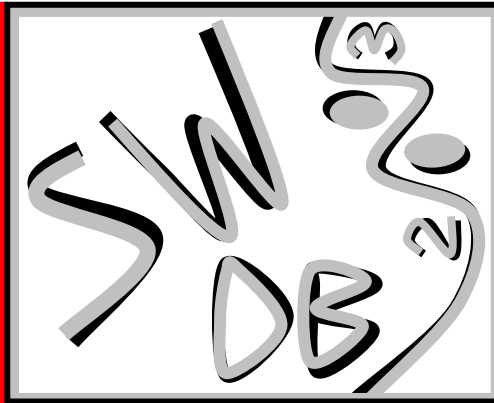**Very
Large
Data
Bases**

Abstracts of

SWDB' 03

**The first International Workshop on
Semantic Web and Databases**

Co-located with VLDB 2003
Humboldt-Universität
Berlin, Germany
September 7-8, 2003

We appreciate the contributions from our sponsors:

OntoWeb Network

Organizers

Program Committee Chairs
Isabel F. Cruz
U. Illinois at Chicago, USA
(ifc@cs.uic.edu)

Vipul Kashyap
National Library of Medicine, NIH, USA
kashyap@nlm.nih.gov)

Proceedings and Publicity Chair
Stefan Decker
USC Information Sciences Institute, USA
stefan@isi.edu

Organization Chair
Rainer Eckstein
Humboldt University, Germany
Rainer. Eckstein@informatik.hu-berlin.de

PC Members

Karl Aberer, EPFL, Switzerland
Sibel Adali, Rensselaer Polytechnic I., USA
Paolo Atzeni, U. Rome Tre, Italy
Alex Borgida, Rutgers U., USA
Olivier Bodenreider, NLM-NIH, USA
Stéphane Bressan, National U. of Singapore
Christoph Bussler, Oracle, USA
Isabel Cruz, U. of Illinois at Chicago, USA
Umesh Dayal, HP Labs, USA
Stefan Decker, USC-ISI, USA
Max Egenhofer, U. Maine, USA
Rainer Eckstein, Humboldt U., Germany
Dieter Fensel, Institut für Informatik, Austria
Mary Fernandez, AT&T Labs - Research, USA
Susan Gauch, U. Kansas
Carole Goble, U. Manchester, UK

Rick Hull, Lucent Technology, USA
Vipul Kashyap, NLM-NIH, USA
Maurizio Lenzerini, U. Rome "La Sapienza", Italy
Ling Liu, Georgia Tech, USA
Robert Meersman, Vrije U., Belgium
John Mylopoulos, U. Toronto, Canada
Aris Ouksel, U. Illinois at Chicago, USA
Dimitris Plexousakis, U. Crete, Greece
Steve Ray, NIST, USA
Amit Sheth, U. Georgia and Semagix, USA
Surya Sripada, Boeing, USA
Munindar Singh, N. Carolina U., USA
V.S. Subrahmanian, U. Maryland, USA
Rudi Studer, U. Karlsruhe, Germany
Ram Sriram, NIST, USA
Clement Yu, U. Illinois at Chicago, USA

