

Alessandro Soro, Eloisa Vargiu, Giuliano Armano, and Gavino Paddeu (Eds.)

Information Retrieval and Mining in Distributed Environments

Studies in Computational Intelligence, Volume 324

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

- Vol. 301. Giuliano Armano, Marco de Gemmis, Giovanni Semeraro, and Eloisa Vargiu (Eds.)
Intelligent Information Access, 2010
ISBN 978-3-642-13999-4
- Vol. 302. Bijaya Ketan Panigrahi, Ajith Abraham, and Swagatam Das (Eds.)
Computational Intelligence in Power Engineering, 2010
ISBN 978-3-642-14012-9
- Vol. 303. Joachim Diederich, Cengiz Gunay, and James M. Hogan
Recruitment Learning, 2010
ISBN 978-3-642-14027-3
- Vol. 304. Anthony Finn and Lakhmi C. Jain (Eds.)
Innovations in Defence Support Systems, 2010
ISBN 978-3-642-14083-9
- Vol. 305. Stefania Montani and Lakhmi C. Jain (Eds.)
Successful Case-Based Reasoning Applications-1, 2010
ISBN 978-3-642-14077-8
- Vol. 306. Tru Hoang Cao
Conceptual Graphs and Fuzzy Logic, 2010
ISBN 978-3-642-14086-0
- Vol. 307. Anupam Shukla, Ritu Tiwari, and Rahul Kala
Towards Hybrid and Adaptive Computing, 2010
ISBN 978-3-642-14343-4
- Vol. 308. Roger Nkambou, Jacqueline Bourdeau, and Riichiro Mizoguchi (Eds.)
Advances in Intelligent Tutoring Systems, 2010
ISBN 978-3-642-14362-5
- Vol. 309. Isabelle Bichindaritz, Lakhmi C. Jain, Sachin Vaidya, and Ashlesha Jain (Eds.)
Computational Intelligence in Healthcare 4, 2010
ISBN 978-3-642-14463-9
- Vol. 310. Dipti Srinivasan and Lakhmi C. Jain (Eds.)
Innovations in Multi-Agent Systems and Applications - 1, 2010
ISBN 978-3-642-14434-9
- Vol. 311. Juan D. Velásquez and Lakhmi C. Jain (Eds.)
Advanced Techniques in Web Intelligence, 2010
ISBN 978-3-642-14460-8
- Vol. 312. Patricia Melin, Janusz Kacprzyk, and Witold Pedrycz (Eds.)
Soft Computing for Recognition based on Biometrics, 2010
ISBN 978-3-642-15110-1

- Vol. 313. Imre J. Rudas, János Fodor, and Janusz Kacprzyk (Eds.)
Computational Intelligence in Engineering, 2010
ISBN 978-3-642-15219-1
- Vol. 314. Lorenzo Magnani, Walter Carnielli, and Claudio Pizzi (Eds.)
Model-Based Reasoning in Science and Technology, 2010
ISBN 978-3-642-15222-1
- Vol. 315. Mohammad Essaaidi, Michele Malgeri, and Costin Badica (Eds.)
Intelligent Distributed Computing IV, 2010
ISBN 978-3-642-15210-8
- Vol. 316. Philipp Wolfrum
Information Routing, Correspondence Finding, and Object Recognition in the Brain, 2010
ISBN 978-3-642-15233-5
- Vol. 317. Roger Lee (Ed.)
Computer and Information Science 2010
ISBN 978-3-642-15404-1
- Vol. 318. Oscar Castillo, Janusz Kacprzyk, and Witold Pedrycz (Eds.)
Soft Computing for Intelligent Control and Mobile Robotics, 2010
ISBN 978-3-642-15533-8
- Vol. 319. Takayuki Ito, Minjie Zhang, Valentin Robu, Shaheen Fatima, Tokuro Matsuo, and Hirofumi Yamaki (Eds.)
Innovations in Agent-Based Complex Automated Negotiations, 2010
ISBN 978-3-642-15611-3
- Vol. 320. xxxx
- Vol. 321. Dimitri Plemenos and Georgios Miaoulis (Eds.)
Intelligent Computer Graphics 2010
ISBN 978-3-642-15689-2
- Vol. 322. Bruno Baruque and Emilio Corchado (Eds.)
Fusion Methods for Unsupervised Learning Ensembles, 2010
ISBN 978-3-642-16204-6
- Vol. 323. Yingxu Wang, Du Zhang, Witold Kinsner (Eds.)
Advances in Cognitive Informatics, 2010
ISBN 978-3-642-16082-0
- Vol. 324. Alessandro Soro, Eloisa Vargiu, Giuliano Armano, and Gavino Paddeu (Eds.)
Information Retrieval and Mining in Distributed Environments, 2010
ISBN 978-3-642-16088-2

