

Giambattista Amati, Claudio Carpineto, Giovanni Semeraro (Eds.)

Proceedings of the
Third Italian Information Retrieval Workshop

IIR 2012

Department of Computer Science, University of Bari Aldo Moro, Italy
January 26-27, 2012

<http://www.di.uniba.it/~swap/iir2012>

This volume is published and copyrighted by:

Giambattista Amati

Claudio Carpineto

Giovanni Semeraro

ISSN 1613-0073

Copyright © 2012 for the individual papers by the papers' authors. Copying permitted only for private and academic purposes. Re-publication of material from this volume requires permission by the copyright owners.

In memoriam of Barbara Asta, her 6 year old twins

Salvatore and Giuseppe and her husband Nunzio,

who could not stand the daily torment related to the

memory of Pizzolungo.

Table of Contents

Preface	vi
Organization	vii
Acknowledgements	ix
Invited Speaker	
Semantic is beautiful: clustering and diversifying search results with graph-based Word Sense Induction <i>Roberto Navigli</i>	1
Ranking	
Estensione dei metodi di ranking mediante analisi dell'interspaziatura fra occorrenze <i>Maria C. Daniele, Claudio Carpineto, and Andrea Bernardini</i>	2
Orthogonal negation for document re-ranking <i>Pierpaolo Basile, Annalina Caputo and Giovanni Semeraro</i>	14
Text Classification	
Hierarchical Text Classification for Supporting Educational Programs <i>Qi Ju, Chiara Ravagni, Alessandro Moschitti and Giampiero Vaschetto</i>	18
Error-Correcting Output Codes for Multi-Label Text Categorization <i>Giuliano Armano, Camelia Chira and Nima Hatami</i>	26
How well do we know Bernoulli? <i>Giorgio Maria Di Nunzio and Alessandro Sordani</i>	38
Investigating the Use of Extractive Summarisation in Sentiment Classification <i>Marco Bonzanini, Miguel Martinez-Alvarez and Thomas Roelleke</i>	45
Evaluation & Geographic IR	
Sull'uso di meno topics nelle iniziative di valutazione per l'information retrieval <i>Andrea Berto and Stefano Mizzaro</i>	53
Classical vs. Crowdsourcing Surveys for Eliciting Geographic Relevance Criteria <i>Stefano De Sabbata, Omar Alonso and Stefano Mizzaro</i>	65
Flexible Querying in Geo-Finder <i>Gloria Bordogna and Giuseppe Psaila</i>	73
Filtering	
Conversational Query Revision with a Finite User Profiles Model <i>Henry Blanco, Francesco Ricci and Derek Bridge</i>	77

Uncertain Graphs meet Collaborative Filtering	89
<i>Claudio Taranto, Nicola Di Mauro and Floriana Esposito</i>	
Movie Recommendation with DBpedia	
<i>Roberto Mirizzi, Tommaso Di Noia, Azzurra Ragone, Vito Claudio Ostuni and Eugenio Di Sciascio</i>	101
Cold Start Problem: a Lightweight Approach at ECML/PKDD 2011 - Discovery Challenge	
<i>Leo Iaquinta and Giovanni Semeraro</i>	113
Comparing Word Sense Disambiguation and Distributional Models for Cross-Language Information Filtering	
<i>Cataldo Musto, Fedelucio Narducci, Pierpaolo Basile, Pasquale Lops, Marco de Gemmis and Giovanni Semeraro</i>	117
Content Analysis	
Using Snippets in Text Summarization: a Comparative Study and an Application	
<i>Giuliano Armano, Alessandro Giuliani and Eloisa Vargiu</i>	121
Grammatical Feature Engineering for fine-grained IR tasks	
<i>Danilo Croce and Roberto Basili</i>	133
Encoding syntactic dependencies using Random Indexing and Wikipedia as a corpus	
<i>Pierpaolo Basile and Annalina Caputo</i>	144
Algebraic compositional models for semantic similarity in ranking and clustering	
<i>Paolo Annesi, Valerio Storch, Danilo Croce and Roberto Basili</i>	155
Applications	
QuestionCube: a framework for Question Answering	
<i>Piero Molino and Pierpaolo Basile</i>	167
TV-Show Retrieval and Classification	
<i>Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Giovanni Semeraro, Marco de Gemmis, Mauro Barbieri, Jan Korst, Verus Pronk and Ramon Clout</i>	179
Un prototipo per la ricerca di opinioni sui blog dedicati alle trasmissioni televisive d'interesse nazionale	
<i>Giambattista Amati, Marco Bianchi and Giuseppe Marcone</i>	183
Tag clouds and retrieved results: The CloudCredo mobile clustering engine and its evaluation	
<i>Stefano Mizzaro, Luca Sartori and Giacomo Strangolino</i>	191
The Collaboration Potential, an index to assess the roles of scientists in their coauthorship networks	
<i>Francesco Giuliani, Michele Pio De Petris and Giovanni Nico</i>	199
Strategie di classificazione per servizi di search della Pubblica Amministrazione	
<i>Marco Bianchi, Mauro Draoli, Giorgio Gambosi, Alessandro Ligi and Marco Serrago</i>	203

Preface

The purpose of the Italian Information Retrieval (IIR) workshop series is to provide an international meeting forum for stimulating and disseminating research in Information Retrieval and related disciplines, where researchers, especially early stage Italian researchers, can exchange ideas and present results in an informal way.

