Supported Cooperative Work, and Personal Technologies.

Prof. Winograd is a longtime advocate for socially responsible computing and is a founding member and past national president (1987-1990) of the Computer Professionals for Social Responsibility, and is on the National Advisory Board of the Association for Software Design. His publications include 'Understanding Natural Language' (1972), 'Understanding Computers and Cognition: A New Foundation for Design' (with Fernando Flores, 1986)[2], and 'Bringing Design to Software' (with John Bennett, Laura De Young, and Bradley Hartfield, 1996, which brings together the perspectives of a number of leading proponents of software design.

II. Terry Winograd and Artificial Intelligence

Prof. Winograd's initial breakthrough in the field of artificial intelligence was with the SHRDLU program, which he wrote at the M.I.T. Artificial Intelligence Laboratory in 1968-1970. SHRDLU is described in his dissertation, issued as MIT AI technical report 235, February 1971 with the title: Procedures as a Representation for data in a Computer Program for Understanding Natural Language. It was published as a full issue of the Journal of Cognitive Psychology, vol. 3, no. 1 (1972), Understanding Natural Language (Academic Press 1972).

SHRDLU understands natural language. It carries on a simple dialog (via teletype) with a user, about a small world of objects (the Blocks World) shown on a display screen. SHRDLU has complete knowledge of the internally represented block world consisting of colored cubes, pyramids and boxes on a flat surface. The program simulates a robot, it accepts commands in English with regard to the block world, carries out the command, and explains how it did it and why certain actions were performed. In addition, it has the ability to learn about new tasks.

SHRDLU uses important ideas about human syntactic semantic and problem solving activities and about their interactions in understanding natural language discourse.

Understanding of English requires an integrated study of syntax semantics and inference. Winograd felt that the best way to experiment with complex models of language was to write a program, which can actually understand language within

some domain. In this case with a robot which has a hand and eye and the ability to manipulate toy blocks. The program attempts;
1) to be a usable language understanding system.
2) to gain a better understanding of what language is and how to put it together.
3) to understand what intelligence is and how it can be put into computers.

[2] In this book they take a critical look at work in artificial intelligence and suggest new directions for the design of computer systems and their integration into human activity.

SHRDLU is head and shoulders above contemporary systems when it comes to intelligent conversation. Although its domain of discourse is restricted to a tabletop world of colored objects SHRDLU really understands this world in terms of the relation between semantics and the physical properties of the blocks and the tabletop. It consists of subsystems that parse interpret and construct sentences, carry out dictionary searches and semantic analyses and makes logical deductions.

SHRDLU uses Halliday's systemic grammar, which emphasizes the limited and highly structured choices made in producing syntactic structure abstracting the features that are important for conveying meaning. The parser is special and interprets the recognition grammars. Meaning is covered by the development of a formalism for concepts within a language user's model of the world representing objects events and relationships. Semantics is represented by a system, which is developed to work in conjunction with the parser, a dictionary and the problem solving programs. It considers not only meaning, but also context.

III. Purpose of Interview

The present interview is one of a number of interviews with people who are or have been involved in AI research. The purpose of these interviews is to learn about their views of the AI-field and the work of other AI scientists.

As it appears from above Prof. Winograd was heavily involved in AI research earlier in his scientific career, but he decided to leave the field of AI. Elsewhere he describes his conversations with Hubert L. Dreyfus as influential to his own “complete shift of research direction, away from artificial intelligence towards a phenomenologically informed perspective on human-computer interaction” (Winograd, 2000). Using this as the point of departure it was assumed that he would provide interesting answers to questions such:

- What attracted him to AI in the first place?
- Being a "second generation" or may be even "third generation" AI-scientist, how does he think that his view on AI differs from the view held by "first generation" AI-scientists?
- What is he view on the work by other AI scientists?
- Why did he choose to leave the field of AI?

And thereby, contribute to our understanding of the history of AI.

