

Proposal for an Information Systems Special Issue on Semantic Web Data Management

Roberto De Virgilio¹, Fausto Giunchiglia², Francesco Guerra³, Letizia Tanca⁴, Yannis Velegrakis²

¹*Dipartimento Informatica e Automazione, Università Roma Tre, Rome, Italy*

²*Department of Information Engineering and Computer Science, University of Trento, Italy*

³*Dipartimento di Economia Aziendale, Università di Modena e Reggio Emilia, Modena, Italy*

⁴*Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milan, Italy*

dvr@dia.uniroma3.it, fausto@dit.unitn.it, francesco.guerra@unimore.it, tanca@elet.polimi.it, velgias@disi.unitn.eu

Key words: RDF, Data Management, Metamodeling, Semantic Web.

1. Introduction

During the last decade we have witnessed a tremendous increase in the amount of data that is available on the Web in almost every field of human activity. Financial information, weather reports, news feeds, product information, and geographical maps are only few examples of such data, all intended to be consumed by the millions of users surfing the Web. The advent of Web 2.0 applications, such as Wikis, social networking sites and mashups have brought new forms of data and have radically changed the nature of modern Web. They have transformed the Web from a publishing-only environment into a vibrant place for information exchange. Web users are not longer plain data consumers but have become active data producers and data dissemination agents, contributing further to the increase of the information plethora on the Web.

For the successful discovery, sharing, distribution and organization of this information, the ability to understand and manage the semantics of the data is of paramount importance. This need gave birth to the Semantic Web vision that aims at the creation of a well-defined, reusable and machine-understandable form of the semantics of the web data. The exponential growth of web data, however, made clear the need for contributions from different disciplines in order to enable the successful implementation of such a plan. Only through interdisciplinary synergies, can advanced and efficient data management techniques for the web be achieved. Two disciplines of major importance are Conceptual Modeling and Data Management.

Both are research fields that have been around for more than three decades, are mature enough and can offer high quality research results alongside efficient implementations that tackle real world application scenarios.

Traditionally, developments in these two fields and also on the Semantic Web have followed independent paths. The modern reality, however, makes clear that research efforts and events have to focus on the intersections of these areas. Conceptual modeling deals with the development of advanced techniques and tools that allow accurate representation of, and reasoning on, artifacts that model real-world objects and information. Data management, on the other hand, deals with the development of techniques for the efficient and effective storage, querying, retrieval and management of large amounts of information of different nature, i.e., relational data, XML documents, blogs, social networking data, multimedia items, etc.

The aim of this special issue on Semantic Web Data Management is to collect contributions that span the above areas and promote the implementation of the Semantic Web vision. To improve the quality and technical depth of the publications we will build on the results of two workshops that we organized on these topics: *Semantic Web Information Management*¹ (SWIM) and *Data Engineering meets the Semantic Web*² (DESWeb).

¹<http://mais.dia.uniroma3.it/SWIM09>

²<http://desweb2010.unimore.it>

The workshops addressed complementary topics having as a final goal the support of data management for the Semantic Web. Our intent is to ask authors of the best papers in the two workshops to contribute to this special issue by producing extended versions of their submissions, containing significant new research material.

Besides the extended papers from the workshops, we will seek original articles examining open problems, research results, and future research directions in semantic web data management. Appropriate topics include:

Semantic Web Data Querying and Reasoning:

- Data search, integration, and analysis on the Semantic Web
- Query languages for ontologies
- Searching and Ranking ontologies
- Reasoning paradigms for Web query languages
- Rule interchange formats and Rule markup languages
- User-friendly interfaces to query and reason on the Web
- Ranked Keyword search on Semantic Web data

Semantic Data Integration:

- Semantic-aware matching and mapping
- Management of incomplete information on the Semantic Web
- Dataspaces and the Semantic Web
- Semantic-aware search engines
- Role of ontologies in data integration and data exchange
- Semantic-based keyword searching in databases
- Probabilistic data and query answering
- Data Management for Social Networking
- Semantic-based lifecycle management
- Database support for social Web2.0 applications
- Entity identification on the Semantic Web
- Tool support for developing and deploying data intensive

Semantic Web application:

- Publishing relational data on the Semantic Web
- Semantic annotation
- Semantic evolution
- Understanding and managing Wiki/Blog data
- Semantic Web technologies for cooperation and collaboration
- Business processes on the Semantic Web
- Semantic-based service discovery and composition
- Interoperation via Semantic Web services
- Query languages for RDF and the Semantic Web
- Semantic Web Services
- Semantic Bioinformatics
- Multimedia management on the Semantic Web

Core Database Technologies for RDF Data:

- Data models and query languages for RDF
- RDF querying processing and optimization
- Indexing and access methods for RDF data
- RDF storage and compression
- Update management and maintenance of RDF data
- Query answering techniques on RDF databases
- Benchmarking RDF Systems and applications

2. Content Contributions

According to the instructions to Guest Editors, we propose a special issue that mixes the features of the two call types *Collecting best papers from a conference* and *Soliciting papers*.

