

Roberto Basili, Fabrizio Sebastiani, Giovanni Semeraro (Eds.)

Proceedings of the  
Fourth Italian Information Retrieval Workshop

**IIR 2013**

National Council of Research campus, Pisa, Italy  
16 - 17 January 2013

<http://iir2013.isti.cnr.it/>

This volume is published and copyrighted by:

Roberto Basili

Fabrizio Sebastiani

Giovanni Semeraro

ISSN 1613-0073

Copyright © 2013 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.

# Table of Contents

<b>Preface</b> .....	v
<b>Organization</b> .....	vi
<b>THEORY</b>	
<b>Are There New BM25 Expectations?</b> <i>Emanuele Di Buccio, Giorgio Maria Di Nunzio</i> .....	1
<b>The Bivariate 2-Poisson Model for IR</b> <i>Giambattista Amati, Giorgio Gambosi</i> .....	13
<b>QUERY LANGUAGES &amp; OPERATIONS</b>	
<b>A Query Expansion Method based on a Weighted Word Pairs Approach</b> <i>Luca Greco, Massimo De Santo, Paolo Napoletano, Francesco Colace</i> .....	17
<b>A Flexible Extension of XQuery Full-Text</b> <i>Emanuele Panzeri, Gabriella Pasi</i> .....	29
<b>Towards a Qualitative Analysis of Diff Algorithms</b> <i>Gioele Barabucci, Paolo Ciancarini, Angelo Di Iorio, Fabio Vitali</i> .....	33
<b>On Suggesting Entities as Web Search Queries</b> <i>Diego Ceccarelli, Sergiu Gordea, Claudio Lucchese, Franco Maria Nardini, Raffaele Perego</i>	37
<b>IMAGE RETRIEVAL</b>	
<b>Visual Features Selection</b> <i>Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro</i> .....	41
<b>Experimenting a Visual Attention Model in the Context of CBIR Systems</b> <i>Franco Alberto Cardillo, Giuseppe Amato, Fabrizio Falchi</i> .....	45
<b>EVALUATION</b>	
<b>Cumulated Relative Position: A Metric for Ranking Evaluation</b> <i>Marco Angelini, Nicola Ferro, Kalervo Järvelin, Heikki Keskustalo, Ari Pirkola, Giuseppe Santucci, Gianmaria Silvello</i> .....	57
<b>Visual Interactive Failure Analysis: Supporting Users in Information Retrieval Evaluation</b> <i>Marco Angelini, Nicola Ferro, Giuseppe Santucci, Gianmaria Silvello</i> .....	61
<b>SOCIAL MEDIA AND INFORMATION RETRIEVAL</b>	
<b>Myusic: a Content-based Music Recommender System based on eVSM and Social Media</b> <i>Cataldo Musto, Fedelucio Narducci, Giovanni Semeraro, Pasquale Lops, Marco de Gemmis</i> .....	65
<b>A Preliminary Study on a Recommender System for the Million Songs Dataset Challenge</b> <i>Fabio Aioli</i> .....	73

**Distributional Models vs. Linked Data: Exploiting Crowdsourcing to Personalize Music Playlists**  
*Cataldo Musto, Fedelucio Narducci, Giovanni Semeraro, Pasquale Lops, Marco de Gemmis* ..... 84

**SEMANTICS, NATURAL LANGUAGE AND APPLICATIONS**

**Opinion and Factivity Analysis of Italian Political Discourse**  
*Rodolfo Delmonte, Rocco Tripodi, Daniela Gifu* ..... 88

**Distributional Semantics for Answer Re-ranking in Question Answering**  
*Piero Molino, Pierpaolo Basile, Annalina Caputo, Pasquale Lops, Giovanni Semeraro* ..... 100

**INSEARCH: A platform for Enterprise Semantic Search**  
*Diego De Cao, Valerio Storch, Danilo Croce, Roberto Basili* ..... 104

**Wikipedia-based Unsupervised Query Classification**  
*Milen Kouylekov, Luca Dini, Alessio Bosca, Marco Trevisan* ..... 116

## Preface

The purpose of the Italian Information Retrieval (IIR) workshop series is to provide a forum for stimulating and disseminating research in information retrieval, where Italian researchers (especially young ones) and researchers affiliated with Italian institutions can network and discuss their research results in an informal way. IIR 2013 took place in Pisa, Italy, at the National Council of Research campus on January 16-17, 2013, following the first three successful editions in Padua (2010), Milan (2011) and Bari (2012).