V. References

Winograd, T. (2000) Foreword. In: M. Wrathall and J. Malpas (eds.) Heidegger, Coping and Cognitive Science – Essays in the honor of Hubert L. Dreyfus – volume 2. Cambridge, MA: MIT Press.
(Available at: http://www.idiom.com/~gdreyfus/70Celebration/Foreword2.html)

5

VI. The Interview[3]

The following conversation took place between Professor Terry A. Winograd (TW) and Morten Thanning Vendelø (MV) on February 29. 2000, in the Department of Computer Science, Stanford University, California.

MV: Could you tell a little about your educational background. I know that you got a BA in Mathematics from The Colorado College, then went to University College in London and studied linguistics there, and then went to MIT for your Ph.D. but is there a storyline in choosing that path for your education?

TW: The storyline is that I went to a liberal arts college that had no engineering and very little mathematics. So my mathematics major didn’t mean that I did a lot of mathematics. I was interested in language, and they didn’t have a linguistics course, but I convinced the anthropology teacher to give me a readings course in linguistics. Also, I became acquainted with computers, not through coursework but outside class during my last couple of years in college. In addition, I saw some writings by Marvin Minsky describing artificial intelligence and thought it was pretty fascinating. So I applied to MIT because of these computer interests, and I also applied for a Fulbright Grant to study linguistics abroad. That was basically opportunistic. People said: “It is wonderful experience to go to Europe for a year on a Fulbright, you should do it.” So the driving force was going to Europe, not the specific content. At that time my belief (which I now realize ironically was false) was that there was no interesting computing work going on outside the United States. If I had known what was going on in computing in England I might have done something different. The only language I had studied in college was Russian, and at that time going to Russia on a Fulbright Grant was not an opportunity. So I was limited to English speaking countries and since I was interested in language I decided to do a year of study in linguistics. So I simply went through the catalogues of the British universities looking for what seemed to be the most interesting linguistics program from the point of view of what I wanted to do. I found the program at University College, London, with a professor Halliday and I applied to it although I did not have a very coherent intellectual plan. I came from Colorado, a small state, and since Fulbright had a quota by state I had an advantage in getting the grant. So that's how I got the Fulbright and studied linguistics for a year in London. If I had not got the grant, I would have gone directly to MIT and not done the linguistics work. Because once you were at MIT, you couldn't do computer science and linguistics at the same time.

When I would meet Chomsky's students at a party and say: "I work in the AI Lab." they would turn around and walk away. That was it. There was no communication, but active hostility between Minsky and Chomsky. So when I came back from the year in London to start at MIT I already had a lot of interest in language and a year of background in linguistics. That is how I ended up doing a language project.

MV: You said that you became interested in computers. What was it about computers that made you interested in them and in AI?

TW: When I was a junior in college a professor in medicine who had been at a larger medical center ended up moving to this small hospital in Colorado Springs. He had previously hired programmers to work on his research. He had a very early computer, which he used for doing calculations on radiation therapy. So he sent a note over to the math department saying: "Do you have any math students that can help me with my computer?" So my first exposure to computers was a personal computer. There was a room and there was this desk-sized thing (a CDC 160) and I was the only person using the computer. I had a great time so that was how I got into computing.

MV: But how did you become interested in AI?

TW: The question was, which field was combining computers and language, because they were the two things I was interested in.

MV: You said that at MIT there were no connections between computer science and linguistics?

TW: Minsky and Chomsky were both very strong personalities. You could go to one or the other but you couldn't be in the middle. That is how MIT is, it has strong boundaries between its departments.

MV: But you were allowed to do computers and linguistics?

TW: Minsky wanted me to do linguistics and prove that his students could do better linguistics than Chomsky's. I was not the first. Dave Waltz, who eventually went on to do work in vision, was doing a research project on linguistics when I got there, as was Gene Charniak. And in the prior generation Daniel Bobrow and Bertram Raphael had both done language projects. So it was a major stream of work in the AI Lab.

