Lecture Notes in Computer Science 6107

Hamish Cunningham Allan Hanbury
Stefan Rüger (Eds.)

Advances in Multidisciplinary Retrieval

First Information Retrieval Facility Conference, IRFC 2010
Vienna, Austria, May 31, 2010
Proceedings

Springer

Volume Editors

Hamish Cunningham
University of Sheffield
Dept. of Computer Science
Sheffield S1 4DP, UK
E-mail: hamishagain@googlemail.com

Allan Hanbury
Information Retrieval Facility
1040 Vienna, Austria
E-mail: a.hanbury@ir-facility.org

Stefan Rüger
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA, UK
E-mail: s.rueger@open.ac.uk

Library of Congress Control Number: 2010926587

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

ISSN 0302-9743
ISBN-10 3-642-13083-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-13083-0 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

These proceedings contain the refereed papers and posters presented at the first Information Retrieval Facility Conference (IRFC), which was held in Vienna on 31 May 2010. The conference provides a multi-disciplinary, scientific forum that aims to bring young researchers into contact with industry at an early stage. IRFC 2010 received 20 high-quality submissions, of which 11 were accepted and appear here. The decision whether a paper was presented orally or as poster was solely based on what we thought was the most suitable form of communication, considering we had only a single day for the event. In particular, the form of presentation bears no relation to the quality of the accepted papers, all of which were thoroughly peer reviewed and had to be endorsed by at least three independent reviewers.

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 whose role is to promote the highest quality in the research supported by the facility. As a non-profit research institution, the IRF provides services to IR science in the form of a reference laboratory, 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. It has at its heart a huge collection of patent documents representing the global archive of ideas and inventions. This data is housed in an environment that allows large-scale scientific experiments on ways to manage and retrieve this knowledge. This collection of real data allows IR researchers from all over the world to experiment for the first time on a realistic data corpus. The quality of search results is reviewed and evaluated in many fields but especially those specialising in patent search.

IRF conferences wish to resonate in particular with young researchers, who are interested in discussing results obtained using the IRF infrastructure and data resources; learning about complementary technologies; applying their research efforts to real business needs; and joining the international research network of the IRF. The first 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

We believe that this first conference has achieved most of these aims and we look forward to many more instances of the IRFC.

Acknowledgements. It is never easy to make a conference happen. Our sincere thanks go out to:

- the IRF executive board: Francisco Eduardo De Sousa Webber, Daniel Schreiber and Sylvia Thal, for their inspiration, for getting the ball rolling and for exceptional organisational talent
- the professional team at the IRF and Matrixware for their help in preparing the conference and this volume: Marie-Pierre Garnier; Katja Mayer; Mihai Lupu; Giovanna Roda; Helmut Berger
- the IRF scientific board and John Tait (IRF CSO) for their guidance
- Niraj Aswani for his help in preparing the proceedings
- the conference programme committee for their hard work reviewing and commenting on the papers:
    - Yannis Avrithis, CERTH, Greece
    - Leif Azzopardi, University of Glasgow, UK
    - Ricardo Baeza-Yates, Yahoo! Research, Spain
    - Jamie Callan, Carnegie Mellon University, USA
    - Paul Clough, University of Sheffield, UK
    - W. Bruce Croft, University of Massachusetts, USA
    - Norbert Fuhr, University Duisburg-Essen, Germany
    - Wilfried Gansterer, University of Vienna, Austria
    - Charles J. Gillan, Queen's University Belfast, UK
    - Gregory Grefenstette, Exalead, France
    - Preben Hansen, Swedish Institute of Computer Science, Sweden
    - David Hawking, Funnelback Internet and Enterprise Search, Australia
    - Joemon Jose, University of Glasgow, UK
    - Noriko Kando, National Institute of Informatics (NII), Japan
    - Philipp Koehn, University of Edinburgh, UK
    - Wessel Kraaij, TNO, The Netherlands
    - Udo Kruschwitz, University of Essex, UK
    - Dominique Maret, Matrixware Information Services, Austria
    - Marie-Francine Moens, Catholic University of Leuven, Belgium
    - Henning Müller, University of Applied Sciences Western Switzerland, Switzerland
    - Walid Najjar, University of California Riverside, USA
    - Arcot Desai Narasimhalu, Singapore Management University, Singapore
    - Fredrik Olsson, Swedish Institute of Computer Science, Sweden
    - Miles Osborne, University of Edinburgh, UK
    - Andreas Rauber, Vienna University of Technology, Austria
    - Magnus Sahlgren, Swedish Institute of Computer Science, Sweden
    - Mark Sanderson, University of Sheffield, UK
    - Frank J. Seinstra, Vrije Universiteit, The Netherlands
    - John Tait, IRF, Austria
    - Benjamin T'sou, City University of Hong Kong, China
    - Christa Womser-Hacker, University of Hildesheim, Germany

– the keynote speakers Mark Sanderson, University of Sheffield, UK, and David Hawking, Funnelback Internet and Enterprise Search, Australia, for providing excellent and inspiring talks
– the sponsors:
  - Matrixware Information Services
  - ESTeam
  - The University of Sheffield
  - STI International
  - Yahoo! Labs
– the Austrian Federal Ministry of Science and Research and Wien Kultur for their support
– BCS — The Chartered Institute for IT for endorsing the conference
– Matt Petrillo for understanding what Hamish was talking about (at least some of the time)

We hope you enjoy the results!

May 2010

Hamish Cunningham, University of Sheffield
http://www.dcs.shef.ac.uk/~hamish
General Chair

Allan Hanbury, Information Retrieval Facility
http://www.ir-facility.org/about/people/staff
Publications Chair

Stefan Rüger, The Open University
http://people.kmi.open.ac.uk/stefan
Programme Chair

Supported by

Endorsed by

Table of Contents