

Giuliano Armano, Marco de Gemmis, Giovanni Semeraro, and Eloisa Vargiu (Eds.)

Intelligent Information Access

Studies in Computational Intelligence, Volume 301

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. 279. Manolis Wallace, Ioannis E. Anagnostopoulos,
Phivos Mylonas, and Maria Bielikova (Eds.)
Semantics in Adaptive and Personalized Services, 2010
ISBN 978-3-642-11683-4

Vol. 280. Chang Wen Chen, Zhu Li, and Shiguo Lian (Eds.)
*Intelligent Multimedia Communication: Techniques and
Applications*, 2010
ISBN 978-3-642-11685-8

Vol. 281. Robert Babuska and Frans C.A. Groen (Eds.)
Interactive Collaborative Information Systems, 2010
ISBN 978-3-642-11687-2

Vol. 282. Husrev Taha Sencar, Sergio Velastin,
Nikolaos Nikolaidis, and Shiguo Lian (Eds.)
*Intelligent Multimedia Analysis for Security
Applications*, 2010
ISBN 978-3-642-11754-1

Vol. 283. Ngoc Thanh Nguyen, Radoslaw Katarzyniak, and
Shyi-Ming Chen (Eds.)
Advances in Intelligent Information and Database Systems,
2010
ISBN 978-3-642-12089-3

Vol. 284. Juan R. González, David Alejandro Pelta,
Carlos Cruz, Germán Terrazas, and Natalio Krasnogor (Eds.)
*Nature Inspired Cooperative Strategies for Optimization
(NICSO 2010)*, 2010
ISBN 978-3-642-12537-9

Vol. 285. Roberto Cipolla, Sebastiano Battiato, and
Giovanni Maria Farinella (Eds.)
Computer Vision, 2010
ISBN 978-3-642-12847-9

Vol. 286. Zeev Volkovich, Alexander Bolshoy, Valery Kirzhner,
and Zeev Barzily
Genome Clustering, 2010
ISBN 978-3-642-12951-3

Vol. 287. Dan Schonfeld, Caifeng Shan, Dacheng Tao, and
Liang Wang (Eds.)
Video Search and Mining, 2010
ISBN 978-3-642-12899-8

Vol. 288. I-Hsien Ting, Hui-Ju Wu, Tien-Hwa Ho (Eds.)
Mining and Analyzing Social Networks, 2010
ISBN 978-3-642-13421-0

Vol. 289. Anne Hökansson, Ronald Hartung, and
Ngoc Thanh Nguyen (Eds.)
*Agent and Multi-agent Technology for Internet and
Enterprise Systems*, 2010
ISBN 978-3-642-13525-5

Vol. 290. Weiliang Xu and John Bronlund
Mastication Robots, 2010
ISBN 978-3-540-93902-3

Vol. 291. Shimon Whiteson
Adaptive Representations for Reinforcement Learning, 2010
ISBN 978-3-642-13931-4

Vol. 292. Fabrice Guillet, Gilbert Ritschard,
Henri Briand, Djamel A. Zighed (Eds.)
Advances in Knowledge Discovery and Management, 2010
ISBN 978-3-642-00579-4

Vol. 293. Anthony Brabazon, Michael O'Neill, and
Dietmar Maringer (Eds.)
Natural Computing in Computational Finance, 2010
ISBN 978-3-642-13949-9

Vol. 294. Manuel F.M. Barros, Jorge M.C. Guilherme, and
Nuno C.G. Horta
*Analog Circuits and Systems Optimization based on
Evolutionary Computation Techniques*, 2010
ISBN 978-3-642-12345-0

Vol. 295. Roger Lee (Ed.)
*Software Engineering, Artificial Intelligence, Networking and
Parallel/Distributed Computing*, 2010
ISBN 978-3-642-13264-3

Vol. 296. Roger Lee (Ed.)
*Software Engineering Research, Management and
Applications*, 2010
ISBN 978-3-642-13272-8

Vol. 297. Tania Tronco (Ed.)
New Network Architectures, 2010
ISBN 978-3-642-13246-9