Alessandro Soro, Eloisa Vargiu, Giuliano Armano,
and Gavino Paddeu (Eds.)

Information Retrieval and Mining in Distributed Environments

 Springer

Alessandro Soro

CRS4, Center of Advanced Studies Research
and Development in Sardinia
Parco Scientifico della Sardegna,
Ed. 1 09010 Loc. Piscinamanna,
Pula, (CA) – Italy
E-mail: asoro@crs4.it

Giuliano Armano

Department of Electrical and
Electronic Engineering
University of Cagliari
Piazza d'Armi
09123 Cagliari – Italy
E-mail: armano@diee.unica.it

Eloisa Vargiu

Department of Electrical and
Electronic Engineering
University of Cagliari
Piazza d'Armi
09123 Cagliari – Italy
E-mail: vargiu@diee.unica.it

Gavino Paddeu

CRS4, Center of Advanced Studies Research
and Development in Sardinia
Parco Scientifico della Sardegna,
Ed. 1 09010 Loc. Piscinamanna,
Pula (CA) – Italy
E-mail: gavino@crs4.it

ISBN 978-3-642-16088-2

e-ISBN 978-3-642-16089-9

DOI 10.1007/978-3-642-16089-9

Studies in Computational Intelligence

ISSN 1860-949X

Library of Congress Control Number: 2010936351

© 2010 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The Web is increasingly becoming a vehicle of shared, structured, and heterogeneous contents. Thus one goal of next generation information retrieval tools will be to support personalization, context awareness and seamless access to highly variable data and messages coming both from document repositories and ubiquitous sensors and devices.

This book is partly a collection of research contributions from the DART 2009 workshop, held in Milan (Italy) in conjunction with the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2009) and Intelligent Agent Technology (IAT 2009). Further contributions have been collected and added to the book following a subsequent call for a chapter on the same topics. At DART 2009 practitioners and researchers working on pervasive and intelligent access to web services and distributed information had the opportunity to compare their work and exchange views on such fascinating topics.

Among the several topics addressed, some emerged as the most intriguing. Community oriented tools and techniques form the necessary infrastructure of the Web 2.0. Solutions in this directions are described in Chapters 1-6.

In Chapter 1, *State-of-the-Art in Group Recommendation and New Approaches for Automatic Identification of Groups*, Boratto and Carta present a comprehensive survey on algorithms and systems for group recommendations. Moreover, they propose a novel approach for group recommendation able to adapt to technological constraints (e.g., bandwidth limitations) by automatically identifying groups of users with similar interests, together with a suitable analysis framework and experimental results that support the authors conclusions.

In the following Chapter 2, *Reputation-based Trust Diffusion in Complex Socio-Economic Networks*, Hauke, Pyka, Borschbach, and Heider present a study on the diffusion of reputation-based trust in complex networks. First, they present relevant related work on trust and reputation, as well as their computational adaptation. Then, an outline of complex networks is provided. Finally, they propose a conceptual distributed trust framework, together with

a simulation that shows how reputation information can be made available in complex social networks.

In Chapter 3, *From Unstructured Web Knowledge to Plan Descriptions*, Addis and Borrajo present a solution aimed at bridging the gap between automatic extraction of information from the web and automated planning. To this end, they propose an architecture, called PAA (Plan Acquisition Architecture), that performs plan and action acquisition starting from semi-structured information (i.e., web pages). The corresponding system is presented through an example taken from WikiHow, a well-known collaborative project that provides how-to guidelines.