MV: What was it that made you make this connection between AI and language, why did you choose this connection between computers and language?

TW: I had the question: "How could you make computers use language?" And by definition the answer is AI, because computers don't do language by themselves. Fortran programs didn't do language, except for Joe Weizenbaum's program, but that was not a standard Fortran program. My direction wasn't because of any philosophical commitment to artificial intelligence. It was because of an interest in understanding how language worked, using computers as the tool.

MV: How would you describe the artificial intelligence environment at MIT at that point in time? If you talk to some people who went to Carnegie Tech in the late 1950s and 1960s they describe it as an intellectual supernova?

TW: It was a very exciting time because people were getting their hands on machines that previously either were available only for serious and highly specialized scientific work. We could try things out that nobody had ever tried before. So people came up with all sorts of interesting things, some of which turned out to be very important and some of which didn't. The excitement is like being in a gold rush or exploring new territory. There are all these nuggets lying on the ground and you can pick them up to see what you can do with them. Also, there was a very strong sense of that we were building the future.

I would not call it an intellectual supernova in that there was no emphasis on deep intellectual thought. The quest was to build stuff and see what it did, and then build more stuff and make new things happen. But it wasn't intellectually grounded. Carnegie was much more grounded in Simon's and Newell's theoretical interests and AI

[3] This transcription of the interview has been with Terry Winograd for review.

was driven by their theory. Whereas at MIT it was more a hacking approach if you can call it that. Those people highly respected in the lab were not cognitive researchers, but virtuoso programmers -- hackers.

MV: So this linking at Carnegie between economics, AI and cognitive science did not exist at all at MIT?

TW: No, it was looked down upon as a kind of waste of effort. "We are building wonderful new stuff, why worry about how people think about economics?"

MV: You completed your Ph.D. at MIT and then you went to Stanford. How would you describe the environment at Stanford compared to MIT?

TW: They were pretty similar. John McCarthy had developed AI Lab here. It was a bit more independent because it had its own building up in the hills -- it wasn't even sharing a building with the rest of computer science. It was funded in the same way as the MIT Lab, which is by large umbrella programs. This meant that individual projects had a lot of latitude to do whatever they felt like and research didn't have to be justified on a project-by-project basis. There were a lot of machines and a lot of work on robotics. As for its emphasis, it was pretty much a sister environment to MIT, and thereby distinct from Carnegie, with its more cognitively grounded environment. Stanford put more emphasis on using formal logic because John McCarthy likes formal logic and Marvin Minsky didn't.

MV: Did you have any contact with Ed Feigenbaum about AI when you arrived here?

TW: It depends on what you mean by contact. He was not in the AI lab. I was up on the hill with McCarthy's lab and Feigenbaum was somewhere on campus with his expert system projects. He was there when they hired me, and he was on the faulty committee and so on, but in terms of day-to-day contact we had very little.

MV: Was it because you had these two different views on AI?

TW: Again it was two different camps. There was Feigenbaum students and McCarthy students, and Feigenbaum faculty and McCarthy faculty. And you really didn't cross those boundaries. It wasn't as hostile as between Minsky and Chomsky, but it was still a very clear divide.

MV: If we look at these two different ways of doing AI, and were to do an evaluation of them today, How would you evaluate them in terms of their progress / contribution to AI?

TW: It is good question. I think that Feigenbaum was much more eager and willing to say: "I want practical and commercial applications and I don't care how deep or interesting the theory is." If you look at his so-called theoretical principles, they are very shallow. The focus was on practical use. If you look at the fifth generation work, it is fairly mundane from an intellectual point of view, but they actually tried to make it commercially relevant. On the other hand, McCarthy is really a pure mathematician. His interests have nothing to do with making money or applying AI. They have to do with coming up with deeper theories. I happen not to agree with McCarthy's theoretical leanings, but I think that it is good that he had them and was trying to drive the program from a conceptual point of view, instead of a kind of opportunistic way.

