PREMIER REFERENCE SOURCE

# Advancing Information Management through Semantic Web Concepts and Ontologies

Patricia Ordóñez de Pablos, Héctor Oscar Nigro, Robert Tennyson,
Sandra Elizabeth González Císaro & Waldemar Karwowski

# Advancing Information Management through Semantic Web Concepts and Ontologies

Patricia Ordóñez de Pablos
*University of Oviedo, Spain*

Héctor Oscar Nigro
*Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*

Robert Tennyson
*University of Minnesota, USA*

Sandra Elizabeth González Císaro
*Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*

Waldemar Karwowski
*University of Central Florida, USA*



Information Science
REFERENCE

| | |
|---|---|
| Managing Director: | Lindsay Johnston |
| Editorial Director: | Joel Gamon |
| Book Production Manager: | Jennifer Romanchak |
| Publishing Systems Analyst: | Adrienne Freeland |
| Development Editor: | Myla Merkel |
| Assistant Acquisitions Editor: | Kayla Wolfe |
| Typesetter: | Travis Gundrum |
| Cover Design: | Nick Newcomer |

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com



Library of Congress Cataloging-in-Publication Data

Advancing information management through Semantic Web concepts and ontologies / Patricia Ordóñez de Pablos ... [et a].
p. cm.
Includes bibliographical references and index.
Summary: "This book provides an analysis and introduction on the concept of combining the areas of semantic web and web mining, emphasizing semantics in technologies, reasoning, content searching and social media"--Provided by publisher.
ISBN 978-1-4666-2494-8 (hardcover) -- ISBN 978-1-4666-2495-5 (ebook) -- ISBN 978-1-4666-2496-2 (print & perpetual access) 1. Semantic Web. 2. Data mining. 3. Ontologies (Information retrieval) 4. Information resources management. I. Ordóñez de Pablos, Patricia, 1975-
TK5105.88815.A528 2012
006.3'12--dc23

2012023343

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

# Editorial Advisory Board

# Preface

## INTRODUCTION

This book provides an introduction and in-depth analysis of the important concepts that underpin this change. It covers semantics: the technologies, the deep Web, reasoning, content searching, Voice-Video-Speech (VVS) searching, multimedia, social media, domain-oriented applications, and others.

The Semantic Web is a vision for the future in which Web content can be manipulated by automated systems for analysis and synthesis. Contrast this with the current Web, where information is published mainly for human consumption; for, although pages are human readable, browsers are able only to interpret HTML mark-up in order to visualize its content. The three most difficult tasks—content interpretation, selection, and management—can currently only be done by humans. The Semantic Web will redress the balance between machine and human, bringing these three most difficult tasks within the remit of the automatic.

Semantic Web mining aims to combine the development of two research areas, namely Semantic Web and Web mining. Web mining extracts information from the content of the pages, its structure of relationships (links), and the users' browsing records.

Both areas are collaborating in different ways:

1. Mining Web techniques can help create a Semantic Web. A very important portion of Semantic Web is the ontologies. The ontology is represented as a set of concepts and their relevant interrelations for certain domains of knowledge. The challenge is to learn ontologies and/or their concepts, in order to make a scalable solution for a wide range of Semantic Web technologies.
2. Knowledge in the form of ontologies, or other forms of representation of knowledge, can be used to improve the process and the results of mining Web. Knowledge provided by the ontology is useful in defining the structure and the scope for Web content mining.

In the interaction between ontologies and Web mining, we can mention the following applications:

1. Sentiment analysis, also called opinion mining, is responsible for classifying words, texts, or documents of opinion, and emotions or feelings that express agreement. It works in tagging texts and their components, which indicate if the expression is positive, negative, or neutral, and in the field of the subjectivity of texts, as well. The area is called affective computing, i.e., the development of means to enable machines to detect and to respond in an appropriate manner to the emotions of users.

2. Optimization in search engines is criteria that must be taken into account in order to plan a campaign of positioning a Web page, its life cycle, and a selection of tools to help analyze the positioning of a site, as well as give keys to improve it. The study of the various basic attributes of an individual Web resource is not sufficient to infer the different strategies of positioning of a search engine. The fundamental problem is the relationship between the different elements of the page and weight that each one brings to the final positioning. The application of various techniques of inductive learning can be a starting point for the optimization of the Web positioning.
3. Web Intelligence (WI) is a research paradigm aimed at exploration of the fundamental interactions between Artificial Intelligence (AI), advanced engineering, and Advanced Information Technology (AIT). Engineering, here, is a general term referring to (a) a new area, for example, informatics of the brain, IA human level, intelligent agents, and intelligence social networks, and (b) classics, such as engineering knowledge, representation, planning, discovery, and data extraction.

The ontologies will give semantic richness to the Web mining process; thus, they will facilitate the understanding of the results obtained and the whole process performance.

The subject area is a combination of Semantic Web, business, management, and information systems.