In Chapter 4, *Semantic Desktop: a Common Gate on Local and Distributed Indexed Resources*, Moulin and Lai describe a Web application designed to organize, share and retrieve documents over the Internet with a desktop-like interaction. They consider communities structured as a network of peers without any centralized support. The proposed solution is based on semantic indexing using concepts of domain ontologies automatically downloaded from the network.

In Chapter 5, *An Agent-Oriented Architecture for Researcher Profiling and Association using Semantic Web Technologies*, Adnan, Tahir, Basharat, and de Cesare describe SEMORA, an architecture that combines agent technologies and Semantic Web in order to acquire information about researchers, so as to enable the retrieval and matching of scored profiles. The overall agent architecture is detailed in the papers, together with use cases.

In Chapter 6, *Integrating Peer-to-Peer and Multi-Agent Technologies for the Realization of Content Sharing Applications*, Poggi and Tomaiuolo describe how the well-known multiagent framework JADE can be extended to take advantage of JXTA networking infrastructure and protocols. To this end, they propose RAIS (Remote Assistant for Information Sharing), a peer-to-peer system that provides a set of advanced services for content sharing and retrieval. In particular, RAIS offers a search power comparable with web search engines, but avoids the burden of publishing the information on the web and ensures controlled and dynamic access to the information. In this context, the adoption of agent technologies simplifies the realization of the main features required by the system.

Chapters 7 and 8 are concerned with the exploitation of agent technology applying it to virtual world scenarios.

In the Chapter *Intelligent Advisor Agents in Distributed Environments*, Augello, Pilato, and Gaglio present a decision support system composed of intelligent conversational agents that play the role of advisors explicitly specialized for the government of a virtual town. After a review of knowledge representation models and agent learning, the authors discuss how their intelligent agents work in distributed environments. The chapter ends illustrating a case study in which a real-world town is simulated.

In the Chapter *Agent-based Search and Retrieval in Virtual World Environments*, Eno, Gauch, and Thompson present an intelligent agent crawler

designed to collect user-generated content in the Second Life and related virtual worlds. In particular, the authors demonstrate that a crawler able to emulate normal user behavior can successfully collect both static and interactive user-created contents.

In Chapter 9, *Contextual Data Management and Retrieval: a Self-organized Approach*, Castelli and Zambonelli discuss the central topic of context aware information retrieval, presenting a self-organizing agent-based approach to autonomously manage distributed contextual data items into sorts of knowledge networks. Services access contextual information via a knowledge network layer, which encapsulates mechanisms and tools to analyze and self-organize contextual information into sorts. A data model is proposed, meant to represent contextual information, together with a suitable programming interface. Experimental results are provided that show an improvement in efficiency with respect to state of the art approaches.

In the next chapter, *A Relational Approach to Sensor Network Data Mining*, Esposito, Di Mauro, Basile, and Ferilli propose a powerful and expressive description language able to represent the spatio-temporal evolution of a sensor network, together with contextual information. Authors extend a previous framework for mining complex patterns expressed in first-order language. They adopt their framework to discover interesting and human-readable patterns by relating spatio-temporal correlations with contextual ones.

Content based information retrieval is the central topic of Chapters 11-14.

In Chapter 11, *Content-based retrieval of distributed multimedia conversational data*, Pallotta discusses in depth multimedia conversational systems, analyzing several real world implementations and providing a framework for their classification along the following dimensions: conversational content, conversational support, information architecture, indexing and retrieval, and usability. Taking earlier research as the starting point, the author shows how the identification of argumentative structure can improve content based search and retrieval on conversational logs.

In the next Chapter, *Multimodal Aggregation and Recommendation Technologies Applied to Informative Content Distribution and Retrieval*, Messina and Montagnuolo also consider multimedia data, presenting a framework for multimodal information fusion. They propose a definition of semantic affinity for heterogeneous information items and a technique for extracting representative elements. Then, they describe a service platform used for aggregating, indexing, retrieving, and browsing news contents taken from different media sources.

In Chapter 13, *Using a network of scalable ontologies for intelligent indexing and retrieval of visual content*, Badii, Lallah, Zhu, and Crouch present the DREAM framework, whose goal is to support indexing, querying and retrieval of video documents based on content, context and search purpose. The overall architecture and usage scenarios are also provided. Usage studies show a good response in terms of accuracy of classifications.

