Vincenzo Pallotta, Alessandro Soro, and Eloisa Vargiu (Eds.)

Advances in Distributed Agent-Based Retrieval Tools

Studies in Computational Intelligence, Volume 361

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. 342. Federico Montesino Pouzols, Diego R. Lopez, and Angel Barriga Barros Mining and Control of Network Traffic by Computational Intelligence, 2011 ISBN 978-3-642-18083-5

Vol. 343. Kurosh Madani, António Dourado Correia, Agostinho Rosa, and Joaquim Filipe (Eds.) Computational Intelligence, 2011 ISBN 978-3-642-20205-6

Vol. 344. Atilla Elçi, Mamadou Tadiou Koné, and Mehmet A. Orgun (Eds.) Semantic Agent Systems, 2011 ISBN 978-3-642-18307-2

Vol. 345. Shi Yu, Léon-Charles Tranchevent, Bart De Moor, and Yves Moreau *Kernel-based Data Fusion for Machine Learning*, 2011 ISBN 978-3-642-19405-4

Vol. 346. Weisi Lin, Dacheng Tao, Janusz Kacprzyk, Zhu Li, Ebroul Izquierdo, and Haohong Wang (Eds.) Multimedia Analysis, Processing and Communications, 2011 ISBN 978-3-642-19550-1

Vol. 347. Sven Helmer, Alexandra Poulovassilis, and Fatos Xhafa Reasoning in Event-Based Distributed Systems, 2011 ISBN 978-3-642-19723-9

Vol. 348. Beniamino Murgante, Giuseppe Borruso, and Alessandra Lapucci (Eds.) Geocomputation, Sustainability and Environmental Planning, 2011 ISBN 978-3-642-19732-1

Vol. 349. Vitor R. Carvalho Modeling Intention in Email, 2011 ISBN 978-3-642-19955-4

Vol. 350. Thanasis Daradoumis, Santi Caballé, Angel A. Juan, and Fatos Xhafa (Eds.) Technology-Enhanced Systems and Tools for Collaborative Learning Scaffolding, 2011 ISBN 978-3-642-19813-7 Vol. 351. Ngoc Thanh Nguyen, Bogdan Trawiński, and Jason J. Jung (Eds.) New Challenges for Intelligent Information and Database Systems, 2011 ISBN 978-3-642-19952-3

Vol. 352. Nik Bessis and Fatos Xhafa (Eds.) Next Generation Data Technologies for Collective Computational Intelligence, 2011 ISBN 978-3-642-20343-5

Vol. 353. Igor Aizenberg Complex-Valued Neural Networks with Multi-Valued Neurons, 2011 ISBN 978-3-642-20352-7

Vol. 354. Ljupco Kocarev and Shiguo Lian (Eds.) Chaos-Based Cryptography, 2011 ISBN 978-3-642-20541-5

Vol. 355. Yan Meng and Yaochu Jin (Eds.) Bio-Inspired Self-Organizing Robotic Systems, 2011 ISBN 978-3-642-20759-4

Vol. 356. Slawomir Koziel and Xin-She Yang (Eds.) Computational Optimization, Methods and Algorithms, 2011 ISBN 978-3-642-20858-4

Vol. 357. Nadia Nedjah, Leandro Santos Coelho, Viviana Cocco Mariani, and Luiza de Macedo Mourelle (Eds.) Innovative Computing Methods and Their Applications to Engineering Problems, 2011 ISBN 978-3-642-20957-4

Vol. 358. Norbert Jankowski, Włodzisław Duch, and Krzysztof Grąbczewski (Eds.) *Meta-Learning in Computational Intelligence*, 2011 ISBN 978-3-642-20979-6

Vol. 359. Xin-She Yang and Slawomir Koziel (Eds.) Computational Optimization and Applications in Engineering and Industry, 2011 ISBN 978-3-642-20985-7

Vol. 360. Mikhail Moshkov and Beata Zielosko Combinatorial Machine Learning, 2011 ISBN 978-3-642-20994-9

Vol. 361. Vincenzo Pallotta, Alessandro Soro, and Eloisa Vargiu (Eds.) Advances in Distributed Agent-Based Retrieval Tools, 2011 ISBN 978-3-642-21383-0 Vincenzo Pallotta, Alessandro Soro, and Eloisa Vargiu (Eds.)

Advances in Distributed Agent-Based Retrieval Tools



Editors

Vincenzo Pallotta InterAnalytics Rue des Savoises, 19 1205 Geneva, Switzerland E-mail: vincenzo.pallotta@interanalytics.ch

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

ISBN 978-3-642-21383-0

e-ISBN 978-3-642-21384-7

DOI 10.1007/978-3-642-21384-7

Studies in Computational Intelligence

ISSN 1860-949X

Library of Congress Control Number: 2011931095

© 2011 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

Nowadays, Internet and the Web are not only vehicles of unstructured and heterogeneous contents. In fact, they are rapidly morphing into a platform for interaction and collaboration. In this scenario, the goal of next generation information retrieval tools will be to support semantics, personalization, context awareness and seamless access to highly variable data and messages coming from document repositories, social media and networks, and ubiquitous sensors and devices.