Semantic Web and Databases

September 7, 2003 (Sunday)	September 8, 2003 (Monday)
8:45-9:00 Welcome	**9:00-10:10 Keynote Talk**
9:00-10:10 Keynote Talk <u>**Can we do better than Google? Using semantics to explore large heterogeneous knowledge sources**</u> ***Anatole Gershman, Accenture Technology Labs***	<u>**From Semantic Search to Analytics and Discovery on Heterogeneous Content: Changing Focus from Documents and Entities to Relationships**</u> ***Amit Sheth, University of Georgia and Semagix, Inc.***
10:10-10:40 Semantic Web at Work Spatially Navigating the Semantic Web for User Adapted Presentations of Cultural Heritage Information in Mobile Environments *Marco Neumann, Dublin Institute of Technology, Ireland.* Text-Based Gene Profiling with Domain-Specific Views. Patrick Glenisson, Bert Coessens, Steven Van Vooren, Yves *Moreau and Bart De Moor, Katholieke Universiteit Leuven, Belgium.*	**10:10-10:40 Web Services** ODE-SWS: A Semantic Web Service Development Environment *Oscar Corcho, Asunción Gómez-Pérez, Mariano Fernández-López, and Manuel Lama, Universidad Politécnica de Madrid, Spain, and Universidad de Santiago de Compostela, Spain.* Applications of PSL to Semantic Web Services *Michael Gruninger, University of Maryland, College Park, USA.*
10:40-11:10 Coffee Break	**10:40-11:10 Coffee Break**
11:10-12:30 Context-Aware Systems Context-Aware Semantic Association Ranking *Boanerges Aleman-Meza, Chris Halaschek, I. Budak Arpinar, and Amit Sheth, University of Georgia, USA.* I know what you mean: semantic issues in Internet-scale publish/subscribe systems *Ioana Burcea, Milenko Petrovic, and Hans-Arno Jacobsen, University of Toronto, Canada.* A Context-Oriented RDF Database *Mohammad-Reza Tazari, Computer Graphics Center, Dept. Mobile Information Visualization, Darmstadt, Germany.* An Adaptable Service Connector Model: *Gang Li, Yanbo Han, Zhuofeng Zhao, Jianwu Wang, Roland Wagner, Chinese Academy of Science, PRC, Fraunhofer, Germany*	**11:10-12:30 Data Mining and Peer-to-Peer Systems** H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems. *S. Castano, A. Ferrara, S. Montanelli, Università degli Studi di Milano, Italy.* A Collaborative Approach for Query Propagation in Peer-to-Peer Systems *Anne Doucet, Nicolas Lumineau, University of Paris 6, France.* OntoMiner: Bootstrapping and Populating Ontologies from Domain Specific Web Sites *Hasan Davulcu, Srinivas Vadrevu, and Saravanakumar Nagarajan, Arizona State University, USA.* Can Data Mining Techniques Ease The Semantic Tagging Burden? *Fabio Forno, Laura Farinetti1, Sean Mehan, Politecnico di Torino, Italy, University of the Highlands and Islands, UK.*
12:30-2:00 Lunch (on your own)	**12:30-2:00 Lunch (on your own)**
2:00-3:10 Keynote Talk <u>**Generic Model Management: A Database Infrastructure for Schema Manipulation**</u> **Phil Bernstein, Microsoft Research, USA**	**2:00-3:30 Formal Querying and Reasoning** Formal aspects of querying RDF databases *Claudio Gutierrez, Carlos Hurtado, and Alberto Mendelzon, Universidad de Chile, Chile, and University of Toronto, Canada.*
3:10-3:40 Modeling Issues Building an integrated Ontology within SEWASIE system, *D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini, Università di Modena e Reggio Emilia, Italy and IEIIT-CNR, Italy.* Ontologies : A contribution to the DL/DB debate *Nadine Cullot, Christine Parent, Stefano Spaccapietra, and Christelle Vangenot, University of Burgundy, France, Swiss Federal Institute of Technology, Lausanne, Switzerland, University of Lausanne, Switzerland.*	Event-Condition-Action Rule Languages for the Semantic Web. *George Papamarkos, Alexandra Poulovassilis, Peter T. Wood, Birkbeck College, UK.* Storing and Querying Ontologies in Logic Databases. *Timo Weithoener, Thorsten Liebig, and Guenther Specht, University of Ulm, Germany.* Design Repositories for the Semantic Web with Description-Logic Enabled Services. *Joseph B. Kopena and William C. Regli, Drexel University, USA.* Mediation of XML Data through Entity Relationship Models. *Irini Fundulaki and Maarten Marx, Bell Laboratories, USA, and University of Amsterdam, The Netherlands.*
3:40-4:10 Coffee Break	**3:30-4:00 Coffee Break**
4:10-5:30 RDF Storage and Implementation Issues Efficient RDF Storage and Retrieval in Jena2 *Kevin Wilkinson, Craig Sayers, and Harumi Kuno, HP Labs, USA.* An Indexing Scheme for RDF and RDF Schema based on Suffix Arrays. *Akiyoshi Matono, Toshiyuki Amagasa, Masatoshi Yoshikawa, and Shunsuke Uemura, Nara Institute of Science and Technology, Japan, and Nagoya University, Japan.* RDF Core: A component for effective management of RDF Models. *Floriana Esposito, Luigi Iannone, Ignazio Palmisano, and Giovanni Semeraro, Università degli Studi di Bari, Italy.* Implementation of a Semantic Network Service (SNS) in the context of the German Environmental Information Network (gein®) *Thomas Bandholtz, Germany.*	**4:00-5:20 Integration and Interaction)** The ICS-FORTH SWIM: A Powerful Semantic Web Integration Middleware *V. Christophides, G. Karvounarakis, I. Koffina, G. Kokkinidis, A. Magkanaraki, D. Plexousakis, G. Serfiotis,and V. Tannen, University of Pennsylvania, USA, and Institute of Computer Science, FORTH, Greece.* Semantic Representation of Contract Knowledge using Multi Tier Ontology *Vandana Kabilan, Paul Johannesson, Stockholm University and Royal Institute of Technology, Sweden.* The Visual Semantic Web: Unifying Human and Machine Semantic Web Representations with Object-Process Methodology *Dov Dori, Technion, Israel and MIT, USA.* Interaction and navigation for a document database: a concrete case study *Isabelle Berrien, François Laburthe, and Jean-David Ruvini, e-lab BOUYGUES SA, France.*
	5:20-5:30 Closing

Table of Contents

Foreword

The Semantic Web is a key initiative being promoted by the World Wide Web Consortium (W3C) as the next generation of the current web. Machine-understandable metadata is emerging as a new foundation for component-based approaches to application development. Within the context of reusable distributed components, Web services represent the latest architectural advancement. Such concepts can be synthesized providing powerful new mechanisms for quickly modeling, creating and deploying complex applications that readily adapt to real world need.

The objective of this workshop is to present database and information system research as they relate to the Semantic Web and more broadly, to gain insight into the Semantic Web technology as it relates to databases and information systems.

Isabel F. Cruz
U. Illinois at Chicago
USA

Vipul Kashyap
National Library of
Medicine, NIH, USA

Stefan Decker
USC Information
Sciences Institute, USA

Rainer Eckstein
Humboldt University,
Germany

Invited Talks

Can we do better than Google? Using semantics to explore large heterogeneous knowledge sources

Anatole Gershman
Accenture Technology Labs
USA

Abstract

Researchers in many fields use dozens of different rapidly growing on-line knowledge sources, each with its own structure and access methods. Successful research often depends on a researcher's ability to discover connections among many different sources of information. The popularity of Google suggests that high-quality indexing would provide a uniform method of access, although it still leaves researchers with vast, undifferentiated lists of results. Hence, the research challenge for semantic web designers: can a knowledge-based approach provide a better way for researchers to explore knowledge and discover useful insights for their research?

In this talk, I will use the example of bio-medical knowledge discovery to explore the key issues in semantic indexing of large amounts of heterogeneous information. I will propose a method and architecture for the creation of practical tools for semantic indexing and exploration.

The example I'll be using is the Knowledge Discovery Tool, or KDT, which contains a knowledge model of a large number of bio-medical concepts and their relationships: from genes, proteins, biological targets and diseases to articles, researchers and research organizations. Based on this model, the KDT index identifies over 2.5 million bio-medical entities with two billion relationships among those entities spanning 15 different knowledge sources. Clearly, the creation and maintenance of such an index cannot be done manually. KDT utilizes an extensive set of rules that cleanse, analyze and integrate data to create a uniform index.