Therefore, the contents of the special issue will be derived from three different sources. The first two will be some selected papers from the two workshops above, since the workshops are excellent concentration points of research publications highly related to the topic of the issue. Moreover, a small number of invited contributions will be solicited from highly distinguished researchers.

Our goal is to receive about 15-20 submissions, and to select among them, by means of the review process, the most interesting papers (5-8).

2.1. Workshop Selected Papers

SWIM The *International Workshop on Semantic Web Information Management (SWIM)* aims at reviewing the most recent data-centered solutions for the Semantic Web. In particular, the workshop ambition is to present and analyze techniques for semantic information management, taking advantage of the synergies between the logical basis of the Semantic Web and the logical foundations of conceptual modeling. A leitmotif of these research areas is to propose the right models and methods for representing and managing structured data, and enabling or facilitating its processing by machines that operate on the web. The long-standing experience of the information modeling community can provide a significant contribution to the substantial issues arising in the Semantic Web community through the use of technologies such as RDF, RDFS and OWL. The workshop calls researchers and practitioners to debate, propose, and elaborate on the foundations for a data-modeling approach to modern Semantic Web problems. It includes presentations of running research projects, both at the theoretical and at the practical level. The topics of the workshop include, but are not limited to: efficient and effective storage of large amounts of semantic data, querying and reasoning about it, and exploiting the semantic data in real world scenarios.

The workshop takes place in conjunction with the IEEE International Conference on Semantic Computing, a highly ranked international forum for researchers and practitioners to present research that advances the state of the art and practice of Semantic Computing. A number of 2-3 papers is expected to be selected from those accepted in the workshop and be considered in the Information Systems issue. These submission will have to go through an evaluation phase.

DESWEB The goal of the *International workshop on Data Engineering meets the Semantic Web (DESWeb)* is to bring together researchers and practitioners from both fields of Data Management and Semantic Web. It aims at investigating the new challenges that Semantic Web technologies have introduced and the new ways through which these technologies can improve existing solutions. Furthermore, it intends to study what data management systems and technologies can offer in order to improve the scalability and performance of Semantic Web applications. The workshop areas of interest are on semantic data integration and on core database technologies for RDF data. The former area, in particular, includes topics such as semantic aware schema matching and mapping, management of incomplete information and probabilistic query answering on the Semantic Web, dataspace and pay-as-you-go integration, semantic search engines, use of ontologies in data integration, semantic-based keyword searching, data management for social networks, entity identification, relational database publishing on the Web, semantic annotation and evolution, and business processes on the Web. The core database technologies for RDF data, on the other hand, include query languages, processing and optimization for RDF and semantic Web data, indexing and access methods for RDF, storage and compression. To capture the dynamic nature of the modern web data, update management and maintenance are also among the interesting topics of the workshop. Since semantic and RDF data management is a relatively new topic, the development of benchmarking techniques for these applications is an interesting issue.

The workshop takes place in conjunction with the IEEE International Conference in Data Engineering, one of the leading conference in Data Management. We expect a number of 2-3 papers to be selected from those accepted in the workshop and be considered in the Information Systems issue. As in the case of SWIM, these papers will also have to go through an evaluation phase.

2.2. Solicited Papers

There is a significant number of distinguished researchers currently performing research work on topics highly related to those of the proposed special issue. Thus, we believe that a small number of papers solicited to high quality research groups will highly advance the quality of the contents of the special issue.

