DIR 2011: the eleventh Dutch-Belgian Information Retrieval Workshop

Corrado Boscarino[1]
c.boscarino@cwi.nl

Katja Hofmann[2]
k.hofmann@uva.nl

Valentin Jijkoun[2]
jijkoun@uva.nl

Edgar Meij[2]
edgar.meij@uva.nl

Maarten de Rijke[2]
derijke@uva.nl

Wouter Weerkamp[2]
w.weerkamp@uva.nl

[1] Centrum Wiskunde & Informatica, The Netherlands

[2] University of Amsterdam, The Netherlands

Abstract

The 11th edition of the annual Dutch-Belgian Information Retrieval workshop (DIR 2011) took place on February 4 in Amsterdam. It was organized by the University of Amsterdam and the Centrum Wiskunde & Informatica. The focus of this year's workshop was on interaction, with the goal of facilitating and increasing interaction, especially within the local research community, and between industry and academia. The scientific program included demos, research papers, and compressed contributions. The keynotes by Nick Belkin and Gabriella Kazai provided intriguing outlooks on the future of IR evaluation.

1 Introduction

The Dutch-Belgian Information Retrieval workshop (DIR) is an annual event that promotes exchange between IR researchers in the region. This year's workshop took place on February 4 in Amsterdam. As in previous years, DIR 2011 reflected a growing IR community in the Netherlands and Belgium. With more than 110 attendees, this was the largest DIR to date. Some new groups were also present besides the regional community, including international attendees, participants from industry, and researchers from related fields such as machine learning. DIR 2011 focused particularly on interaction. The organizers reached out to a more varied audience, and created a program that allowed space for interaction around a varied and high-quality scientific program.

Below we highlight some aspects of the program. The full program and the proceedings are available on the workshop's website: `http://ilps.science.uva.nl/dir2011`.

2 Keynotes

The two keynotes provided insights into how IR evaluation could change in the future, in particular with regarsd to human aspects of evaluation.

The first keynote was given by Nick Belkin of Rutgers university. He argued that the evaluation paradigm that focuses on recall and precision—developed with a view of information seeking in library systems more than 50 years ago—is not sufficient for evaluating many of today's retrieval systems. Instead, he argued for "usefulness" as a criterion for IR evaluation. Belkin demonstrated how this measure could be used to evaluate interactive systems, even over complete information seeking episodes.

The second keynote was given by Gabriella Kazai of Microsoft Research Cambridge, and focused on crowdsourcing. Crowdsourcing has become increasingly popular for collecting data where human input is required. Work is split up into small tasks that are completed by workers across the world, who are compensated for their work with small amounts of money per completed task. Crowdsourcing is an interesting way of obtaining human relevance labels for search evaluation quickly and relatively inexpensively. Kazai discussed her recent and ongoing work concerning quality assurance of the data that can be obtained in this way.

3 Research Contributions

An important aspect of DIR is to provide junior researchers with a forum for presenting new research ideas and discussing work in progress. Five original research contributions were accepted.

Carter, Tsagkias, and Weerkamp addressed the problem of language identification in microblogs. They found that state-of-the-art language identifiers show suboptimal performance in this setting, and that performance can be improved using blogger-based and link-based priors.

Dadvar, Hauff, and de Jong addressed the problem of negation detection for analyzing the polarity of movie reviews. They find that traditional negation detection approaches cannot capture implicitly expressed negations that often occur in the studied domain and suggest further study of the lexical features that are used for implicit sentiment expressions.

Hinne, van der Heijden, Verberne, and Kraaij presented a new model for characterizing search intent. The authors suggest a multi-dimensional model that takes into account action, form, and mode of the result, and provide examples of how this classification can be used to analyze a search log.

Lubell-Doughtie and Hofmann study the use of Probabilistic Latent Semantic Analysis for result diversification. They find that interpolating pLSA-based clusters and document relevance can improve result diversity, but that the topics induces by pLSI are often suboptimal.

Tigelaar and Hiemstra study caching as a way to reduce delays in peer-to-peer IR networks. In realistic simulations based on live traffic, the authors find that caching may be an effective way of reducing delays, with an estimated cache hits ratio of 30% or more.

In future editions of DIR, we would like to encourage more junior researchers in the IR community to submit their work.

4 Compressed Contributions

Compressed contributions were first introduced at DIR 2010. These summaries of papers that were published at international conferences in in international journals allow attendants

to become familiar with some of the most interesting IR research that has recently been completed in the area. For DIR 2011, 14 submissions were accepted.

The accepted contributions covered a broad range of recent work in IR that had been previously published in JASIS&T and at SIGIR'10, CIKM'10, ACL'10, ECIR'10, SAC'11, and AWIC'11. Contributions addressed the use of anchor text or tags for retrieval, entity retrieval, sentiment analysis, graph measures for query similarity, annotating web pages, diversification, language models and word clouds, transaction log analysis, analysis of news comments, and retrieval in context.

Particularly intriguing was the paper by Yilmaz, Kanoulas, Robertson, and Aslam, who combined several of their recent publications on constructing collections for learning to rank. Assuming that a fixed amount of resources was available for building a collection, the authors found that obtaining shallow judgments for a large number of queries leads to better test collections, deep judgments for a smaller set of queries. The authors also discussed which method was most appropriate for building such a collection and that a learning algorithm should be optimized for more *informative* measures, such as average precision or graded average precision.

5   Demos

The demo session was particularly successful, with 10 high-quality submissions. Both participants from industry and from research groups presented and explained their most recent developments. The demos included fuzzy faceted search, browsing topic structure for videos, an IR system for children, analysis of Dutch-language news, annotation-based IR for multimedia, natural language search, semantic search for recruiters, associative catalogue search, a term extractor for dutch, and online sentiment monitoring.

6   Conclusion

DIR 2011 was a very successful event, with the largest attendance yet. The high-quality program enabled interaction between researchers from the local and international community, as well as industry and government representatives with an interest in IR. The proceedings of the workshop can be downloaded from the workshop website. The 12th edition of DIR will take place in Ghent, Belgium.

Acknowledgements