The DART workshop series on Distributed Agent-Based Retrieval Tools was launched in 2006 and reached its 4^{th} edition in 2010. It was originally aimed at putting together practitioners and researchers working on novel retrieval tools for distributed systems and environments. The main goal was to contribute to the discussion about advances in pervasive and intelligent access to Web services and distributed information systems. Over time, the aim of the workshop has been extended to the role of semantics and distributed information retrieval to support content-based multimedia indexing and search.

This book collects revised and extended versions of articles submitted and accepted for presentation to the DART'10 workshop. The 4^{th} edition of DART was held in Geneva at the Webster University international campus, on June 2010.

Social media and collaboration are the topics of the first two chapters.

In Chapter 1, *Rethinking Search Engines in Social Network Vision*, Angioni et al. focus on re-contextualizing the current search engine technology to social networks. In particular, they study the integration of Semantic Web and Natural Language Processing (NLP) technology in a unifying framework. In so doing, user-centric points of views of socially connected users can be used to generate a more effective notion of relevance.

In Chapter 2, A Collaborative Web Application for Supporting Researchers in the Task of Generating Protein Datasets, Armano and Manconi describe a collaborative web application for supporting the generation of protein datasets. In fact, the lack of specific datasets is a major problem for bioinformaticians. ProDaMa-C, the collaborative web application presented in this Chapter, helps bioinformaticians in generating those datasets. Improving search engines is the main topic of the next three chapters.

In Chapter 3, *RefGen: Identifying Reference Chains to Detect Topics*, Longo and Todirascu propose advanced NLP techniques for improving search engine performance through topic indexing. Their approach for topic detection is based on the identification of references chains.

In Chapter 4, Synonym Acquisition Across Domains and Languages, Van der Plas et al. contribute to improve quality of search by proposing an automatic method for acquiring synonyms over different domains and languages. Synonyms can then be used for query expansion for enlarging the search scope by exploiting semantics.

Improving the quality of search is also the goal of Chapter 5, *Linguistically-based Reranking of Googles Snippets with GreG*, in which Delmonte and Tripodi propose a re-ranking approach to standard search engine. The approach is based on the linguistic analysis of unconstrained natural language queries.

The next two chapters are concerned with Sentiment Analysis and Opinion Mining.

In Chapter 6, Opinion Mining and Sentiment Analysis Need Text Understanding, Delmonte and Pallotta advocate for the need of full linguistic processing of input in order to achieve accurate and robust sentiment analysis of product reviews.

At the opposite extreme, in Chapter 7, Sentiment Analysis of French Movie Reviews, Ghorbel and Jacot describe a Machine-Learning approach focused on classifying the polarity (positive, negative) of conveyed opinions. The authors present a supervised classification of French movie reviews based on some shallow linguistic features such as part-of-speech tagging and word semantic orientation.

Distributed information retrieval is tackled in Chapter 8 and 9.

In Chapter 8, *Query Building in a Distributed Semantic Indexing System*, Moulin and Lai propose a query expansion mechanism to deal with terminology variation of semantic description of shared resources in peer-to-peer networks.

In Chapter 9, Building Distributed and Pervasive Information Management Systems with HDS, Bergenti and Poggi present the Heterogeneous Distributed System (HDS) framework built on top of JADE, one of the most established agent framework. The chapter also present the RAIS (Remote Assistant for Information Sharing) application built with HDS whose goal is to support the sharing of information among a community of connected users.

Pervasive intelligence is the main topic of the last two chapters.

In Chapter 10, Sensor Mining for User Behavior Profiling in Intelligent Environments, Augello et al. consider the problem of mining information from sensor networks for user profiling purposes. They report the results of an experiment carried out in office aimed at building a probabilistic model that could be used to predict users behavior and optimize the deployment of building resources.

Pervasive intelligence is also used to build a context-aware recommendation system for exhibition tours planning. In Chapter 11, *Motivating Serendipitous Encounters in Museum Recommendations*, Iaquinta et al. evaluate the impact in users satisfaction of introducing novelty and diversity in personalized museum tours.

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 four editions of the DART workshop, for their continuous encouragement, support and suggestions.

March 2011

Vincenzo Pallotta Alessandro Soro Eloisa Vargiu

Contents

Rethinking Search Engines in Social Network Vision Manuela Angioni, Emanuela De Vita, Cristian Lai, Ivan Marcialis, Gavino Paddeu, Franco G. Tuveri	1
A Collaborative Web Application for Supporting Researchers in the Task of Generating Protein Datasets Giuliano Armano, Andrea Manconi	13
RefGen: Identifying Reference Chains to Detect Topics Laurence Longo, Amalia Todiraşcu	27
Synonym Acquisition across Domains and Languages Lonneke van der Plas, Jörg Tiedemann, Jean-Luc Manguin	41
Linguistically-Based Reranking of Google's Snippets with GreG Rodolfo Delmonte, Rocco Tripodi	59
Opinion Mining and Sentiment Analysis Need Text Understanding Rodolfo Delmonte, Vincenzo Pallotta	81
Sentiment Analysis of French Movie Reviews	97
Query Building in a Distributed Semantic Indexing System Claude Moulin, Cristian Lai	109
Building Distributed and Pervasive Information Management Systems with HDS Federico Bergenti, Agostino Poggi	129

Sensor Mining for User Behavior Profiling in Intelligent Environments	143
Motivating Serendipitous Encounters in Museum Recommendations Leo Iaquinta, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, Piero Molino	159
Author Index	169