Using its index, KDT presents the user with a uniform graphical browsing space integrating all underlying knowledge sources. This space is "warped" and filtered based on domain-specific rules customized for the needs of various groups of users, such as pharmaceutical researchers, clinicians, etc. Another customized set of rules discovers and graphically highlights potential indirect relationships among various entities that might be worth exploring (e.g., relationships between genes or between diseases). Finally, the tool enables several modes of collaboration among its users from annotations to activities tracking.

Currently, KDT is undergoing testing in two pilot settings: an early stage of the drug discovery process in a pharmaceutical company and a bio-medial academic research group.

About The Speaker
Anatole Gershman joined Accenture Technology Labs in 1989 and in 1997 became its overall Director of Research. Under his leadership, research at the laboratories is focusing on early identification of potential business opportunities and the design of innovative applications for the home, commerce and work place of the future. These include electronic commerce, high-performance virtual enterprise, knowledge management, and human performance support. To achieve these goals, the laboratories are conducting research in the areas of ubiquitous computing, human-computer interaction, interactive multimedia, information access and visualization, intelligent agents, and simulation and modeling.

Prior to joining Accenture, Anatole spent over 15 years conducting research and building commercial systems based on Artificial Intelligence and Natural Language processing technology. He held R&D positions at Coopers & Lybrand, Cognitive Systems, Inc., Schlumberger, and Bell Laboratories. In 1997, Anatole was named among the top 100 technologists in the Chicago area by Crain's Chicago Business. In 2000, Industry Week named Anatole one of the "R&D stars to watch."

Anatole studied Mathematics and Computer Science at Moscow State Pedagogical University and received his Ph.D. in Computer Science from Yale University in 1979.

Generic Model Management: A Database Infrastructure for Schema Manipulation

Philip A. Bernstein
Microsoft Research
USA

Abstract

Meta data management problems are pervasive in the development and maintenance of semantic web applications. Although solutions to these problems are similar to each other, today they are solved in an application-specific way and usually require much object-at-a-time programming. To make solutions more generic and easier to program, we propose a higher level interface, called Model Management. The main abstractions are models and mappings between models. It treats these abstractions as bulk objects and offers such operators as Match, Merge, Diff, Compose, Extract, and ModelGen. We will present an overview of Model Management and recent results about some of the operators.

About The Speaker

Phil Bernstein is a researcher at Microsoft Corporation. Over the past 25 years, he has been a product architect at Microsoft and at Digital Equipment Corp., a professor at Harvard University and Wang Institute of Graduate Studies, and a VP Software at Sequoia Systems. During that time, he has published over 100 articles on the theory and implementation of database systems, and coauthored three books, the latest of which is "Principles of Transaction Processing for the System Professional" (Morgan Kaufmann, 1997). He holds a B.S. from Cornell University and a Ph.D. from University of Toronto. A summary of his current research on meta data management can be found at http://www.research.microsoft.com/~philbe.

**From Semantic Search to Analytics and Discovery on Heterogeneous Content:
Changing Focus from Documents and Entities to Relationships**

Amit Sheth
LSDIS Lab,
The University of Georgia and Semagix, Inc.
USA

Abstract

Research in search techniques was a critical component of the first generation of the Web, and has gone from academe to mainstream. Research and products supporting Semantic Search also look promising.

A second generation "Semantic Web" is being realized in one form of a scalable ontology-driven information system, where semantic metadata allow software to associate meaning with heterogeneous content. This is enabling a fundamental shift in focus from documents and entities within documents to discovering and reasoning about relationships. And it will transform the hunt for documents that humans can examine or analyze into a more automated content analysis, resulting in actionable information and insights into heterogeneous content. In this talk, we juxtapose the following shifts, to paint the exciting new possibilities:

- From documents and entities to relationships
- From techniques that focus on either unstructured data (text) or structured content to both types and semi-structured data
- From directly analyzing data to ontology based processes of creating high quality metadata and analyzing metadata
- From search and browsing for delivering relevant documents and locating entities within contents to discovering complex relationships and delivering actionable information with insights; from semantic search to analytics and discovery-based semantic applications

This talk will interleave academic research with state-of-the-art commercial uses, including tools and real-world applications and experiences. The critical challenge in dealing with the Web scale of ontologies (with huge description base/assertion set), metadata (very large RDF graphs), and their analysis in discovering relationship will be discussed.

About The Speakers

Amit Sheth is a Professor at the University of Georgia and CTO of Semagix, Inc. He started the LSDIS lab at Georgia in 1994. Earlier he served in R&D groups at Bellcore, Unisys, and Honeywell. He founded his second company, Taalee, in 1999 based on technology developed at the LSDIS lab, and managed it as CEO until June 2001. Following Taalee's acquisition/merger, he currently serves as CTO and a co-founder of Semagix, Inc. His research has led to three significant commercial products, several deployed applications and over 150 publications. More: http://lsdis.cs.uga.ed/~amit

Spatially Navigating the Semantic Web for User Adapted Presentations of Cultural Heritage Information in Mobile Environments

Marco Neumann

Digital Media Centre, Dublin Institute of Technology
Dublin 2, Ireland
`marco.neumann@dit.ie`

Abstract. The integration of local and global information is an essential requirement for future location-based services. The development of two technologies for mobile devices, namely positioning devices like GPS and wireless communication networks, is encouraging the development of new kinds of spatial- and context-aware applications. The CHI project investigates the applicability of these technologies for context-aware mobile computing applications that take advantage of new metadata-standards to enable semantic, user and device adapted services in the field of Tourism and Cultural Heritage management and presentation.

Text-Based Gene Profiling
with Domain-Specific Views

Patrick Glenisson, Bert Coessens, Steven Van Vooren, Yves Moreau, and Bart
De Moor

Departement Elektrotechniek, Katholieke Universiteit Leuven, Kasteelpark Arenberg
10, 3001 Leuven (Heverlee)
{`pgleniss, bcoessen`}`@esat.kuleuven.ac.be`

The current tendency in the life sciences to spawn ever growing amounts
of high-throughput assays has led to the situation were the interpretation of
data and the formulation of hypotheses lag the pace with which information
is produced. Although the first generation of statistical algorithms scrutinizing
single, large-scale data sets found their way into the biological community, the
great challenge to connect their results to the existing knowledge still remains.
Despite the fairly large number of biological databases that is currently available,
we find a lot of relevant information presented in free-text format (such as textual
annotations, scientific abstracts, and full publications). Moreover, many of the
public interfaces do not allow queries with a broader scope than a single biological
entity (gene or protein).

