

# Introduction to special issue on the second international conference on the theory of information retrieval

Leif Azzopardi · Dawei Song · Gabriella Kazai · Stephen Robertson ·  
Stefan Ruger · Milad Shokouhi · Emine Yilmaz

Published online: 27 August 2010  
© Springer Science+Business Media, LLC 2010

## 1 Introduction

This special issue of the Journal of Information Retrieval contains selected and substantially extended works presented and published at the second International Conference on the Theory of Information Retrieval (ICTIR), held at Microsoft Research in Cambridge, UK, on 10–11 September 2009.

The ICTIR conference series is a biennial international conference that provides an opportunity for the presentation of the latest work describing advances in the theoretical and formal aspects of Information Retrieval (IR). The conference is run under the auspices of the British Computer Society’s Information Retrieval Specialist Group and the proceedings are now being published in the Lecture Notes in Computer Science (LNCS) series by Springer. The first ICTIR was held in Budapest in October 2007, organized by Sandor

---

L. Azzopardi  
Department of Computing Science, University of Glasgow, Glasgow, Scotland, UK  
e-mail: leif@dcs.gla.ac.uk

D. Song (✉)  
School of Computing, The Robert Gordon University, Aberdeen, Scotland, UK  
e-mail: d.song@rgu.ac.uk

G. Kazai · S. Robertson · M. Shokouhi · E. Yilmaz  
Microsoft Research Cambridge, Cambridge, UK  
e-mail: v-gabkaz@microsoft.com

S. Robertson  
e-mail: ser@microsoft.com

M. Shokouhi  
e-mail: milads@microsoft.com

E. Yilmaz  
e-mail: eminey@microsoft.com

S. Ruger  
Knowledge Media Institute, The Open University, Milton Keynes, UK  
e-mail: s.rueger@open.ac.uk

Dominich, Sándor Darányi, Ferenc Kiss, and Keith van Rijsbergen (Dominich et al. 2007). It was brought about by the growing interest in the consecutive workshops ran at the ACM SIGIR conference each year from 2000 until 2005 on Mathematical and Formal Methods in IR (Athens, Greece, 2000; New Orleans, USA, 2001; Tampere, Finland, 2002; Toronto, Canada, 2003; Sheffield, UK, 2004; Salvador, Brazil, 2005). This sustained initiative was largely down to the determination and passion of the late Sándor Dominich for theoretical and formal models.

In recognition of his contributions, ICTIR 2009 was dedicated to Sándor Dominich. The conference boasted a high quality programme covering a diverse range of topics (Azzopardi et al. 2009). The papers accepted for publication and presentation were selected from a total of 82 submissions, which were received from Continental Europe (39%), UK (21%), North America (18%), Asia and Australasia (10%), Middle East and Africa (12%). The submissions were assessed by at least three expert reviewers in a double blind review process, and were ranked according to their scientific quality, originality, and contribution to the theory of IR. In total, 18 full papers (22%), 14 short papers (17%), and 11 posters (13%) were accepted. The accepted contributions were categorized into four main themes: novel IR models, evaluation, efficiency, and new perspectives in IR. Most of the papers fell into the general theme of novel IR models, ranging from retrieval models, query and term selection models to Web IR models, along with developments in novelty and diversity and user aspect modeling. Papers on new evaluation methodologies were focused on modeling score distributions, evaluation over sessions, and an axiomatic framework for XML retrieval evaluation. Other papers tackled problems of efficiency, offering solutions to improve the tractability of algorithms such as PageRank, or cleansing data before training classifiers, as well as the development of approximate search algorithms for distributed IR. Finally, a number of papers examined new perspectives of IR, such as the application and adoption of quantum theory in IR.

The authors of seven papers, selected based on the review reports and the presentations at the conference, were invited to submit a substantially expanded version to this Journal of Information Retrieval special issue. Each paper underwent two rounds of reviewing by at least three external experts, including at least one reviewer of the original conference submission. Five papers were accepted for inclusion in this special issue. The accepted papers cover the following aspects of IR theory:

- (1) the development and extension of the theory underlying retrieval models,
- (2) the modeling of score distributions (considered by two papers),
- (3) the evaluation of structured retrieval in a theoretical manner, and
- (4) a detailed analysis of evaluation measures used in Novelty and Diversity.

Below we provide a short summary of each of the papers, grouped according to these aspects.