## OBJECTIVE OF THE BOOK

The book intends to be an international platform to bring together practitioners, academics, researchers, decision makers, those in government, policy makers, and practitioners from different backgrounds to share new theories, research findings, and case studies, enhancing understanding and collaboration in Semantic Web and the role of information technologies and analyses of recent developments in theory and practice.

In addition, this publication comprises the state-of-the-art, innovative theoretical supports, advanced and successful implementations, as well as the latest empirical research findings in the iteration field of ontologies and Web mining.

## CONTENTS OF THE BOOK

Currently, key technologies include those for the fine-grained description of electronically served content. With fine-grained descriptions, automated processing is possible. One proposed vocabulary is RDF, which enables the description of resources and the relationships between them. The main point of RDF is that the relationships between resources are globally identified by URIs. The Semantic Web approach breaks the concept of a Web page as the smallest unit of information to enable the creation of resource descriptions with finer granularity. For example, instead of the homepage of a person, it would be possible to refer to the phone number of that person. In the case of textual information, that break-up is more affordable, although not easy, given that it is possible to access to paragraphs, words, etc. facilitating syntactic searches from keywords. In this way, traditional search engines do a reasonably good job to access text-based information.

Novel and interesting researches in the interaction of Web mining and ontologies are presented, such as ODIX platform, a meta-search tool to deliver search results structured according to the specific users' interests, ontology-based optimization for positioning websites in search engines, improvement of the relevance of Web search results via search-term disambiguation, and ontological filtering of search results based on socially constructed search concepts, and sentiment analysis or opinion mining with ontologies.

## Target Audience

The audience for such a book includes politicians, policy makers, government officers, students (undergraduate, postgraduate, master, PhD level), corporate heads of firms, senior general managers, managing directors, board directors, academics and researchers in the field both in universities and business schools, information technology directors and managers, and libraries and information centres serving the needs of the above, among others.

## Recommended Topics

Topics include, but are not limited to, the following: digital libraries, hypertext and hypermedia semantic, information security in semantic processing, knowledge management, languages (RDF, RDF Schema, OWL, etc.), learning (individual, group, organizational level, industry-academia), ontology for semantic interoperability, ontology learning, ontology mapping and visualization, ontology-based evaluation, progressive ontologies, scalability to the Web level, semantic applications/platforms/tools, semantic blogs, semantic content searching, theoretic foundations on Web mining ontologies-based approach, ontologies in Web mining process, ontology-based interpretation and validation of mined knowledge, interaction from ontologies to Web mining, ontology-based optimization in search engines, opinion mining with ontologies, real study cases, implementing Web mining based on ontologies, and other topics.

*Patricia Ordoñez de Pablos*
*University of Oviedo, Spain*

*Héctor Oscar Nigro*
*Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*

*Robert D. Tennyson*
*University of Minnesota, USA*

*Sandra Elizabeth González Císaro*
*Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*

*Waldemar Karwowski*
*University of Central Florida, USA*

# Acknowledgment

We wish to express our sincere gratitude to Dr. Mehdi Khosrow-Pour and Jan Travers, whose enthusiasm motivated us to accept their invitation for taking on this project, for the opportunity they have given us.

Special thanks also go to all the staff at IGI Global, whose contributions throughout the whole process from inception of the initial idea to final publication have been invaluable. In particular, to Joel A. Gamon, Development Manager, and Myla R. Merkel, Development Editor, for their help in advising us how to solve some problems and for their guidance and professional support. As well, we thank our first assigned Assistant Development Editors, Hannah Abelbeck and Christine Bufton.

Most of the authors of the chapters included in this book also served as referees for articles written by other authors, without whose support the project could not have been satisfactorily completed. Thanks go to all those who provided constructive and comprehensive reviews.

Once again, as editors we would like to thank all the contributing authors for their excellent papers and patience with the process. Making this kind of compilation is a huge responsibility and involves people with very different experiences and analyses.

Finally, we want to thank our families and friends for their love and support throughout this project.

*Patricia Ordoñez de Pablos*
*University of Oviedo, Spain*

*Héctor Oscar Nigro*
*Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*

*Robert D. Tennyson*
*University of Minnesota, USA*

*Sandra Elizabeth González Císaro*
*Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina*

*Waldemar Karwowski*
*University of Central Florida, USA*

*May 2012*

# Table of Contents

# Detailed Table of Contents

a specific domain. On the other hand, folksonomies show a lack of formality because of their implicit ambiguity and flexibility by definition. The main objective of this chapter is to outline and evaluate a new way to exploit Web information resources and tags for bridging the gap between ontology modeling and folksonomies.