Vol. 298. Adam Wierzbicki
Trust and Fairness in Open, Distributed Systems, 2010
ISBN 978-3-642-13450-0

Vol. 299. Vassil Sgurev, Mincho Hadjiski, and
Janusz Kacprzyk (Eds.)
Intelligent Systems: From Theory to Practice, 2010
ISBN 978-3-642-13427-2

Vol. 300. Baoding Liu (Ed.)
Uncertainty Theory, 2010
ISBN 978-3-642-13958-1

Vol. 301. Giuliano Armano, Marco de Gemmis,
Giovanni Semeraro, and Eloisa Vargiu (Eds.)
Intelligent Information Access, 2010
ISBN 978-3-642-13999-4

Giuliano Armano, Marco de Gemmis,
Giovanni Semeraro, and Eloisa Vargiu (Eds.)

Intelligent Information Access

Giuliano Armano
Department of Electrical and
Electronic Engineering
University of Cagliari
Piazza d'Armi - I09123 Cagliari
Italy
E-mail: armano@diee.unica.it

Giovanni Semeraro
Department of Informatics
University of Bari "Aldo Moro"
Via E. Orabona, 4
70126 - Bari
Italy
E-mail: semeraro@di.uniba.it

Marco de Gemmis
Department of Informatics
University of Bari "Aldo Moro"
Via E. Orabona, 4
70126 - Bari
Italy
E-mail : degemmis@di.uniba.it

Eloisa Vargiu
Department of Electrical and
Electronic Engineering
University of Cagliari
Piazza d'Armi - I09123 Cagliari
Italy
E-mail: vargiu@diee.unica.it

ISBN 978-3-642-13999-4

e-ISBN 978-3-642-14000-6

DOI 10.1007/978-3-642-14000-6

Studies in Computational Intelligence

ISSN 1860-949X

Library of Congress Control Number: 2010929476

© 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

Intelligent Information Access techniques attempt to overcome the limitations of current search devices by providing personalized information items and product/service recommendations. They normally utilize direct or indirect user input and facilitate the information search and decision processes, according to user needs, preferences and usage patterns. Recent developments at the intersection of Information Retrieval, Information Filtering, Machine Learning, User Modelling, Natural Language Processing and Human-Computer Interaction offer novel solutions that empower users to go beyond single-session lookup tasks and that aim at serving the more complex requirement: “Tell me what I don’t know that I need to know”. Information filtering systems, specifically recommender systems, have been revolutionizing the way information seekers find what they want, because they effectively prune large information spaces and help users in selecting items that best meet their needs and preferences. Recommender systems rely strongly on the use of various machine learning tools and algorithms for learning how to rank, or predict user evaluation, of items. Information Retrieval systems, on the other hand, also attempt to address similar filtering and ranking problems for pieces of information such as links, pages, and documents. But they generally focus on the development of global retrieval techniques, often neglecting individual user needs and preferences.

The book aims to investigate current developments and new insights into methods, techniques and technologies for intelligent information access from a multidisciplinary perspective. It comprises six chapters authored by participants in the research event *Intelligent Information Access*, held in Cagliari (Italy) in December 2008.

In Chapter 1, *Enhancing Conversational Access to Information through a Socially Intelligent Agent*, Berardina De Carolis, Irene Mazzotta and Nicole Novielli emphasize the role of Embodied Conversational Agents (ECAs) as a natural interaction metaphor for personalized and context-adapted access to information. They propose a scalable architecture for the development of ECAs able to exhibit an emotional state and/or social signs.

The automatic detection of emotions in text is the problem investigated in Chapter 2, *Annotating and Identifying Emotions in Text*, by Carlo Strapparava and Rada Mihalcea. The authors describe the “Affective Text” task, presented at SEMEVAL- 2007. The task focused on classifying emotions in news headlines, and was intended to explore the connection between emotions and lexical semantics. After illustrating the data set, the rationale of the task and a brief description of the participating systems, several experiments on the automatic annotation of emotions in text are presented. The practical applications of the task are very important. Consider for example opinion mining and market analysis, affective computing, natural language interfaces for e-learning environments or educational games.