MV: You said that you didn't interact with the linguists at MIT. But did the work by, for instance, Noam Chomsky inspire you?

TW: It did indirectly, but my direct source of inspiration came from the year in London where I studied with Halliday. If you look at what Halliday was doing in the context of the larger picture of generative grammar, he was trying to adapt his theories, which came more from a social perspective, to a generative form. In my opinion, the merger never worked. If you look at the subsequent work of Halliday in linguistics, the attempts to do generative grammar have been much less successful than the return to systemic grammar's roots as a socio-linguistic analysis. But certainly nevertheless the training was important, I took a course in transformational grammar during that year, so I had learned it. Chomsky's basic insight was that synchronistic language can be described fairly well with a generative rule-based system. That was at the heart not only of his work but also of the work by anybody who was trying to use computers, because if you cannot put language into a rule-based system, then you cannot program it. So in that sense I think it was very much along the lines of Chomsky. The differences are at the next level where Chomsky posited a structure of transformations, which was not computationally implemented while those in AI were focused on finding appropriate algorithms. So I was on the algorithm side and not the formal grammar side.

MV: Returning to your time at MIT. Would you say that your view of AI is similar to that of Marvin Minsky's? How would you position yourself in relation to him?

TW: We differ in a couple of major ways. One is his basic faith that the intelligence embodied in organisms, for instance people, is very similar in nature to the programs we write in AI. Not necessarily in detail but in basic nature. I disagree with that fundamental assumption about symbolic processing, which I critiqued in the book I wrote with Fernando Flores. The other key differences are Minsky's view on the role of human values and the role of machines. If you ask: "What is Minsky's religion?" Then it is some abstract notion of the progress of intelligence. The values he was pursuing had to do with a drive towards higher intelligence as a value in itself. I have always been a much more political, social oriented kind of person for whom other human values take priority.

MV: In my first e-mail I mentioned that you might be what we can call a second or may be even a third generation AI scientist. How would describe yourself in contrast to the first or may be the second generation?

TW: People like Minsky and Newell and so on were the immigrants. None of them started their intellectual life in artificial intelligence, as it didn't exist. So they went in and made it happen. Then the second generation, which I would include myself in, followed in their labs. At MIT there were Tom Evans, Daniel Bobrow, Bertram Raphael and Adolfo Guzman and so on, and then in a later round there were Gerald Sussman, Carl Hewitt, Dave Waltz, me, Eugene Charniak, Pat Winston, and others. But there wasn't any major transition between those two. I think it was more gradually increasing machinery and more sophistication, but it was pretty much the same spirit. We did not have to develop the context. The context was there, the machines were there, LISP was there and so on. So we could just take a problem and apply those things.

MV: And then your own idea for your block-moving program SHRDLU, how did it come about?

TW: Basically the block-moving problem was chosen because I wanted to do something more concrete. Bill Woods at Harvard was using airline reservation systems as a language domain, but I didn't find them very interesting. Minsky had the idea that I should do something about stories for children, because he thought they were much more simple than stories for adults. I was not so sure about that and wanted to do something else. There were others at the lab building robot hand-eye systems that actually moved blocks around on a tabletop. So I don't know if it was me or Minsky, but we arrived with the idea of doing a system that conversed about block moving. But there was never any real robot arm connected to my work.

MV: It was common for AI scientists from Carnegie Tech to go to RAND and many people perceive RAND as very important to the development of AI. Did you ever go to RAND?

TW: No and I cannot remember anyone from RAND coming to MIT, and I cannot remember that anyone from MIT went to RAND. I remember it as a pure Carnegie connection.

MV: In the middle of the 1980s you wrote and published a book with Fernando Flores, where you articulate a more critical view of AI. How did you meet Fernando Flores?