Conversely, the process of successfully gaining insight into complex genetic
mechanisms will increasingly depend on a complementary use of a variety of
resources, including biological databases, ontologies, and specialized literature on
the one hand, and the expert's knowledge on the other. We therefore consider the
knowledge discovery process as cyclic, i.e., requiring several iterations between
various information sources to extract a reliable hypothesis.

We implemented a methodology that covers various public biological re-
sources in a flexible text-mining system named TextGate that is designed to-
wards the analysis and summarization of groups of genes based on text. We
exemplify how structured term- and concept-centric profiles complement each
other by discussing 1) how genes involved in transscriptional up- and downreg-
ulation can be readily distinguished based on their term profiles, and 2) how
a group of genes related to colon cancer can be extended to other genes via
colinkage.

Through the use of domain vocabularies that are composed of standardized
gene or term nomenclatures, we demonstrate that our generated summaries are
suited for further connecting them to a variety of databases and ontologies,
hereby closing the analysis cycle. Finally, the core functionality of the TextGate
system is conceived as a SOAP service so it can be seamlessly integrated with
other systems.

Context-Aware Semantic Association Ranking

[Boanerges Aleman-Meza](), [Chris Halaschek](), [I. Budak Arpinar](), and [Amit Sheth]()

[Large Scale Distributed Information Systems (LSDIS) Lab](),
Computer Science Department, University of Georgia, Athens, GA 30602-7404
{boanerg, ch, budak, amit}@cs.uga.edu

Abstract. Discovering complex and meaningful relationships, which we call Semantic Associations, is an important challenge. Just as ranking of documents is a critical component of today's search engines, ranking of relationships will be essential in tomorrow's semantic search engines that would support discovery and mining of the Semantic Web. Building upon our recent work on specifying types of Semantic Associations in RDF graphs, which are possible to create through semantic metadata extraction and annotation, we discuss a framework where ranking techniques can be used to identify more interesting and more relevant Semantic Associations. Our techniques utilize alternative ways of specifying the context using ontology. This enables capturing users' interests more precisely and better quality results in relevance ranking.

I know what you mean: semantic issues in Internet-scale publish/subscribe systems

Ioana Burcea, Milenko Petrovic, and Hans-Arno Jacobsen

Department of Electrical and Computer Engineering
Department of Computer Science
University of Toronto, Canada
{ioana,petrovi,jacobsen}@eecg.toronto.edu

Abstract. In recent years, the amount of information on the Internet has increased exponentially developing great interest in selective information dissemination systems. The publish/subscribe paradigm is particularly suited for designing systems for routing information and requests according to their content throughout wide-area network of brokers. Current publish/subscribe systems use limited syntax content-based routing. Since publishers and subscribers are anonymous and decoupled in time, space and location, often over wide-area network boundaries, they do not necessarily speak the same language of use the same data and language format. Consequently, adding semantics to current publish/subscribe systems is important. In this paper we identify and examine the issues in developing semantic-aware content-based routing for publish/subscribe broker networks.

A Context-Oriented RDF Database

Mohammad-Reza Tazari

Computer Graphics Center, Dept. Mobile Information Visualization
Fraunhoferstraße 5, 64283 Darmstadt, Germany
Saied.Tazari@zgdv.de

Abstract. The importance of contextual knowledge in knowledge management and organizational memory is shown in topical literature. Even in an initial visionary scenario for the Semantic Web, one can immediately encounter the contextual knowledge needed to realize the necessary services. Hence, it is not inappropriate to claim that context management is an integral service of the Semantic Web. After discussing the distributed nature of contextual knowledge, we define some requirements for a context-oriented database service and then introduce CORD as a service satisfying those requirements based on the Semantic Web technologies. Selected features of CORD that provide some contribution to the discussions within the Semantic Web research community, like embedded resources, query language, and definition of rules, are discussed in some detail.

An Adaptable Service Connector Model[1]

Gang Li[1], Yanbo Han[1], Zhuofeng Zhao[1], Jianwu Wang[1], and Roland M. Wagner[2]

[1] Software Division, ICT, Chinese Academy of Science, PRC
{ligang, yhan}@ict.ac.cn {zhaozf, wjw}@software.ict.ac.cn
[2] Fraunhofer ISST, Dortmund, Germany
roland.wagner@isst.fhg.de

Abstract. Service oriented computing is gaining popularity. In a typical contemporary service-oriented application, service connections are pragmatically implemented using protocols like SOAP. Through this kind of connection, services can be composed into applications. Service composition is regarded as a new approach for developing applications in network environments.

However, service composition still faces serious challenges due to the openness and dynamism of network environment, such as grids. Let us take service grids as an example. Firstly, services freely join in or quit from a grid and most services in a grid continue evolving over time. Secondly, user requirements are subject to dynamic changes in a virtual enterprise environment. All these require service connections to be adaptable, so that service interactions can be easily reconfigured and involved services can be changed dynamically, while changes of service resources and user requirements take place. In this paper, we focus our research on how to make a service connection adapt to those changes.

To make the connection adaptable, the connection structure ought to be changeable. As a semantic concept, role provides an organizing mechanism through which the abstraction of services with common functions is derived and marked by role features. With this mechanism, a role offers flexible connection structure, enabling service connection adaptation by reconfiguration. In this paper, we treat service connections as individual components called service connectors and present an adaptable service connector model that adopts a role mechanism to adjust connections between services. A role is an abstraction of services with common functionalities. It offers a changeable connector structure, enables reconfiguration of service interaction and encapsulates changes in interacting participants, by modifying the feature and related service references.