Personalization of the ranking computed by search engines and recommender systems is the main topic of Chapter 3, *Improving Ranking by Respecting the Multidimensionality and Uncertainty of User Preferences*, by Bettina Berendt and Veit Koppen. The research question addressed by the authors is whether system ranking is the “right ranking” for the user, based on the context in which she/he operates. A general conceptualization of the ranking-evaluation task is proposed: the comparison between the ranking generated by a computational system, and the “user’s ideal ranking”. Eight challenges to this simple model are discussed, leading to the conclusion that approaches for dealing with multidimensional, and often only partial, preference orders are required and that randomness could be a beneficial feature of system rankings.

In Chapter 4, Hotho reviews the state of the art in the new research area of *data mining on folksonomies*. The first part describes the basics of folksonomies, summarizing del.icio.us, the most popular social bookmarking system, and illustrates in detail BibSonomy, a very successful online service for social bookmarking and publication sharing. Starting from these systems, the author discusses in greater depth the main issues regarding folksonomies, proposing a formal model and presenting their most important network properties. In the second part, the author illustrates three applications: spam detection, ranking and recommendation. Regarding spam detection, the author develops techniques, based on binary classifiers, which prevent spammers from publishing in social bookmarking systems. As far as ranking is concerned, a new algorithm is proposed, namely FolkRank, which takes into account the folksonomy structure for ranking users, tags and resources. For recommendation, the author evaluates a tag recommender based on Collaborative Filtering, a graph based recommender using FolkRank and several simple approaches based on tag counts. In the third part, a possible link between folksonomies and ontologies is suggested, paving the way to some very promising strategies for detecting organizational principles hidden within folksonomies.

Amati, Amodeo, Bianchi, Gaibisso and Gambosi propose, in Chapter 5, *A Uniform Theoretic Approach to Opinion and Information Retrieval*, an application of the Divergence From Randomness (DFR) model to the

opinion finding task, the task of retrieving opinionated blog posts, relevant for a given topic, from a large collection. The opinion finding task can be seen as a search in which, after the standard retrieval of ranked documents, documents are re-ranked according to the presence of opinions within the selected documents. This task can be handled by a supervised or unsupervised method. The authors propose a method for creating a lexicon of opinionated terms for re-ranking the documents, using a supervised algorithm. The first part introduces the statistical basis underpinning the proposed approach and its adoption in opinion retrieval. In particular, two information-theoretic functions are defined, opinion entropy and average opinion entropy. The authors also formally describe their lightweight opinion retrieval algorithm. Lastly, the authors discuss the effectiveness of their approach for creating a dictionary of polarity-bearing terms. They also describe some preliminary experiments and propose alternative ways to approach the polarity detection problem.

In Chapter 6, *A Suite of Semantic Web Tools Supporting Development of Multilingual Ontologies*, Pazienza, Stellato and Turbati propose a suite of software libraries, tools and ontologies to support multilingual development of Semantic Web ontologies. The three tools illustrated in this Chapter are Semantic Turkey, The Linguistic Watermark, and Ontoling. Semantic Turkey is aimed at providing innovative solutions for web browsing and for gathering and organizing the information observed when surfing the net. The novel aspect of Semantic Turkey is its ability to provide a clear separation between acquired data and web links. The Linguistic Watermark is an ontological and software framework for describing and managing heterogeneous linguistic resources and for using their contents for ontological-driven document enrichment. Ontoling is a generic architecture for extending ontology development tools with functionalities for enriching ontological knowledge with linguistic content. The tools presented implicitly embed a new way of rethinking the development of ontologies in terms of making their content reusable and comprehensible. Furthermore, they represent living proof of software engineering principles associated with software reuse, documentation, modularity, interaction analysis, applied to the domain of Knowledge Management Software.

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 the preparation of the manuscripts.