The contributions to IIR 2013 mainly address six relevant topics:

- theory
- query languages and operations
- image retrieval
- evaluation
- social media and information retrieval
- semantics, natural language and applications

Most submitted papers were from PhD students and early stage researchers. All the 24 submissions, both full and short original papers presenting new research results, as well as extended abstracts containing descriptions of ongoing projects or presenting already published results, were reviewed by two members of the Program Committee and 18 contributions were selected for presentation on the basis of originality, technical depth, style of presentation, and impact. Additionally to the presentations of these 18 submitted papers, IIR 2013 featured two special events. The first was an invited talk by Renato Soru, CEO of Tiscali SpA, in which the speaker addressed past, present, and future efforts by Tiscali to enter the Web search market. In particular, Soru highlighted some new features of “istella”, the soon-to-be-announced Web search engine by Tiscali, mainly addressed at covering the Italian Web space, with a special emphasis on making Italy’s cultural heritage digitally available to a wide audience. The second special event was a panel on EVALITA, an evaluation campaign which has been running biennially since 2007 and whose main goal is the evaluation of natural language processing tools for Italian. Several EVALITA task organizers have presented the main results obtained in the recent editions of the campaign and have discussed the unresolved challenges that still lie ahead of researchers, with the aim of generating awareness about the state-of-the-art in Italian NLP among IR researchers and of strengthening the relationships between the two communities.

The present proceedings include the papers that were presented at IIR 2013. We hope they represent an interesting contribution to IR research in Italy, and to IR research in general.

### The Workshop Organisers

Roberto Basili  
*University of Roma “Tor Vergata” (Program co-Chair)*

Fabrizio Sebastiani  
*ISTI-CNR (General Chair)*

Giovanni Semeraro  
*University of Bari Aldo Moro (Program co-Chair)*

## **Organization**

### **General Chair**

Fabrizio Sebastiani (ISTI-CNR)

### **Program Chairs**

Roberto Basili (University of Rome “Tor Vergata”)

Giovanni Semeraro (University of Bari Aldo Moro)

### **IIR Steering Committee**

Gianni Amati (Fondazione Ugo Bordoni)

Claudio Carpineto (Fondazione Ugo Bordoni)

Massimo Melucci (University of Padua)

Stefano Mizzaro (University of Udine)

Gabriella Pasi (University of Milano Bicocca)

Giovanni Semeraro (University of Bari Aldo Moro)

### **Program Committee**

Giambattista Amati (Fondazione Ugo Bordoni)

Giuseppe Amodeo (Almawave srl)

Pierpaolo Basile (University of Bari Aldo Moro)

Giacomo Berardi (ISTI-CNR, Pisa)

Gloria Bordogna (IDPA-CNR Dalmine, Bergamo)

Claudio Carpineto (Fondazione Ugo Bordoni)

Fabio Crestani (Università della Svizzera Italiana)

Danilo Croce (University of Rome “Tor Vergata”)

Marco de Gemmis (University of Bari Aldo Moro)

Pasquale De Meo (VU University, Amsterdam)

Giorgio Maria Di Nunzio (University of Padua)

Giorgio Gambosi (University of Rome “Tor Vergata”)

Marco Gori (University of Siena)

Antonio Gulli (Microsoft)

Pasquale Lops (University of Bari Aldo Moro)

Marco Maggini (University of Siena)

Massimo Melucci (University of Padua)

Stefano Mizzaro (University of Udine)

Alessandro Moschitti (University of Trento)

Salvatore Orlando (University of Venezia)

Gabriella Pasi (University of Milano Bicocca)

Raffaele Perego (ISTI-CNR, Pisa)

Francesco Ricci (Free University of Bozen-Bolzano)

Fabrizio Silvestri (ISTI-CNR, Pisa)

### **Organizing Committee**

Adriana Lazzaroni, IIT-CNR (Local Arrangements Chair)

Patrizia Andronico, IIT-CNR

Giacomo Berardi, ISTI-CNR (Webmaster)

Catherine Bosio, ISTI-CNR

Raffaella Casarosa, IIT-CNR

Giulio Galesi, ISTI-CNR

### **Additional Reviewers**

Annalina Caputo (University of Bari Aldo Moro)

Piero Molino (University of Bari Aldo Moro)

Fedelucio Narducci (University of Milano Bicocca)