In the next Chapter, *Integrating Sense Discrimination in a Semantic Information Retrieval System*, Basile, Caputo, and Semeraro propose an information retrieval system that integrates sense discrimination to overcome the problem of word ambiguity. The chapter has a dual goal: (i) to evaluate the effectiveness of an information retrieval system based on Semantic Vectors, and (ii) to describe how they have been integrated into a semantic information retrieval framework to build semantic spaces of words and documents. The authors' main motivation for focusing on the evaluation of disambiguation and discrimination systems is that word ambiguity resolution can improve the performance of information retrieval systems.

Finally, in Chapter 15, *Intelligent Information Processing in Smart Grids and Consumption Dynamics*, Simonov, Zich, and Mussetta describe an industrial application of intelligent information retrieval. The authors describe a distributed environment and discuss the application of data mining and knowledge management techniques to the information available in smart grids, outlining their industrial and commercial potential. The concept of digital energy is introduced here and a system for distributed event delivery is described.

We would like to thank all the authors for their excellent contributions and the reviewers for their careful revision and suggestions for improving them. We are grateful to the Springer-Verlag Team for their assistance during preparation of the manuscripts.

We are also indebted to all the participants and scientific committee members of the three editions of the DART workshop, for their continuous encouragement, support and suggestions.

Cagliari (Italy)
May 2010

Alessandro Soro, Eloisa Vargiu
Giuliano Armano, Gavino Paddeu

Contents

| | |
|--|-----|
| State-of-the-Art in Group Recommendation and New Approaches for Automatic Identification of Groups | 1 |
| <i>Ludovico Boratto, Salvatore Carta</i> | |
| Reputation-Based Trust Diffusion in Complex Socio-Economic Networks | 21 |
| <i>Sascha Hauke, Martin Pyka, Markus Borschbach, Dominik Heider</i> | |
| From Unstructured Web Knowledge to Plan Descriptions ... | 41 |
| <i>Andrea Addis, Daniel Borrajo</i> | |
| Semantic Desktop: A Common Gate on Local and Distributed Indexed Resources | 61 |
| <i>Claude Moulin, Cristian Lai</i> | |
| An Agent-Oriented Architecture for Researcher Profiling and Association Using Semantic Web Technologies | 77 |
| <i>Sadaf Adnan, Amal Tahir, Amna Basharat, Sergio de Cesare</i> | |
| Integrating Peer-to-Peer and Multi-agent Technologies for the Realization of Content Sharing Applications | 93 |
| <i>Agostino Poggi, Michele Tomaiuolo</i> | |
| Intelligent Advisor Agents in Distributed Environments | 109 |
| <i>Agnese Augello, Giovanni Pilato, Salvatore Gaglio</i> | |
| Agent-Based Search and Retrieval in Virtual World Environments | 125 |
| <i>Joshua Eno, Susan Gauch, Craig W. Thompson</i> | |
| Contextual Data Management and Retrieval: A Self-organized Approach | 145 |
| <i>Gabriella Castelli, Franco Zambonelli</i> | |

| | |
|---|------------|
| A Relational Approach to Sensor Network Data Mining | 163 |
| <i>Floriana Esposito, Teresa M.A. Basile, Nicola Di Mauro, Stefano Ferilli</i> | |
| Content-Based Retrieval of Distributed Multimedia Conversational Data | 183 |
| <i>Vincenzo Pallotta</i> | |
| Multimodal Aggregation and Recommendation Technologies Applied to Informative Content Distribution and Retrieval | 213 |
| <i>Alberto Messina, Maurizio Montagnuolo</i> | |
| Using a Network of Scalable Ontologies for Intelligent Indexing and Retrieval of Visual Content | 233 |
| <i>Atta Badii, Chattun Lallah, Meng Zhu, Michael Crouch</i> | |
| Integrating Sense Discrimination in a Semantic Information Retrieval System | 249 |
| <i>Pierpaolo Basile, Annalina Caputo, Giovanni Semeraro</i> | |
| Information Processing in Smart Grids and Consumption Dynamics | 267 |
| <i>Mikhail Simonov, Riccardo Zich, Marco Mussetta</i> | |
| Author Index | 287 |