This book is dedicated to the memory of Fiorella de Rosis in recognition of her contribution to user modeling. She was a pioneer in the field of affective computing, a leader in research on modeling emotions and constructing embodied animated agents. She produced key contributions in intelligent user interfaces, in particular on user-adapted generation of natural language and multimedia messages, uncertainty in user models, and presentation of medical explanations and clinical guidelines. During her teaching and research activities she mentored many students who have become established researchers. These research and teaching activities didn't prevent her from being an active

member of the ACM, of the International Society for Research on Emotions, of the European Network of Excellence on Emotions (HUMAINE), of the editorial boards of UMUI and co-chair of many international conferences. All the people acquainted with Fiorella have appreciated her scientific and human value and are grateful for her friendship.

February 2010

Giuliano Armano
Marco de Gemmis
Giovanni Semeraro
Eloisa Vargiu

Contents

Enhancing Conversational Access to Information through a Socially Intelligent Agent	1
<i>Berardina De Carolis, Irene Mazzotta, Nicole Novielli</i>	
Annotating and Identifying Emotions in Text	21
<i>Carlo Strapparava, Rada Mihalcea</i>	
Improving Ranking by Respecting the Multidimensionality and Uncertainty of User Preferences	39
<i>Bettina Berendt, Veit Köppen</i>	
Data Mining on Folksonomies	57
<i>Andreas Hotho</i>	
A Uniform Theoretic Approach to Opinion and Information Retrieval	83
<i>G. Amati, G. Amodeo, M. Bianchi, C. Gaibisso, G. Gambosi</i>	
A Suite of Semantic Web Tools Supporting Development of Multilingual Ontologies	109
<i>Maria Teresa Pazienza, Armando Stellato, Andrea Turbati</i>	
Author Index	137

List of Contributors

Giambattista Amati

Fondazione Ugo Bordoni,
Rome, Italy
gba@fub.it

Giuseppe Amodeo

Dept. of Computer Science,
University of L'Aquila, Italy
gamodeo@fub.it

Bettina Berendt

K.U. Leuven,
Dept. of Computer Science,
Leuven, Belgium
<http://www.cs.kuleuven.be/berendt>

Marco Bianchi

Istituto di Analisi dei Sistemi
ed Informatica
“Antonio Ruberti” - CNR,
Rome, Italy
surname@iasi.cnr.it

Berardina De Carolis

Dipartimento di Informatica,
Università degli Studi di Bari,
Via Orabona,
4 - 70125 - Bari, Italy
decarolis@di.uniba.it

Carlo Gaibisso

Istituto di Analisi dei
Sistemi ed Informatica
“Antonio Ruberti” - CNR,
Rome, Italy
surname@iasi.cnr.it

Giorgio Gambosi

Dept. of Mathematics,
University of Rome
“Tor Vergata”, Italy,
gambosi@mat.uniroma2.it

Andreas Hotho

Knowledge & Data
Engineering Group,
University of Kassel,
34121 Kassel, Germany
hotho@cs.uni-kassel.de

Veit Köppen

Otto-von-Guericke-Universität
Magdeburg,
Dept. of Technical & Business
Information Systems,
Magdeburg, Germany
<http://www.veit-koeppen.de/>

Irene Mazzotta

Dipartimento di Informatica,
Università degli Studi di Bari,

Via Orabona,
4 - 70125 - Bari, Italy
mazzotta@di.uniba.it

Rada Mihalcea
University of North Texas
rada@cs.unt.edu

Nicole Novielli
Dipartimento di Informatica,
Università degli Studi di Bari,
Via Orabona,
4 - 70125 - Bari,
Italy
novielli@di.uniba.it

Maria Teresa Pazienza
ART Group,
Dept. of Computer Science,
Systems and Production
University of Rome,
Tor Vergata
Via del Politecnico 1,

00133 Rome, Italy
pazienza@info.uniroma2.it

Armando Stellato
ART Group,
Dept. of Computer Science,
Systems and Production
University of Rome,
Tor Vergata
Via del Politecnico 1,
00133 Rome, Italy
stellato@info.uniroma2.it

Carlo Strapparava
FBK-IRST
strappa@fbk.edu

Andrea Turbati
ART Group,
Dept. of Computer Science,
Systems and Production
University of Rome,
Tor Vergata
Via del Politecnico 1,
00133 Rome, Italy
turbati@info.uniroma2.it