IIR 2012 took place in Bari, Italy, at the Department of Computer Science, University of Bari Aldo Moro, on January 26-27, 2012, following the first two successful editions in Padua (2010) and Milan (2011).

We received 37 submissions, including full and short original papers with new research results, as well as short papers describing ongoing projects or presenting already published results. Most contributors to IIR 2012 were PhD students and early stage researchers. Each submission was reviewed by at least two members of the Program Committee, and 24 papers were selected on the basis of originality, technical depth, style of presentation, and impact.

The 24 papers published in these proceedings cover six main topics: ranking, text classification, evaluation and geographic information retrieval, filtering, content analysis, and information retrieval applications. Twenty papers are written in English and four in Italian. We also include an abstract of the invited talk given by Roberto Navigli (Department of Computer Science, University of Rome “La Sapienza”), who presented a novel approach to Web search result clustering based on the automated discovery of word senses from raw text.

The Editors of the Conference Proceedings

Giambattista Amati

Fondazione Ugo Bordoni

Claudio Carpineto

Fondazione Ugo Bordoni

Giovanni Semeraro

University of Bari Aldo Moro

Organization

General Chair

- Giovanni Semeraro (University of Bari Aldo Moro)

Program Chairs

- Giambattista Amati (Fondazione Ugo Bordoni)
- Claudio Carpineto (Fondazione Ugo Bordoni)

Steering Committee

- Massimo Melucci (University of Padua)
- Stefano Mizzaro (University of Udine)
- Gabriella Pasi (University of Milano Bicocca)

Program Committee

- Giuseppe Amodeo (Fondazione Ugo Bordoni)
- Roberto Basili (University of Rome “Tor Vergata”)
- Marco Bianchi (Fondazione Ugo Bordoni)
- Gloria Bordogna (IDPA-CNR Dalmine, Bergamo)
- Fabio Crestani (Università della Svizzera Italiana, Lugano)
- Marco de Gemmis (University of Bari Aldo Moro)
- Pasquale De Meo (University of Messina)
- Giorgio Di Nunzio (University of Padua)
- Giorgio Gambosi (University of Rome “Tor Vergata”)
- Antonio Gulli (Search Technology Center, Bing Search, Microsoft)
- Monica Landoni (University of Strathclyde, Glasgow)
- Pasquale Lops (University of Bari Aldo Moro)
- Marco Maggini (University of Siena)
- Massimo Melucci (University of Padua)
- Alessandro Micarelli (University of Roma Tre)
- Stefano Mizzaro (University of Udine)
- Alessandro Moschitti (University of Trento)
- Roberto Navigli (University of Rome “La Sapienza”)
- Salvatore Orlando (University of Venice “Ca’ Foscari”)
- Gabriella Pasi (University of Milano Bicocca)
- Raffaele Perego (ISTI-CNR, Pisa)
- Francesco Ricci (Free University of Bozen-Bolzano)
- Fabrizio Sebastiani (ISTI-CNR, Pisa)
- Fabrizio Silvestri (ISTI-CNR, Pisa)

Organizing Committee

- Pierpaolo Basile (University of Bari Aldo Moro)
- Annalina Caputo (University of Bari Aldo Moro)
- Marco de Gemmis (University of Bari Aldo Moro)
- Leo Iaquinta (University of Bari Aldo Moro)
- Pasquale Lops (University of Bari Aldo Moro)
- Cataldo Musto (University of Bari Aldo Moro)
- Fedelucio Narducci (University of Bari Aldo Moro)

Acknowledgments

The workshop was supported by:

- Department of Computer Science of the University of Bari Aldo Moro
<http://www.di.uniba.it>
- Distretto Produttivo dell'Informatica della Regione Puglia
<http://www.distrettoinformatica.it>
- Informatici senza Frontiere
<http://www.informaticisenzafrontiere.org>



and was sponsored by:



www.ethicasystem.com



www.exprivia.it



www.fub.it



www.linksmt.it



www.murexcs.it



www.nealogic.it



www.openworkbpm.com



www.questioncube.com



www.sudsistemi.it