In *Retrieval Constraints and Word Frequency Distribution: A Log-logistic model for IR*, Clinchant and Gaussier, perform a formal analysis on various typical retrieval models using heuristic retrieval constraints. Their work also includes an empirical word distribution analysis, where they make an important observation regarding the “burstiness” of word frequency. They show that the retrieval constraints can be satisfied by an information-based retrieval function when its underlying probability distribution is bursty. Based on this observation and a formal definition of burstiness, the authors propose a new retrieval model based on the log-logistic distribution, which naturally encodes the burstiness of word frequency. Experimental results further validate the effectiveness of the proposed model in comparison with a number of major probabilistic models.

In *Modeling Score Distributions in Information Retrieval*, Arampatzis and Robertson provide a reflective account of the history of modeling score distributions. They focus on analyzing the popular Normal-Exponential mixture model, where relevant documents are characterized by the normal (or Gaussian) distribution and non-relevant documents are characterized by an exponential distribution. In their work, they examine whether hypotheses, such as the Recall-Fallout Convexity Hypothesis, hold, and on what type of retrieval models the Normal-Exponential model of score distributions is suitable/valid. From their analysis they make a number of interesting theoretical contributions to the development of score distribution models. While it is shown that some of the consequences do not have a significant impact upon performance in certain application areas, this work raises a number of empirical and theoretical questions which are worth considering.

In *Variational Bayes for Modeling Score Distributions*, Dai, Kanoulas, Pavlu and Aslam propose modeling the scores of relevant and non-relevant documents using a mixture of Gaussians and Gamma distributions, respectively. Key to their approach is the use of Variational Bayes to estimate the score distributions where the complexity of the model can be balanced against the goodness-of-fit of the models. Through extensive empirical testing, they show that these distributions provide a better fit than previously proposed models (such as the Gaussian-Exponential<sup>1</sup> model which was analyzed by Arampatzis and Robertson). This theoretical development leads to more accurate estimations of performance, however, estimating such models in practice remains a challenging direction for future work.

Blanke and Lalmas, in their paper *Specificity Aboutness in XML Retrieval*, present a theoretical evaluation methodology for XML retrieval models. The work is rooted in various aboutness inference systems, where different aboutness decision making strategies underlying different retrieval models can be captured by different aboutness properties (as a set of essential axioms). These are extended to form a *specificity aboutness* framework for XML retrieval filters. A formal and in-depth theoretical analysis of a number of representative retrieval models (vector space model vs. language model) and filters (brute-force vs. re-ranking) are carried out by showing what aboutness axioms they support, conditionally support and do not support, and how these are correlated to the empirical results in INEX.

In *An Analysis of NP-Completeness in Novelty and Diversity Ranking*, Carterette investigates measures of novelty and diversity by performing a detailed analysis of the problems of computing such measures. In an analytical analysis, Carterette shows that the calculation of novelty and diversity measures is NP-Complete. As a result, approximations are used (i.e. often with a greedy algorithm) but he points out that this will sometimes lead to non-optimal solutions that results in over-estimation or under-estimation of performance. To determine the extent of this problem in practice, he conducts a simulation to show the implications of using such approximations. The paper provides an excellent overview of the different measures along with an extensive investigation into the problems of calculating such measures and warns that such measures may introduce systematic errors into the evaluation process.

Finally, we would like to note that this special issue is intended to highlight only some of the excellent research presented at ICTIR 2009, and we recommend that the interested reader refers to the conference proceedings for a complete overview (Azzopardi et al. 2009). With the foundations of the conference now established, the ICTIR series will provide an ongoing forum for the presentation and discussion of Information Retrieval

---

<sup>1</sup> Referred to as Normal-Exponential.

Theory and its advancement. In 2011, the third ICTIR will be held in Bertinoro, Italy and will be organized by Gianni Amati and Fabio Crestani.

**Acknowledgments** We would like to thank all the reviewers of the ICTIR 2009 conference and the reviewers of this special issue. We would also like to thank all the local organizers, student volunteers, sponsors and participants for their contribution to the success of the ICTIR 2009 conference.

## References

- Azzopardi, L., Kazai, G., Robertson, S. E., Rüger, S. M., Shokouhi, M., Song, D., et al. (Eds.). (2009). *Advances in information retrieval theory, second international conference on the theory of information retrieval, ICTIR 2009, Cambridge, UK, September 10–12, 2009 Proceedings, Lecture Notes in Computer Science*, vol. 5766. Berlin: Springer.
- Dominich S., & Kiss, F. (Eds.) (2007). *Proceedings of the 1st international conference on the theory of information retrieval (ICTIR 2007)*, Alma Mater Series. Infota (Foundation for Information Society, Budapest).