Building an integrated Ontology within SEWASIE system[*]

Domenico Beneventano[1,2], Sonia Bergamaschi[1,2], Francesco Guerra[1], and Maurizio Vincini[1]

[1] Dipartimento di Ingegneria dell'Informazione
Università di Modena e Reggio Emilia
Via Vignolese 905 - Modena
{lastname.firstname}@unimo.it
[2] IEIIT-CNR Istituto di Elettronica e di Ingegneria
dell'Informazione e delle Telecomunicazioni
Viale Risorgimento 2 - Bologna

Abstract. The SEWASIE (SEmantic Webs and AgentS in Integrated Economies) project (IST-2001-34825) is an European research project that aims at designing and implementing an advanced search engine enabling intelligent access to heterogeneous data sources on the web.
In this paper we focus on the Ontology Builder component of the SEWASIE system, that is a framework for information extraction and integration of heterogeneous structured and semi-structured information sources, built upon the MOMIS (Mediator envirOnment for Multiple Information Sources) system. The result of the integration process is a Global Virtual View (in short GVV) which is a set of (global) classes that represent the information contained in the sources being used. In particular, we present the application of our integration concerning a specific type of source (i.e. web documents), and show the extension of a built-up GVV by the addition of another source.

Ontologies : A contribution to the DL/DB debate.

Nadine Cullot[1], Christine Parent[3], Stefano Spaccapietra[2], and Christelle Vangenot[2]

[1] LE2I Laboratory, University of Burgundy,
BP 47870, 21078 Dijon Cedex, France
`nadine.cullot@u-bourgogne.fr`,
[2] Database Laboratory, Swiss Federal Institute of Technology,
CH-1015 Lausanne, Switzerland
`stefano.spaccapietra@epfl.ch,christelle.vangenot@epfl.ch`
[3] University of Lausanne, CH-1015 Lausanne, Switzerland
`christine.parent@unil.ch`

Abstract. The move to global economy has emphasized the need for intelligent information sharing, and turned ontologies into a kernel issue for the next generation of semantic information services. The push towards an effective use of ontologies as a means to achieve semantic interoperability is, in our opinion, shifting the focus from purely taxonomic ontologies to more descriptive ontologies. These would namely provide agreed descriptions of the data structures representing the complex organization of objects and links of interest within the targeted domain. This paper analyzes the requirements for such descriptive ontologies, and contrasts the requirements to the functionality provided by some current representative approaches that have been proposed for ontology management. Selected approaches originate from research in artificial intelligence, knowledge representation and database conceptual modeling. The paper concludes that extending rich semantic data models with support for reasoning is an interesting alternative to extending description logics with data management functionality.

Efficient RDF Storage and Retrieval in Jena2

Kevin Wilkinson[1], Craig Sayers[1], Harumi Kuno[1], Dave Reynolds[2]

HP Laboratories

[1] 1501 Page Mill Road
Palo Alto, CA, 94304   USA

[2] Filton Road, Stoke Gifford
Bristol BS34 8QZ   United Kingdom
firstName.lastName@hp.com

Abstract. RDF and related Semantic Web technologies have been the recent focus of much research activity. This work has led to new specifications for RDF and OWL. However, efficient implementations of these standards are needed to realize the vision of a world-wide semantic Web. In particular, implementations that scale to large, enterprise-class data sets are required. Jena2 is the second generation of Jena, a leading semantic web programmers' toolkit. This paper describes the persistence subsystem of Jena2 which is intended to support large datasets. This paper describes its features, the changes from Jena1, relevant details of the implementation and performance tuning issues. Query optimization for RDF is identified as a promising area for future research.

An Indexing Scheme for RDF and RDF Schema

based on Suffix Arrays

Akiyoshi MATONO[1], Toshiyuki AMAGASA[1], Masatoshi YOSHIKAWA[2], and
Shunsuke UEMURA[1]

[1] Graduate School of Information Science, Nara Institute of Science and Technology

8916–5 Takayama-cho, Ikoma-shi, Nara 630–0192, Japan

{akiyo-ma,amagasa,uemura}@is.aist-nara.ac.jp

[2] Information Technology Center, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi 464–8601, Japan

yosikawa@itc.nagoya-u.ac.jp

Abstract

The Semantic Web is a candidate for the next generation of the World Wide Web. It is anticipated that the number of metadata written in RDF (Resource Description Framework) and RDF Schema will increase as the Semantic Web becomes popular. In such a situation, demand for querying metadata described with RDF and RDF Schema will also increase, and therefore effective query retrieval of RDF data is important. To this end, we propose an indexing scheme for RDF and RDF Schema. In our (proposed) scheme, we first extract four kinds of DAGs (Directed Acyclic Graphs) from an RDF data, and extract all path expressions from the DAGs. Then, we generate four kinds of suffix arrays based on the path expressions. Using the indices, we can achieve efficient processing of query retrievals on RDF data including schematic information defined by RDF Schema (for example, classes and/or properties).

RDF Core: A component for effective management of RDF Models

FLORIANA ESPOSITO, LUIGI IANNONE, IGNAZIO PALMISANO AND
GIOVANNI SEMERARO

Dipartimento di Informatica
Università degli Studi di Bari
Via Orabona, 4
Bari, 70125, ITALY
+39 080 544 2299
{esposito,iannone,semeraro}@di.uniba.it, ignazio_io@yahoo.it

Abstract.

In order to make Semantic Web effective, the first step was the development of languages that could support data portability, namely XML, metadata descriptions, namely RDF, and ontology management and inference, such as DAML+OIL, OWL etc. Those languages have to be manipulated by applications and many Application Programming Interfaces (APIs) have been developed in order to accomplish this task. Obviously, they differ in implementation details. Moreover, developers often would like to exploit more than an API at a time. Another issue is that a developer would be very advantaged if he could have a uniform support for some services across these frameworks (such as query languages), despite the lack of standards. In this paper, we present a component, called *RDFCore,* developed in order to overcome these problems. We will also illustrate the added value that our framework provides to RDF in order to exploit the full potentiality of the language and to employ it in research as well as in real world applications. Consequently we will provide some test results on the performances of the presented framework.

