

# Report on the 3rd Dutch-Belgian Information Retrieval Workshop (DIR-2002)

Rik De Busser  
Interdisciplinary Centre for Law and IT  
Katholieke Universiteit Leuven  
Leuven, Belgium  
rik.debusser@law.kuleuven.ac.be

## 1 Introduction

In the Low Countries, interest in information retrieval, the discipline that is mainly concerned with identifying information in document or multimedia collections, has been modest but steady throughout the years. In 2000, this led to the first Dutch-Belgian Information Retrieval Workshop (DIR) at the University of Maastricht (the Netherlands). Two years later, the third edition of DIR shows that the IR community in Belgium and the Netherlands is more alive than ever. Organized by ICRI/LIIR (the research group Legal Informatics and IR of the Interdisciplinary Centre for Law and IT, K.U.Leuven, Belgium) in cooperation with SIKS (Dutch Research School for Information and Knowledge Systems) and the IWT Flanders, DIR-2002 received considerable attention from researchers and students from the Netherlands, Belgium, and – notwithstanding the local focus – various other countries. Altogether, more than 80 participants registered for the workshop. The local organizers were Marie-Francine Moens (K.U.Leuven, Belgium), Djoerd Hiemstra (University of Twente, the Netherlands) and Wessel Kraaij (TNO Delft, the Netherlands).

Unlike DIR-2001, which focussed on user interaction and adaptive methods in IR, the organizers of the 2002 workshop had opted for keeping the subject matter as diverse as possible. Fourteen presentations and a keynote speech were crammed into a tight one-day schedule, covering an array of IR subjects as diverse as XML, video retrieval, cross-language IR, and IR from historical corpora. Candidate speakers were invited to submit an extended abstract, which was checked by an international committee of reviewers. We were very happy that all speakers later also submitted a full paper, although this was by no means a prerequisite for the workshop. All papers and an abstract of the keynote speech are bundled into an electronic volume, which is available at [http://www.law.kuleuven.ac.be/icri/proceedings\\_dir.php](http://www.law.kuleuven.ac.be/icri/proceedings_dir.php).

## 2 The presentations

The morning started off with the keynote lecture of Karen Sparck Jones (Cambridge, UK). Unfortunately, some grave circumstances prevented Prof. Sparck Jones to attend the meeting, but she managed to record a vivid speech on videotape – *Language and information: old ideas revisited and new ones considered* – in which she discussed the pros and cons of statistical techniques for language information processing (LIP) in a historical context ([10]). She argued that, although some radical progress has been made in the field of statistical language processing during the previous decade, these developments were to a considerable extent a continuation of ideas that had already been developed in the 1950s and 1960s and that there was still considerable room for innovation, both in purely statistical LIP and for hybrid techniques.

Apart from the keynote speech, the five talks in the forenoon focussed on IR from semi-structured documents and multimedia information retrieval, both of which are issues that currently seem to be very much in the spotlight in the IR world. The first three talks ([9], [7] and [6]) covered the former subject and affirmed the increasing importance of XML for information retrieval; the latter two ([2] and [12]) dealt with content-based image retrieval. During the second session of five talks in the afternoon – after a copious though rather hasty lunch in a historical farmstead – the theme shifted towards the use of natural language processing in IR ([5], [11], [1], [13] and [8]). Roughly, presentations either dealt with evaluating the effectiveness of complementing traditional IR techniques with NLP or with issues concerning lexical cohesion. The last session of the day started off with three talks in the field of cross-lingual IR ([4], [14] and [15]) and the day closed with a highly interesting presentation on information retrieval from historical corpora ([3]).

### 3 Conclusion

All in all, the Third Dutch-Belgian Information Retrieval Workshop was a success. Attendance was a lot higher than was expected, as was the quality of the submissions. Maybe the large number of presentations demanded some perseverance of the attendants, but it also ensured a diverse view on the Dutch and Belgian IR scene. We can only advise all that are interested to keep an eye on the next DIR, which will be organized by the University of Amsterdam (the Netherlands). Further information on DIR-2002 can be found on <http://www.law.kuleuven.ac.be/icri/seminars.php?id=7&where=>.

### 4 Acknowledgements

The organisers would like to thank the sponsors – the Dutch Research School for Information and Knowledge Systems (SIKS) and IWT Flanders – for their kind support. They would also like to express their thanks to the team of reviewers for the effort they put into correcting the submissions.

### References

- [1] Angheluta, R. & M.-F. Moens. A study of synonym replacement in news corpora. *Ibid.*, 58-63.
- [2] Boldareva, L.V.; D. Hiemstra & W. Jonker. A scalable and efficient content-based multimedia retrieval system. *Ibid.*, 28-37.
- [3] Braun, L.; F. Wiesman & I. Sprinkhuizen-Kuyper. Information retrieval from historical corpora. *Ibid.*, 106-12.
- [4] Diekema, A. Spurious matches in Dutch cross-language information retrieval: Lexical ambiguity, vocabulary mismatch, and other causes of translation error. *Ibid.*, 88-95.
- [5] Kamps, J.; C. Monz & M. de Rijke. Combining morphological and n-gram evidence for monolingual document retrieval. *Ibid.*, 48-52.
- [6] Kamps, J.; M. Marx; C. Monz & M. de Rijke. Exploiting structure for information retrieval. *Ibid.*, 20-7.
- [7] Kosala, R.; M. Bruynooghe; J. Van den Busche & H. Blockeel. Information extraction from web pages based on k-testable tree automaton induction. *Ibid.*, 13-9.
- [8] Koster, C.H.A. & M. Seutter. Making phrases work. *Ibid.*, 69-87.
- [9] List, J. & A.P. de Vries. XML-IR: Coverage as part of relevance. *Ibid.*, 7-12.

- [10] Sparck Jones, K. Language and information: old ideas revisited and new ones considered. 5-6. In M.-F. Moens et al. (eds.), *Proceedings of the Third Dutch-Belgian Information Retrieval Workshop (DIR-2002)*. Leuven: ICRI, 2002, 5-6.
- [11] Van Belleghem, K.; A. Vandecandelaere & L. Dehaspe. AlphaDMax: an integrated tool for biomedical information retrieval and extraction. *Ibid.*, 53-7.
- [12] van den Broek, E.; L. Vuurpijl; P. Kisters & J.C.M. von Schmid. Object-based image retrieval: color-selection exploited. *Ibid.*, 38-47.
- [13] van Gils, B. & H. Paijmans. Creating document surrogates with lexical cohesion. *Ibid.*, 64-8.
- [14] Vervenne, D. et al. URUK, a platform for causal text retrieval. *Ibid.*, 96-9.
- [15] Volk, M. & P. Buitelaar. A systematic evaluation of concept-based cross-language information retrieval in the medical domain. 100-5.