

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Allan Hanbury Andreas Rauber
Arjen P. de Vries (Eds.)

Multidisciplinary Information Retrieval

Second Information Retrieval Facility Conference
IRFC 2011
Vienna, Austria, June 6, 2011
Proceedings

Volume Editors

Allan Hanbury
Information Retrieval Facility
Donau City Str. 1, 1220 Vienna, Austria
E-mail: a.hanbury@ir-facility.org

Andreas Rauber
Vienna University of Technology
Favoritenstr. 9-11/188, 1040 Vienna, Austria
E-mail: rauber@ifs.tuwien.ac.at

Arjen P. de Vries
Centrum Wiskunde and Informatica
Science Park 123, 1098 XG Amsterdam, The Netherlands
E-mail: arjen@acm.org

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21352-6 e-ISBN 978-3-642-21353-3
DOI 10.1007/978-3-642-21353-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.3, I.2.4, H.5, H.4, I.2, C.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

© Springer-Verlag Berlin Heidelberg 2011

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, re-use of illustrations, recitation, broadcasting, reproduction on microfilms 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.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

These proceedings contain the refereed papers presented at the Second Information Retrieval Facility Conference (IRFC), which was held in Vienna on June 6, 2011. The Second IRFC aimed to tackle four complementary research areas:

- Information retrieval
- Semantic Web technologies for IR
- Natural language processing for IR
- Large-scale or distributed computing for the above areas

IRFC 2011 received 19 high-quality submissions, of which 10 were accepted for presentation at the conference and appear here. Three of these fall into the natural language processing for IR area, one into the distributed computing area, while the remaining six are IR papers.

The IRF Conference provides a multi-disciplinary, scientific forum that aims to bring young researchers into contact with industry at an early stage. The link to industry was emphasized in 2011 with the addition of a panel discussion, where industry representatives considered the practical implications of the work presented.

The Information Retrieval Facility (IRF) is an open IR research institution, managed by a scientific board drawn from a panel of international experts in the field. As a non-profit research institution, the IRF provides services to IR science in the form of a reference laboratory and hardware and software infrastructure. Committed to Open Science concepts, the IRF promotes publication of recent scientific results and newly developed methods, both in traditional paper form and as data sets freely available to IRF members. Such transparency ensures objective evaluation and comparability of results and consequently diversity and sustainability of their further development.

The IRF is unique in providing a powerful supercomputing infrastructure that is exclusively dedicated to semantic processing of text. The first standardized dataset made available by the IRF is MAREC, a collection of 19 million patent documents. These data are housed in an environment that allows large-scale scientific experiments on ways to manage and retrieve this knowledge. Subsets of MAREC are used in the CLEF-IP and TREC-CHEM evaluation campaigns, providing researchers with the opportunity to experiment with realistic retrieval tasks on a real-world data corpus. Two of the papers in this volume make use of the CLEF-IP data.

Our sincere thanks go out to:

- The professional team at the IRF and Vienna University of Technology for their help in preparing the conference and this volume: Marie-Pierre Garnier; Katja Mayer; Mihai Lupu; Veronika Stefanov; Matthias Samwald; Linda Andersson

- The IRF executive board: Francisco Eduardo De Sousa Webber, Daniel Schreiber and Sylvia Thal, for their exceptional support
- Bruce Croft for his support as General Chair of the conference
- The IRF scientific board and John Tait (IRF CSO) for their guidance
- Maarten de Rijke, University of Amsterdam, the Netherlands, for accepting to give a keynote
- The panelists
- The members of the Program Committee and the additional reviewers for their thorough reviews
- The sponsors: Google; CEPIS-EIRSG, the European Information Retrieval Specialist Group
- BCS — The Chartered Institute for IT and the Association for Computing Machinery (ACM) for endorsing the conference
- The Vienna University of Technology and the Austrian Computer Society for their organizational assistance

We hope that you enjoy the results!

June 2011

Allan Hanbury
Andreas Rauber
Arjen P. de Vries

Organization

Program Committee

Maristella Agosti	University of Padua, Italy
Yannis Avrithis	National Technical University of Athens, Greece
Christos Bouganis	Imperial College London, UK
Pavel Braslavski	Yandex, Russia
Jamie Callan	Carnegie Mellon University, USA
Kilnam Chon	Keio University, Japan
Philipp Cimiano	University of Bielefeld, Germany
Paul Clough	University of Sheffield, UK
Fabio Crestani	University of Lugano, Italy
Hamish Cunningham	University of Sheffield, UK
Gideon Dror	Yahoo! Research
Norbert Fuhr	University of Duisburg-Essen, Germany
Wilfried Gansterer	University of Vienna, Austria
Gregory Grefenstette	Exalead, France
David Hawking	Funnelback, Australia
Noriko Kando	National Institute of Informatics, Japan
Gabriella Kazai	Microsoft Research, UK
Udo Kruschwitz	University of Essex, UK
Marie-Francine Moens	Katholieke Universiteit Leuven, Belgium
Henning Müller	University of Applied Sciences Western Switzerland
Patrick Ruch	University of Applied Sciences Geneva, Switzerland
Stefan Rueger	The Open University, UK
Frank Seinstra	VU University, Amsterdam, The Netherlands
Oren Somekh	Yahoo! Research, Israel
John Tait	Information Retrieval Facility, Austria
Dolf Trieschnigg	University of Twente, The Netherlands
Howard Turtle	Syracuse University, USA
Keith Van Rijsbergen	University of Glasgow, UK
Suzan Verberne	Radboud University Nijmegen, The Netherlands
Thijs Westerveld	Teezir Search Solutions, The Netherlands

Additional Reviewers

Niels Drost	Jacopo Urbani
Jaap Kamps	Timo Van Kessel
Parvaz Mahdabi	

Sponsored by



CEPIS:EIRSG
European Information
Retrieval Specialist Group

In cooperation with



Endorsed by



Table of Contents

Keynote

- Adapting Rankers Online 1
Katja Hofmann, Shimon Whiteson, and Maarten de Rijke

Patents and Multilinguality

- Building Queries for Prior-Art Search 3
*Parvaz Mahdabi, Mostafa Keikha, Shima Gerani,
Monica Landoni, and Fabio Crestani*
- Expanding Queries with Term and Phrase Translations in Patent
Retrieval 16
Charles Jochim, Christina Lioma, and Hinrich Schütze
- Supporting Arabic Cross-Lingual Retrieval Using Contextual
Information 30
Farag Ahmed, Andreas Nürnberger, and Marcus Nitsche

Interactive Retrieval Support

- Combining Interaction and Content for Feedback-Based Ranking 46
Emanuele Di Buccio, Massimo Melucci, and Dawei Song
- Query Expansion for Language Modeling Using Sentence Similarities ... 62
Debasis Ganguly, Johannes Leveling, and Gareth J.F. Jones
- Word Clouds of Multiple Search Results 78
Rianne Kaptein and Jaap Kamps

IR and the Net

- Free-Text Search over Complex Web Forms 94
Kien Tjin-Kam-Jet, Dolf Trieschnigg, and Djoerd Hiemstra
- Multilingual Document Clustering Using Wikipedia as External
Knowledge 108
Kiran Kumar N., Santosh G.S.K., and Vasudeva Varma

Applying Web Usage Mining for Adaptive Intranet Navigation	118
<i>Sharhida Zawani Saad and Udo Kruschwitz</i>	
Search Result Caching in Peer-to-Peer Information Retrieval Networks	134
<i>Almer S. Tigelaar, Djoerd Hiemstra, and Dolf Trieschnigg</i>	
Author Index	149