Sharing Ontology by Web Services:

Implementation of a Semantic Network Service (SNS) in the context of the German Environmental Information Network (gein®)

Thomas Bandholtz

Consultant, Karl-F.-Schinkelstr. 2, 53127 Bonn, Germany
(formerly: Solutions Manager Knowledge Technologies, SchlumbergerSema)
thomas@bandholtz.info

Abstract. A thesaurus, a gazetteer and a chronology have been integrated in a consolidated ontology on the basis of the Topic Map pattern. The result has been made accessible to a working information community of 89 environmental authorities in Germany by Web Services technology. A semantically shared ontology can be shared physically in the Web.

ODE-SWS: A Semantic Web Service Development Environment

Óscar Corcho[1], Asunción Gómez-Pérez[1], and Mariano Fernández-López[1]
Manuel Lama[2]

[1] Departamento de Inteligencia Artificial. Facultad de Informática.
Campus de Montegancedo, s/n. Universidad Politécnica de Madrid.
28660 Boadilla del Monte, Madrid, Spain.
{ocorcho,mfernandez,asun}@fi.upm.es
[2] Departmento de Electrónica y Computación. Facultad de Física.
Campus Sur, s/n. Universidad de Santiago de Compostela.
15782 Santiago de Compostela, A Coruna, Spain.
lama@dec.usc.es, davidal@usc.es

Abstract. Web Services (WS) are software modules that perform operations that are network-accessible through XML messaging. Web Services in the Semantic Web, that is, Semantic Web Services (SWS), should describe semantically their structure and capabilities to enable its automatic discovery, invocation and composition. In this work we present a development environment to design SWS in a language-independent manner. This environment is based on a framework that defines an ontology set to characterize how a SWS should be specified. The core ontology of this framework describes the SWS problem-solving behaviour and enables the SWS design at a conceptual level. Considering this framework, the SWS development environment is composed of (1) a graphical interface, in which the conceptual design of SWSs is performed, and (2) a tool set, which instantiates the framework ontologies according to the graphical model created by the user, verifies the completeness and consistency of the SWS through instance evaluation, and translates the SWS conceptual model description into SWS (and WS) languages, such as DAML-S, WSDL or UDDI. This tool set is integrated in the WebODE ontology engineering workbench in order to take advantage of its reasoning and ontology translation capabilities.

APPLICATIONS OF PSL TO SEMANTIC WEB SERVICES

MICHAEL GRÜNINGER

ABSTRACT. In this paper we will show how the ontology of the Process Specification Language can be used as an upper-level process ontology that serves as the semantic foundation for the DAML-S ontology for web services.

1

H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems [*]

S. Castano, A. Ferrara, and S. Montanelli

Università degli Studi di Milano
DICO - Via Comelico, 39, 20135 Milano - Italy
{castano,ferrara,montanelli}@dico.unimi.it

Abstract

In this paper, we present H-MATCH, an algorithm for dynamically matching distributed ontologies. By exploiting ontology knowledge descriptions, H-MATCH can be used to dynamically perform ontology matching at different levels of depth, with different degrees of flexibility and accuracy. H-MATCH has been developed in the HELIOS framework, conceived for supporting knowledge sharing and ontology-addressable content retrieval in peer-based systems.

A Collaborative Approach for Query Propagation in Peer-to-Peer Systems[1]

Anne Doucet, Nicolas Lumineau

LIP6 Laboratory
University of Paris 6
8, Rue du Capitaine Scott, 75015 Paris, France
FirstName.LastName@lip6.fr

Abstract. Sharing resources on a world-wide scale is a current research topic. Nowadays, peer-to-peer architecture is considered as a scalable solution to this issue. However, a lot of problems to extend this architecture for data storage remain open. When information is not highly replicated, localization of nodes storing relevant data becomes essential to avoid covering completely the network. To this purpose, we try to propagate a query towards nodes potentially storing relevant data. Information about nodes relevance is obtained by users' experiences. In an applicative context where "communities of interest" exist, we create logical and semantic links not only to specify the nodes relevant for a community, but also to link communities with a related interest. The proposed pattern has been created with respect to Peer-to-Peer philosophy and so as to consider the evolution of communities.

OntoMiner: Bootstrapping and Populating Ontologies From Domain Specific Web Sites

Hasan Davulcu, Srinivas Vadrevu, and Saravanakumar Nagarajan

Department of Computer Science and Engineering,
Arizona State University,
Tempe, AZ, 85287, USA
{hdavulcu, svadrevu, nrsaravana}@asu.edu

Abstract. RDF/XML has been widely recognized as the standard for annotating online Web documents and for transforming the HTML Web to the so called Semantic Web. In order to enable widespread usability for the Semantic Web there is a need to bootstrap large, rich and up-to-date domain ontologies that organize most relevant concepts, their relationships and instances. In this paper, we present automated techniques for bootstrapping and populating specialized domain ontologies by organizing and mining a set of relevant Web sites provided by the user. We develop algorithms that detect and utilize HTML regularities in the Web documents to turn them into hierarchical semantic structures encoded as XML. Next, we present tree-mining algorithms that identify key domain concepts and their taxonomical relationships. We also extract semi-structured concept instances annotated with their labels whenever they are available. Experimental evaluation for the News and Hotels domain indicates that our algorithms can bootstrap and populate domain specific ontologies with high precision and recall.

Can Data Mining Techniques Ease The Semantic Tagging Burden?