TW: I went to a meeting where I met a Chilean scientist, Francisco Varela, who said: "How is Fernando Flores?" I said: "Fernando who?" And he said: "Fernando Flores. He is at Stanford." And I said: "Oh well Stanford is a big place and I don't know where he is." And he said: "No, he is in the Computer Science Department at Stanford," and I said: "No, this cannot be, I am in the Computer Science Department I go to all the faculty meetings and I have never heard of Fernando Flores." He said: "I know Fernando Flores is in the Computer Science Department." So when I got back I looked him up, and it turned out that he was indeed there. After being in the government of Salvador Allende that was overthrown by a coup, Fernando had been imprisoned for three years. The San Francisco Chapter of Amnesty International took up him as a "prisoner of conscience," and one of the conditions for his release that the Chilean government put up was that he had to have a job waiting for him outside of the country. Two professors from our department, Bob Floyd and George Dantzig, had a position they could support from various grants that were close enough to what he had done before going into government that they could justify hiring him. So without having a specific demand on his work they created a research associate position for him at Stanford so that he could have a job, so that they could get him out of prison. But he had just been in prison in Chile for three years and had been flown to California, where he was getting oriented and reconnecting with his family, so he was not spending his time in the Computer Science Department.

I looked him up, and he is a tremendously intellectual guy, probably the most intellectual, in some deep sense, that I have ever known. He is also practical, but he is the kind of thinker who can read six philosophy books before breakfast. He has this incredible mind always thinking and always looking for more, so he wanted to find out who at the department were interested in talking to him. We started talking in a casual way, then he handed me a book on philosophy of science and said "You should read this." I read it, and we started talking about it, and we decided to write a paper about it that turned into a monograph that turned into a book. It was a gradual process of finding him interesting, and finding the stuff we were talking about intellectually stimulating. He only kept his job at Stanford for a short time, and then went to Berkeley as a Ph.D.-student, because he wanted to have a Ph.D. In Chile he had gone into politics before he had finished his Ph.D. I was officially on his committee there, but it was never a question who was leading and who was following.

MV: You are also one of the founders of the organization called: Computer Professionals for Social Responsibility. Did it have any connection with your decision to leave the field of AI?

TW: No, the two issues are really separate. My interest in CPSR grew out of nuclear war issues. In those years any organization that had social responsibility in its name was basically trying to prevent nuclear holocaust. The American government seemed to be headed down this track, and every group said: How can we help? First there was Physicians for Social Responsibility, then Educators for Social Responsibility, Architects for Social Responsibility, Psychologists for Social Responsibility, and so on. It was clear that there were a lot of computing issues involved in warfare and when the Star Wars program (SDI) came along it was very clear that it was based on assumptions about computing that were not valid. We picked up on this as the key issue. The only place where it touched on the areas of AI was to the degree that military funding and military applications were based on practical applications of AI. But as to whether the theory of AI in general was right or wrong, a good idea or a bad idea, was just never in the purview of that kind of politically oriented activity. CPSR was not philosophically oriented but politically oriented.

MV: And then after AI you moved on to HCI, what let you in that direction?

TW: I think that I was always doing it. I just didn't realize it. When I was writing systems to use language there were two somewhat independent motivations. One is to model and understand human language, and the other is to make computers easier to use. What I realized was that making computers easier to use, was not

the same as making them use ordinary language. There are many other issues involved in how to get people and computers to interact well. For a lot of the cognitive studies, you can turn them around. Instead of thinking of how to model people with the computer you use cognitive analyses of people, to better fit the computer to them. – how to interface with them, rather than how to duplicate them. Much of what we have learned in AI is very relevant to HCI.

There was a period of about five years where I had no label for my work. Because I was out of AI and I was doing a sort of philosophical writing, but I wasn't a philosopher. I realized that really HCI was more compatible if I wanted to be in the university environment. The kind of cognitive-philosophical reflection that I was doing with Flores was not a viable program for graduate students in a Computer Science Department. There needed to be something much more concrete that the concepts could be applied to. If you read the book with Flores you see a kind of open promissory note at the end, which says all of this theory should guide you in the design of systems. So that is really how I got off into Human Computer Interaction -- asking how to make some of these more philosophical considerations guide us in the design of systems that work with people.