3. Potential Content

3.1. Workshop selected papers

DESWEb

- **Processing Online News Streams for Large-Scale Semantic Analysis** [*Milos Krstaji, Florian Mansmann, Andreas Stoffel, Martin Atkinson, Daniel A. Keim*] The paper provides a framework for the semantic analysis of news streams. It describes in details the emerging challenges and the developed solutions that enable processing of incoming articles close to real-time. It also provide two case studies.
- **An Ontology-Based Retrieval System Using Semantic Indexing** [*Soner Kara, Ozgur Alan, Orkunt Sabuncu, Samet Akpınar, Nihan K. Cicekli, Ferda N. Alpaslan*] In this paper, the authors present an ontology-based information extraction and retrieval system and its application to the soccer domain. The three main issues studied are the usability, scalability and the performance of the semantic search.
- **Extensions to the Pig Data Processing Platform for Scalable RDF Data Processing Using Hadoop** [*Yusuke Tanimura, Akiyoshi Matono, Steven Lynden, Isao Kojima*] In order to effectively handle the growing amount of available RDF data, The paper is about a a scalable and flexible RDF data processing framework. It extends the Pig data processing platform on top of the Hadoop infrastructure and present an evaluation study of its performance.

SWIM

- **Building a Semantic Web of Things: issues and perspectives in information compression.** [*Floriano Scioscia and Michele Ruta*]. The paper discusses a novel framework to compress semantic data by using a path oriented representation. The novelty is to be able to query the compressed data by a XPath-like language (without unzip the document). The language can be mapped easily into sparql.
- **A Semantic Web enabled System for Resume Composition and Publication.** [*Roberto Mirizzi, Tommaso Di Noia, Eugenio Di Sciascio and Michelantonio Trizio*]. The authors present a system to automatically produce a semantically annotated resume exploiting domain knowledge modeled with respect to a domain ontology. This application exploits an approach to annotate web pages by using RDFa format.

3.2. Solicited Papers

The list below contains a number of potential researchers and /or research groups to be contacted for contribution, grouped by research areas.

Semantic Web Data Querying and Reasoning:

1. Marcelo Arenas, Pontificia Universidad Catolica de Chile, Chile
2. Francois Bry, University of Munich, Germany
3. Georg Gottlob, Computing Science at Oxford University
4. Jan Hidders, TU Delft

5. Ian Horrocks, Computing Science at Oxford University

Semantic Data Integration:

1. Paolo Atzeni, Università di Roma 3, Italy
2. Phil Bernstein, Microsoft Research, USA
3. Silvana Castano, Università degli Studi di Milano, Italy
4. Howard Ho, IBM Almaden Research Center, USA
5. Maurizio Lenzerini, Sapienza, Università di Roma, Italy
6. Felix Naumann, University of Potsdam, Germany

Semantic Web applications:

1. Valeria De Antonellis, Università di Brescia, Italy
2. Christian Bizer, Freie Universität Berlin, Germany
3. Marco Colombetti, Politecnico di Milano, Italy

Core Database Technologies for RDF Data

1. Vassilis Christophides, ICS-FORTH, Greece
2. Georg Lausen, University of Freiburg, Institute for Computer Science
3. Gerhard Weikum, Max-Planck-Institut für Informatik, Germany

4. Reviewing process

The review process will be scheduled according to the deadlines stated by Information Systems. We will use the Elsevier web system to manage the entire review process, which allows reviewers to read the abstract and the submission before accepting the review. This helps to guarantee high quality reviews and the selection of the best papers, which must be reviewed by at least three reviewers. Our previous experience shows an acceptance rate of papers ranging from 25% to 40%.

5. Timetable

The is the tentative scheduling for this special issue:

March/April 2010: Invitation distribution

12 September 2010: Submission

20 November 2010: 1st round review

18 December 2010: 2nd round review and notification

April/May 2011 Publication

6. Editors Information

Roberto De Virgilio

Research Assistant

Dip. di Informatica e Autom., Università Roma Tre, Via della Vasca Navale, 79, 00146 Rome, Italy.

Tel: (+39) 06 5733 3229, Fax: (+39) 06 5733 3612

E-mail: dvr@dia.uniroma3.it

Homepage: <http://mais.dia.uniroma3.it/rodevirg>

Roberto De Virgilio is a Research Assistant at the University of "ROMA TRE" under the supervision of Prof. Riccardo Torlone. His PhD referred to Adaptation of Web Information Systems respect to the context

of a client. The last years his research focuses on Semantic Web Information Management at different levels of abstraction. He is the author of several papers on Web Information Systems management and Semantic Web published in international journals and conferences, a chapter of the book “*Mobile Information Systems: Infrastructure and Design for Adaptivity, and Flexibility*” edited by Springer and editor of a Springer-Verlag book on Semantic Data Management which will be published by December 2009.

His research interests are: Data integration, adaptive information systems, Web based information systems, XML data management, Semantic Web, data modeling and database design, model management, Web Information Systems, Personalization.

Fausto Giunchiglia

Full professor

University of Trento, Via Sommarive 14, 38123 Trento, Italy.