Fabio Forno[1], Laura Farinetti[1], Sean Mehan[2]

[1] Politecnico di Torino, Dipartimento di Automatica ed Informatica, Torino, Italy
fabio.forno@polito.it, laura.farinetti@polito.it
[2] SMO, University of the Highlands and Islands, Sleat, Isle of Skye, UK
sean@smo.uhi.ac.uk

Abstract

The effective implementation of the Semantic Web vision is highly dependent upon the widespread availability of large collections of semantically rich resources which are trustworthy and meaningful. Since semantic classification is dependent upon complex ontologies, a recognized difficulty is the steep learning curve presented to human classifiers when attempting to utilize such ontologies. One important method to foster an increase in web accessible, semantically tagged resources is to make available tools which allow users to explore and understand relevant ontologies and to present relevant categories with which to tag new data.

In this paper we first describe an experiment of classification, evidencing some of the observed problems that human classifiers have when trying to catalogue a web resource, especially when they are expert in the site domain but not generally expert in classification tasks and are not familiar with the domain ontologies. Subsequently we give a numerical interpretation of the observed quantities in order to build a model capable of measuring the difficulties involved in classification tasks. Towards this end, we introduce a codebook entropy analogy for the the selected keywords, and we interpret the quantity of information given by the keywords as disorder and disagreement between the classifiers.

In the second part we describe preliminary investigations of how an important and powerful data mining technique, Latent Semantic Indexing (LSI), might help in the design and implementation of tools that guide users in semantic tagging tasks. We applied LSI to a large portion of the Open Directory Project (ODP) catalogue, one of the largest repositories of semantically tagged resources available today. We computed statistical information concerning category relationships in the ODP data set, and we incorporated structural information by modifying the LSI space construction process. We built several modified LSI spaces using different techniques, and we have found that spaces containing structure information perform better both in terms of precision and recall when extracting meaningful topics for documents that must be classified.

Formal aspects of querying RDF databases

Claudio Gutierrez[1], Carlos Hurtado[1], and Alberto Mendelzon[2]

[1] Department of Computer Science, Universidad de Chile
{cgutierr,churtado}@dcc.uchile.cl
[2] Department of Computer Science, University of Toronto
mendel@cs.toronto.edu

Abstract We study formal aspects of querying databases containing RDF data. We present a formal definition of a query language for RDF and compare it with other proposals. Our language is intended to make it easy to formalize and prove results about its properties. We study novel features of query languages derived from the presence of blank nodes and reification. Finally we provide complexity results for query processing, static optimization of queries, and redundancy elimination in answers.

Event-Condition-Action Rule Languages for the Semantic Web

George Papamarkos, Alexandra Poulovassilis, Peter T. Wood

School of Computer Science and Information Systems, Birkbeck College, University
of London, London WC1E 7HX
email: {gpapa05,ap,ptw}@dcs.bbk.ac.uk

Abstract

XML and RDF are becoming dominant standards for storing and exchanging
information on the World Wide Web. With their increasing use in dynamic
applications such as e-commerce and e-learning there is a need for the support
of reactive functionality on XML and RDF repositories. Event-condition-action
(ECA) rules are a natural candidate for this.

ECA rules automatically perform actions in response to events provided that
stated conditions hold. They allow an application's reactive functionality to be
defined and managed within a single rule base rather than being encoded in
diverse programs, thus enhancing the modularity and maintainability of the
application. Also, ECA rules have a high-level, declarative syntax and so are
amenable to analysis and optimisation techniques which could not be applied if
the same functionality were expressed directly in programming language code.

An ECA rule has the general syntax

on *event* `if` *condition* `do` *actions*

The event part specifies when the rule should be triggered, the condition part is
a query which determines if the database is in particular state, and the action
part states the actions to be performed automatically if the condition holds.
Executing a rule's actions may in turn trigger further ECA rules, and the rule
execution proceeds until no more rules are triggered.

This paper discusses the provision of ECA rules on XML and RDF reposito-
ries. We begin with a brief review of related work on ECA rules for XML. We
then describe our language for specifying ECA rules on XML repositories, and
present a prototype implementation of it. This language can be used for RDF
data which has been serialised as XML but we are also exploring ECA rule lan-
guages for RDF that will operate on a graph/triple representation. We present
an archetypal such language and also an architecture for distributed deployment
of such RDF ECA rules. Along the way, we discuss directions of further work
for both languages.

The work reported here is part of the ongoing SeLeNe project, which is
investigating techniques for managing RDF repositories of educational metadata,
and providing syndication, personalisation and notification services over this
metadata (see `http://www.dcs.bbk.ac.uk/selene`).

Storing and Querying Ontologies in Logic Databases

Timo Weithöner[1], Thorsten Liebig[2], and Günther Specht[1]

[1] Dept. of Databases and Information Systems
University of Ulm
D-89069 Ulm
`{timo.weithoener|specht}@informatik.uni-ulm.de`
[2] Dept. of Artificial Intelligence
University of Ulm
D-89069 Ulm
`liebig@informatik.uni-ulm.de`

Abstract. The intersection of Description Logic inspired ontology languages with Logic Programs has been recently analyzed in [GHVD03]. The resulting language, called Description Logic Programs, covers RDF Schema and a notable portion of OWL Lite. However, the proposed mapping in [GHVD03] from the corresponding OWL fragment into Logic Programs has shown scalability as well as representational deficits within our experiments and analysis. In this paper we propose an alternative mapping resulting in lower computational complexity and more representational flexibility. We also present benchmarking results for both mappings with ontologies of different size and complexity.

DESIGN REPOSITORIES ON THE SEMANTIC WEB WITH DESCRIPTION-LOGIC ENABLED SERVICES

Joseph B. Kopena* and William C. Regli**

Geometric and Intelligent Computing Laboratory
Department of Computer Science, College of Engineering
Drexel University
`http://gicl.cs.drexel.edu/`

Abstract. All engineering firms maintain archives of previously designed artifacts, often in the form of databases of computer aided design (CAD) data. Design repositories are an evolution of such databases to include more heterogenous information and to provide enhanced capabilities through the application of knowledge representation techniques. This paper introduces on-going work on applying description logic and the Semantic Web to constructing such design repositories.