MV: Another issue related to this is that a lot of people tend to view AI as the ultimate science in the sense that the big question we have here is: What is the nature of human intelligence? And therefore, it is one of the biggest questions that you can work on as a scientist, and most else come second. Is this an opinion that you share?

TW: I think that my sense of it, and this is in a broad sense my opinion for scientific research, is that for a scientific topic to be good to work on there have to be at least three things that come together. One is resources -- funding. One is that it is an interesting problem, and the third is that you are at the right point in intellectual and technical history to have some leverage on the problem. Scientists a hundred or two hundred years ago could have thought that it was a fascinating problem to understand how the brain works, and they might have built models of it. But they wouldn't have gotten very far because they didn't have the background knowledge. My sense of AI is that we are not there yet. If I were a young person going into science today, I might very well go into biology or neuroscience, because I think there are all sorts of things that we yet don't know about real brains and nervous systems. We need to develop better fundamental principles of how information systems and biologic information systems work, without trying to pretend that they are what we have in silicon, just sloppier. After another 10, 20 or 50 years of great research in that area it will be the moment when somebody can say: "Ah, we can put this together." My own hunch about where we are is that we are still missing the basic knowledge that will make AI a productive science.

MV: So you don't share these more optimistic views on AI?

TW: Well, what I just said is that it will happen. I just think that we are much further away than many AI proponents think we are. I certainly don't believe that the basic science that we need is there, and that it is just a matter of putting together the mechanisms. It is going to take quite some time before we get there, and it is going to come out of the biologic science and not out of more computer development at this point.

MV: So more work is needed, but in different areas than many AI-scientists believe. Do you share the reductionistic view of AI that some AI-scientists pursue?

TW: There are two different levels. I share the physicalist view. If you say: "Is there anything going on in my mind which cannot be explained by the motion of molecules in my brain?" I don't believe it. I believe that thinking is purely a mechanical process. If by reductionistic you mean, when I am thinking about a dog, will that be correspond to activating some kind of computer-like symbol for the concept of dog, then I don't believe it. So it is a physical reductionism but not a symbolic reductionism, if you want to distinguish among those two.

MV: Given that the needed biological knowledge is in place, how far do you think we can go with artificial intelligence?

TW: Starting from scratch and building up a mind will not be the way it goes. It will be more likely to incorporate hybrids that combine real biological systems with some amount of artificial stuff. We will be able to engineer changes to real biological systems genetically or chemically. So the question: "Could you build intelligence?" will not be the interesting question. It will be more a question of extending and adapting intelligence. I don't think that there are ultimate limits in the sense that we can never achieve a certain kind of intelligence. I don't know where they are, and I have no particular reason to believe that they are not hundred or thousand years away. But in principle the brain is a mechanical system, just a very complex one.

MV: If you look 20 years or so ahead, do you then have any hopes for artificial intelligence. In terms of what would be a good outcome of the work being done in the field?

TW: I think that what will happen in twenty years is a continuation of what we see now. Consider speech understanding. When I was a graduate student there were heated arguments as to whether a computer needed full logical understanding to understand speech. It is easy to come up with words that sound alike and whole phrases that sound alike but mean different things, and you have to understand this. What has happened is, that as computers were able to process more and more data and compute higher level statistics and more analysis, we have got programs for speech dictation that achieve accuracies up to 90% without any logical understanding. You have programs listening to you, which don't have any logical understanding of what you are saying and they do it pretty well. I think that we are going to see those boundaries pushed more in various applications such as visual recognition. It is not necessarily going to be because of new insights into how people do these tasks, which would be scientifically interesting. Instead it is going to be because you can just throw enough processing power at it to do an acceptable job. So I think that a lot of things which people in the early days though of as proofs of deep AI, including chess as an the obvious example, will just be handled by not deep AI but by a lot of processing power. We have almost seen success at chess that way. Computers haven't quite been beating Kasparov consistently yet, but they are doing well. AI success will be in a number of applications, none of which will have the flavor that the original AI people though about, which is: "This machine is thinking like a person." Instead, it will use extensive computation and incorporate statistical methods. When I was a student in the AI Lab nobody there even studied statistics, nobody even mentioned statistics. Today non-symbolic techniques are critical in most branches of AI. As for any major imminent breakthroughs in AI, I don't see them, but usually you don't see a breakthrough before it happens.