**Chapter 3**



The chapter reports the implementation and validation of the knowledge process visualization technology that extends the basic capabilities of the Semantic MediaWiki platform. The Design Project Visualizer has been developed in the ACTIVE Integrating Project of the Seventh Framework Programme of the European Union for the case study in the engineering design in Microelectronics and Integrated Circuits. The concept of knowledge process visualization is based on the paradigm of project navigation. The knowledge workers in this domain are design project managers, designers, and design support engineers. The visualization suggests optimized performance, points to the bottlenecks in executions, and fosters collaboration in development teams. The authors describe the software prototype architecture and implementation. The components and the solution for process knowledge transformation between ontological representations as well as the visualization are presented in detail. Validation results indicate that the solution is helpful in providing expert assistance to design project managers performing their typical tasks of project planning and execution management.

**Chapter 4**



The multimedia information system represents a specific form of information system. This research area suffered many changes in direction due to technology shifts. The general problem is that few years back, multimedia technologies had been limited to relatively simple, stand-alone applications, but multimedia systems, particularly Web-based systems grew in complexity and intervened with many critical issues for development. In this chapter, a specific focus will be cast on existing methodology approaches, their upsides and downsides, and on the surveys and research done by distinguished authors in this area on what sort of methodologies are used in practice. Afterwards, the focus of this chapter will be on whether existing development methodologies can be applied to multimedia systems and if there is any need to adapt them for that specific purpose.

**Chapter 5**



This chapter examines the ethical questions and actions emerging from academic social networking. Academics have always been involved in rigorous discourse across multiple contexts, involving generation, exploration, analysis, evaluation, and application of ideas through a process of thought, research, peer

validation, and publication. The argument is that the concept of collective intelligence is changing the traditional hierarchical "rules" associated with academic dialogue. Collective intelligence is defined as a mix of formal and informal conversational contexts, and the storing and sharing of ideas and information through multiple public online contexts. The meta-concept of collective intelligence presents a number of ethical dilemmas and questions related to privacy, and ownership and control of net-generated data, ideas, and information. The purpose of this chapter is to identify and describe these ethical issues and actions in relation to academic social networking.

**Chapter 6**



The increasing strength and usefulness of semantic technologies have led to innovative decision support processes and management of partners and R&D call for proposals. This book chapter introduces the SEM-IDi project, an architecture that integrates R&D processes with project management software and semantic customized environments. This SEM-IDi platform is composed by two main modules: General Management Module (GMM) which will be responsible for general management of diverse initiatives and projects, and the Semantic and Competence Module (SCM) which will provide functionalities related mainly to decision making.

**Chapter 7**



Cloud Computing is finding its way into the architecture of current IT landscapes. The present chapter depicts an algorithm-based methodology supporting the Root-Cause-Analysis in the context of malfunctioning Federated ERP (FERP) software in Enterprise Clouds. The challenge is to standardize the error-finding procedure and increase the efficiency. For a given error symptom it is shown that the error location is approximated iteratively with help of generic operators in a semiautomatic manner. This approach of Semantic Debugging outperforms classical methods of Technical Debugging in efficiency regarding prerequisite knowledge and time consumption. Semantic integration and maintainability correlate strongly. The Delta-Operator enables the reconstruction of semantic FERP integration in the course of the error reproduction session. In combination with the Join-Operator, the defect approximation can be performed along the dependencies of semantic artifacts.

study is provided of higher education students from three countries working together using a wiki to construct knowledge about teaching and learning. The authors present these two frameworks in order to increase understanding of the knowledge and skills needed by students in higher education to engage with the affordances of collective intelligence systems.

**Chapter 11**



Natural Language Processing (NLP) provides tools to extract explicitly stated information from text documents. These tools include Named Entity Recognition (NER) and Parts-Of-Speech (POS). The extracted information represents discrete entities in the text and some relationships that may exist among them. To perform intelligent analysis on the extracted information a context has to exist in which this information is placed. The context provides an environment to link information that is extracted from multiple documents and offers a big picture of the domain. Analysis can then be provided by adding inference capabilities to the environment. The ODIX platform provides an environment for bringing together information extraction, ontology, and intelligent analysis. The platform design relies on existing NLP tools to provide the information extraction capabilities. It also utilizes a Web crawler to collect text documents from the Web. The context is provided by a domain ontology that is loaded at run time. The ontology offers limited inference capabilities and external intelligent agents offer more advanced reasoning capabilities. User involvement is key to the success of the analysis process. At every step of the process, the user has the opportunity to direct the system, set selection criteria, correct errors, or add additional information.

**Chapter 12**



The chapter presents a meta-search tool developed in order to deliver search results structured according to the specific interests of users. Meta-search means that for a specific query, several search mechanisms could be simultaneously applied. Using the clustering process, thematically homogenous groups are built up from the initial list provided by the standard search mechanisms. The results are more user-oriented, thanks to the ontological approach of the clustering process. After the initial search made on multiple search engines, the results are pre-processed and transformed into vectors of words. These vectors are mapped into vectors of concepts, by calling an educational ontology and using the WordNet lexical database. The vectors of concepts are refined through concept space graphs and projection mechanisms, before applying the clustering procedure. The chapter describes the proposed solution in the framework of other existent clustering search solutions. Implementation details and early experimentation results are also provided.