Tel: (+39) 0461 88 15 33, Fax: (+39) 0461 88 20 93

E-mail: fausto@dit.unitn.itHomepage: <http://www.dit.unitn.it/~fausto/>

Fausto Giunchiglia, ECCAI Fellow, is a Professor of Computer Science. He has done research in various related areas including knowledge management, data and knowledge representation, reasoning with context and formal methods. He has been program or conference chair various events, including: IJCAI 2005, Context 2003, AOSE 2002, Coopis 2001, KR&R 2000. He has been editor or editorial board member of around ten journals, including: Journal of Autonomous Agents and Multi-agent Systems, Journal of applied non Classical Logics, Journal of Software Tools for Technology Transfer, Journal of Artificial Intelligence Research. He has been Member of the ECCAI Fellows Selection Committee, of the IJCAI Board of Trustees member (01-11), President of IJCAI (05-07), President of KR, Inc. (02-04), Advisory Board member of KR, Inc., Steering Committee member of the CONTEXT conference. Relevant to the topic of this proposal, he has developed influential work on context,, ontology (semantic) matching, and data and knowledge management, and lately also on semantic Web and Web 2.0 related issues (e.g., lightweight ontologies).

Francesco Guerra

Assistant Professor

Universita di Modena and Reggio Emilia, V.le Berengario 51, Italy.

Tel: (+39) 059 205 6264, Fax: (+39) 059 205 6129

E-mail: francesco.guerra@unimore.itHomepage: <http://www.dbgroup.unimo.it/~guerra/>

Francesco Guerra received his PhD in information engineering at the University of Modena and Reggio Emilia in 2003. Since November 2005, he has been an assistant professor of Computer Engineering at the Faculty of Economics at the University of Modena and Reggio Emilia, where he teaches computer science and enterprise information systems. His main research interests include integration of heterogeneous information sources, ontologies, and the Semantic Web. He is part of the program committee of international conferences and workshops, he was co-chair of the SWAE workshop (Semantic Web Architectures For Enterprises); he was reviewer for the Information Sciences Journal (ELSEVIER), for the Electronic Commerce Research and Applications journal (ELSEVIER), for the IEEE/ACM Transactions on Computational Biology and Bioinformatics, for the International Journal of Metadata, Semantics and Ontologies (IJMSO), and he is guest editor of the special issue on Search using Metadata, Semantic, and Ontologies of the IJMSO journal. He will be guest editor of the special issue on Information Overload of the IEEE Internet Computing Review (2010).

Letizia Tanca

Full professor

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Via Ponzio 34, 20133, Milan, Italy.

Tel: (+39) 02 2399 3531, Fax: (+39) 02 2399 3411

E-mail: tanca@elet.polimi.itHomepage: <http://tanca.dei.polimi.it/>

Letizia Tanca is a Full Professor of Technologies for Information Systems and of Database Systems and an author of about 100 papers published in international journal and conference proceedings, on databases and database theory, author of the book “Logic Programming and Databases”, coauthored with S. Ceri and G. Gottlob, and editor of a Springer-Verlag book on Semantic Data Management which will be published by December 2009. She has been involved in European and Italian research projects, often as the local coordinator. Her research interests range over all database theory, especially on deductive, active and object oriented databases, graph-based languages for semistructured data. Her most recent research interests include context-aware database design and integration, mobile databases and semantic-based data integration. Letizia Tanca has been a referee of several international journals, and a member of the program committee of a large number of international conferences, and is a member of the board of the Informatics Europe association.

Yannis Velegarakis

Assistant professor

University of Trento, Via sommarive 14, 38123 Trento, Italy

Tel: (+39) 0461 883986

E-mail: velgias@disi.unitn.euHomepage: <http://www.disi.unitn.eu/~velgias>

Yannis Velegarakis is a faculty member of the Department of Information Engineering and Computer Science of the University of Trento. His research area of expertise includes tool building for schema and ontology mapping, interoperability, data translation, information integration, data exchange, view updates, view maintenance and meta-data management. Prior to joining the University of Trento, he held a researcher position at AT&T Research Labs in the US. He has also spent time as a visitor at the University of California, Santa-Cruz, the IBM Almaden Research Center, and the Center of Advanced Studies of the IBM Toronto Lab. He was a member of the committee for the CIMI cultural profile of the ANSI/NISO Z39.50 standard. He has served in program committees of many national and international conferences, has been a reviewer for numerous international journals, and holds 3 US patents. In 2006 he became Marie Currie Reintegration fellow.