MV: OK, this was my last question, so thank you for your time.

VII. Selected works by Terry A. Winograd

VII.i Books

Winograd, T. with J. Bennett, L. De Young and B. Hartfield (eds.) (1996) Bringing Design to Software. Menlo Park, CA: Addison-Wesley

Adler, P., and Winograd, T. (eds.) (1992) Usability: Turning Technologies into Tools. New York, NY: Oxford University Press.

Friedman, B., and Winograd, T. (eds.) (1989) Computing and Social Responsibility: A Collection of Course Syllabi. Palo Alto, CA: Computer Professionals for Social Responsibility.

Winograd, T., and Flores, F. (1986) Understanding Computers and Cognition – A New Foundation for Design. Menlo Park, CA: Addison-Wesley.

Winograd, T. (1972) Understanding Natural Language. New York, NY: Academic Press.

VII.ii Journal Articles

Oviatt, S. L., Cohen, P. R., Wu, L.,Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., and Ferro, D. (2000) Designing the User Interface for Multimodal Speech and Gesture Applications: State-of-the-Art Systems and Research Directions. Human Computer Interaction, vol. 15, no. 4, pp. 263-322.

Winograd, T. (1995) Environments for Software Design. Communications of the ACM, vol. 38, no. 6, pp. 65-74.

Winograd, T. (1994) Designing the Designer. Human-Computer Interaction, vol. 9, no. 1, pp. 128-132.

Flores, F., Graves, M., Hartfield, B., and Winograd, T. (1988) Computer Systems and the Design of Organizational Interaction. ACM Transactions on Office Information Systems, vol. 6, no. 2, pp. 153-172.

Winograd, T., and Flores, F. (1987) Response to Reviews of Understanding Computers and Cognition. Artificial Intelligence, vol. 31, pp. 250-261.

Winograd, T. (1980) Extended Inference Modes in Reasoning by Computer Systems. Artificial Intelligence, vol. 13, no. 1, pp. 5- 26.

Winograd, T. (1980) What Does It Mean to Understand Language? Cognitive Science, vol. 4, no. 3, pp. 209-242.

VII.iii Book Chapters, Reviews, and Popular Articles

Winograd, T. (1995) Heidegger and the Design of Computer Systems. In: A. Feenberg and A. Hannay (eds.) Technology and the Politics of Knowledge. Bloomington, IN: Indiana University Press, pp. 108-127.

Winograd, T. (1994) The Norbert Wiener Award for Social and Professional Responsibility. Cybernetica, vol. 37, no. 3/4, pp. 387-392.

Winograd, T. (1991) Thinking Machines: Can There Be? Are We? In: J. Sheehan and M. Sosna (eds.) The Boundaries of Humanity: Humans, Animals, Machines. Berkeley, CA: University of California Press, pp. 198-223.

Davis, R. (ed.), Dreyfus, S., and Winograd, T. (1989) Expert Systems: How Far Can They Go? AI Magazine, Spring, pp. 61-67.

Winograd, T. (1984) Some Thoughts on Military Funding. CPSR Newsletter, vol. 2, no. 2, pp. 1-3.

Winograd, T. (1976) Artificial Intelligence and Language Comprehension. Lead article in a report of the same title by the National Institute of Education, February, pp. 1-26.

Winograd, T. (1976) Computer Memories--A Metaphor for Human Memory. In: C. Cofer (ed.) The Structure of Human Memory. Freeman, pp. 133-161.