---

* At the National Institute of Standards and Technology during preparation of this document.
** Also of the Department of Mechanical Engineering; Email: `regli@drexel.edu`

Mediation of XML Data through Entity Relationship Models

Irini Fundulaki[1][*] and Maarten Marx[2][**]

[1] Bell Laboratories, Lucent Technologies, USA and
INRIA-Rocquencourt, France.
`fundulaki@research.bell-labs.com`
[2] Institute for Logic Language and Computation,
University of Amsterdam, The Netherlands.
`marx@science.uva.nl`

Abstract. This paper describes an approach for the querying of heterogeneous XML resources using an ontology-based mediator. Here an ontology is an Entity-Relationship schema defined independently of the schemas of the data sources. The sources are described to the mediator by means of mapping rules as in the Local-As-View approach to data integration. User queries are conjunctive queries formulated in terms of the ontology, and answers to these queries are obtained by rewriting them to XQuery expressions and evaluating these on the data sources. A formal semantics for queries is defined by interpreting XML sources into ER models. As there can be many such interpretations, a certain answer to a query is one which is true in all of them. We describe the rewriting algorithm and we show its completeness and correctness with respect to the given semantics. We also give an algorithm for producing a canonical model of the ontology and the interpreted data sources. It is shown that the certain answers can also be obtained by evaluating the query to just this one model.

The ICS-FORTH SWIM: A Powerful Semantic Web Integration Middleware[*]

V. Christophides[1], G. Karvounarakis[1], I. Koffina[1], G. Kokkinidis[1], A. Magkanaraki[1], D. Plexousakis[1], G. Serfiotis[1], and V. Tannen[2][**]

[1] Institute of Computer Science - FORTH
Vassilika Vouton, PO Box 1385, 71110, Heraklion, Greece
{christop, gregkar, koffina, kokkinid, aimilia, dp,
serfioti}@ics.forth.gr
[2] Department of Computer and Information Science, University of Pennsylvania
200 South 33rd Street, Philadelphia, PA 19104-6389, USA
val@cis.upenn.edu

Abstract. Semantic Web (SW) technology aims to facilitate the integration of legacy data sources spread worldwide. Despite the plethora of SW languages (e.g., RDF/S, DAML+OIL, OWL) recently proposed for supporting large scale information interoperation, the vast majority of legacy sources still rely on relational databases (RDB) published on the Web or corporate intranets as *virtual* XML. In this paper, we advocate a Datalog framework for mediating high-level queries to relational and/or XML sources using community ontologies expressed in a SW language such as RDF/S. We describe the architecture and the reasoning services of our SW integration middleware, called SWIM, and we present the main design choices and techniques for supporting powerful mappings between different data models, as well as, reformulation and optimization of queries expressed against mediation schemas and views.

Semantic Representation of Contract Knowledge using Multi Tier Ontology

Vandana Kabilan

Paul Johannesson

Department of Computer and System Sciences,
Stockholm University and Royal Institute of Technology
FORUM 100 , SE 164 40,Kista, Sweden
{Vandana, Pajo}@dsv.su.se

Business contract knowledge exists dispersed in different domains. For successful business process functioning, a precise, clear understanding and interpretation of contractual terms and conditions is required. A semantic interpretation of *contract obligations* and their required *performances* to *fulfill* the obligations, is aimed to bridge the existing gap between business process management and contract management. The increasing impact of e-commerce also necessitates the requirement for centralized, reusable knowledge bases. This paper presents conceptual models and an ontological representation methodology for capturing semantic interpretations of business contracts in a *Multi Tier Contract Ontology*.

The Visual Semantic Web: Unifying Human and Machine Semantic Web Representations with Object-Process Methodology

Dov Dori

Technion, Israel Institute of Technology, Haifa 32000, Israel
dori@ie.technion.ac.il, and
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
dori@mit.edu

Abstract

The Visual Semantic Web (ViSWeb) paradigm enhances human accessibility to the current Semantic Web technology by enabling the visualization of knowledge. Arguing against the claim that humans and machines need to look at different knowledge representation formats, Object-Process Methodology (OPM) is shown to enable modeling of systems in a single graphic and textual model. ViSWeb provides for representation of knowledge over the Web in a unified way that caters to humans as well as machines. ViSWeb is developed as an OPM-based layer on top of XML/RDF/OWL to express knowledge visually and in natural language. Both the graphic and the textual representations are strictly equivalent. Being intuitive yet formal, they are not only understandable to humans, but are also amenable to computer processing. The advantages of the ViSWeb approach include equivalent graphic-text knowledge representation, visual navigability, semantic sentence interpretation, specification of system dynamics, and complexity management. The ability to use such bimodal knowledge representation that is both human understandable and machine processable is a major step forward in the evolution of the Semantic Web.

Interaction and navigation for a document database: a concrete case study

Isabelle Berrien, François Laburthe and Jean-David Ruvini

e-lab BOUYGUES SA
1, avenue Eugène Freyssinet,
78061 Saint Quentin en Yvelines, FRANCE
Mail :iberrien@bouygues.com
Tél : +33 1 30 60 53 66 Fax : +33 1 30 60 22 15

Abstract

In this article, we present the application Wishbone, an innovative search engine to explore complex document database. Dedicated to the intranet document site of our society, we present in this paper how we managed to improve the request process by avoiding the "all or nothing" syndrome for a too fuzzy or too demanding request respectively, offering also the maximum of smoothness in the way the user formulates the query. We particularly focused our attention on the interaction between a web database and its users, either during a consultation or an interrogation process. This web context can be found in multiple environments like an intranet, a FAQs site, self-assistance, sale catalogues or document data bases. First we analyse document consultation sites in regard to quality criteria. In a second step, we describe our approach to furnish a software response to the problematic of settling, management and interaction of an online document database. All this work was accomplished using as example an enterprise document database with various documents like contracts, business notes and bug reports.