Winograd, T. (1974) Artificial Intelligence - When Will Computers Understand People? Psychology Today, vol. 7, no. 12, pp. 73-79.

Winograd, T. (1973) A Process Model of Language Understanding. In: Schank and Colby (eds.) Computer Models of Thought and Language. Freeman, pp. 152-186.

Winograd, T. (1973) Language and the Nature of Intelligence. In: G. J. Dalenoort (ed.) Process Models for Psychology. Rotterdam, NL: Rotterdam University Press, pp. 249-285.

AIS SIG on Semantic Web and Information Systems,
AIS SIG SEMIS Bulletin, 3(1) 2006

Special Issue on Semantic Web for Life Sciences
(especially for PhD students)

http://www.sigsemis.org

CALL FOR SHORT ARTICLES

Deadline for submission: Feb 15th, 2006.

Regural CFP, CFP for the special issue will be announced shortly

We invite submissions that are related (but not limited to) to the following topics:
Semantic Web Services
Intelligent Systems
Semantic E-Business
Semantic KM
Semantic E-learning
Semantic E-Government
Semantic Web & Business Intelligence
Semantic Web & Enterprise Application Integration
Semantic Web languages
Ontologies
Agents
Semantic Information Processing
Semantic Web & Multimedia
Semantic Web Standards
Open Source
Peer to Peer
Semantic Web and Mobile & wireless Technologies
Semantic Web and IS Research & Methodology
SW Curricula
Knowledge Society
SW Industry
SW and Culture

Submission procedure:

1. The articles in the bulletin can be from 1000 to 3500 words.
2. The manuscripts should be either in Word or RTF format.
3. Please send the manuscripts by email as attachment to Dr. Miltiadis D. Lytras lytras@ceid.upatras.gr ,

- Discussion papers.

We invite submission of RIP papers that will be hosted also in **SIGSEMIS**.org and a discussion will be organized in order to strengthen the authors' approach. Only two papers max will be selected in each bulletin and must be from PhD students.

JOIN AIS SIGSEMIS

You can support the activities of the SIG by joining as an AIS SIG member.
This can be done in one of the following ways:

- ➔ For AIS members: Go to the AIS SIG page (http://www.aisnet.org/sigs.shtml) and join SIGSEMIS after log in with your AIS username and password (10$ membership fees)
- ➔ For non-AIS members: Go to the AIS membership page (http://www.aisnet.org/join.shtml) and select the SIGSEMIS option while joining.

Join as supporters

You can state your support to AIS SIGSEMIS activities by sending through email to Miltiadis Lytras mdl@eltrun.gr a short bio in order to update a relevant section in our portal.

OPEN RESEARCH SOCIETY

http://www.open-research-society.org

The **Open Research Society (ORS)** is a non-profit international organization dedicated to the promotion of the "open research" paradigm in scholarly publishing, research cooperation and dissemination of research results in the field of **Information Technology** in a broad sense.

The initial effort of the society will be that of contributing to the **open access movement** in scholarly communication with a community-based approach to scientific journals and research monographs.

ORS is open to any individual that is interested in the activities of the society. **Membership in the Society is free**. Fees will be only required to support specific open research initiatives, in which participation will take place by volunteering.

Any individual or institution interested in the concept of the ORS can subscribe the *Yahoo Group* to keep informed on the projects of the ORS: http://groups.yahoo.com/group/open-research-society/

The following are the two journals planned for 2006 in an initial inception phase.

THE OPEN RESEARCH JOURNAL

A journal devoted to the economic, social and ethical aspects of the paradigm of "Open Research"

Open Research ON INFORMATION SYSTEMS

A peer-reviewed journal on any aspects of Information Systems research

The second phase will be planned early 2006. Information about opportunities to collaborate will be provided to members of the ORS.

Free Registration: Use this form: http://www.open-research-society.org/preregform.doc

We can make it TOGETHER