

Progress in Information Retrieval

Mounia Lalmas¹, Stefan R uger², Theodora Tsirikla¹, and Alexei Yavlinsky²

¹ Department of Computer Science, Queen Mary, University of London, UK
{mounia, theodora}@dcs.qmul.ac.uk, <http://qmir.dcs.qmul.ac.uk>

² Department of Computing, Imperial College London, UK
{s.rueger, alexei.yavlinsky}@imperial.ac.uk, <http://mmis.doc.ic.ac.uk>

Abstract. This paper summarises the scientific work presented at the 28th European Conference on Information Retrieval and demonstrates that the field has not only significantly progressed over the last year but has also continued to make inroads into areas such as Genomics, Multimedia, Peer-to-Peer and XML retrieval.

Introduction. Information Retrieval is certainly one of those thriving research fields that — despite being relatively old and established — still generates an enormous amount of new interest: the last decade has not only seen eminently successful commercial applications that survived the dot-com bubble but also new challenges such as the hidden and visible Web, enterprise repositories, digital libraries, multimedia, semi-structured documents and new theoretical approaches. It appears that, now more than ever, new and diversified approaches are necessary to stem the tide of information that engulfs us in all shapes and forms. Document collections that increase in size, vary in type and prompt different user needs have driven the information retrieval community to revisit established formal models and create new ones; to look at appropriate document & query representation and increase automated text understanding; to rethink experimental design and evaluation; to improve topic identification and news retrieval; to research user interests and workspaces; to develop specific clustering and classification approaches; to work on query refinement and feedback; to increase performance and design peer-to-peer networks; to contribute to Web search approaches; to create novel XML query approaches; to shift paradigms for multimedia retrieval; to refine methods in cross-language retrieval; and to establish more methods for the new field of genomic Information Retrieval. All this proves that Information Retrieval as a research field continues to push the boundaries of the scientific state of the art and proceeds to positively change the way we all browse, search, select, assess and evaluate, ie, ultimately access and use information.

Formal models. Amati [1] introduces three hypergeometric models, namely KL, DLH and DLLH, using the Divergence from Randomness approach, and compares these models to other relevant models of Information Retrieval. Experiments show that these models have an excellent performance with small and very large collections. Azzopardi and Losada [2] present an efficient implementation of the multiple-Bernoulli language model, which makes it comparable in

speed to traditional term matching algorithms. Fernández *et al.* [3] argue that the performance of score-based aggregation is affected by artificial deviations consistently occurring in the input score distributions. They propose to rectify this by normalising the scores to a common distribution before combination. Wan and Yang [4] propose a novel document similarity measure based on the Proportional Transportation Distance. They show, using the TDT-3 data, that this measure improves on the previously proposed similarity measure based on optimal matching by allowing many-to-many matching between subtopics of documents. Wang *et al.* [5] propose a probabilistic user-item relevance model to re-formulate the problem of implicit acquisition of user preferences for log-based collaborative filtering to perform recommendations. They show that the approach provides a better recommendation performance on a music play-list data set.

Document & query representation and text understanding. Chang and Poon [6] propose a common phrase index as an efficient index structure to support phrase queries in a very large text database. The structure is an extension of previous index structures for phrases and achieves better query efficiency with negligible extra storage cost. Ernst-Gerlach and Fuhr [7] describe a new approach for retrieval in texts with non-standard spelling, which is important for historic texts in English or German. The approach is based on a new algorithm for generating search term variants in ancient orthography and is shown to outperform competing methods. Karbasi and Boughanem [8] develop a method to assess the potential role of the term frequency-inverse document frequency measures commonly used in text retrieval systems. They identify a novel factor, which is shown to be significant for retrieving relevant documents, especially in large collections.

Kane *et al.* [9] use a machine learning approach for classifying document readability based on a simple set of features that attempt to measure the syntactic complexity of text. Meyer zu Eissen and Stein [10] propose a novel plagiarism detection method that identifies potentially plagiarised passages by analysing a single document with respect to variations in writing style; they also identify new features for the quantification of style aspects for this purpose.

Design and evaluation. Clough *et al.* [11] argue that, within the framework of geographic information retrieval, spatial relevance should be considered independently from thematic relevance. They suggest that spatial relevance requires greater assessor effort and more localised geographic knowledge than judging thematic relevance. Demartini and Mizzaro [12] classify 44 different information retrieval evaluation metrics according to the notions of document relevance and of retrieval. Kirsch *et al.* [13] research whether the inclusion of information about a user's social environment and his or her position in the social network of his or her peers leads to an improvement in search effectiveness. Liang *et al.* [14] present a new metric for measuring summary quality based on representativeness and judgeability. They argue that the elements that make up an evaluation methodology are interdependent, and the way in which they are combined is critical to its effectiveness. Della Mea *et al.* [15] carry out a number of text retrieval

experiments using the Average Distance Measure and show that it is highly correlated with traditional effectiveness metrics. Mooney *et al.* [16] carry out a physiological user study that shows that users exhibit galvanic skin response when watching movies engaging in interactive tasks. They examine how these data might be exploited for indexing of data for search and within the search process itself. Vinay *et al.* [17] investigate the problem of some text collections being more difficult to search or more complex to organise into topics than others. Using the Cox-Lewis statistic to measure this complexity, they demonstrate that this analysis is useful in text retrieval. Wen *et al.* [18] investigate the effect of topic familiarity on users' relevance judgements and find that users employ different relevance criteria when searching on less familiar topics.

Topic identification and news retrieval. Parapar and Barreiro [19] present two sentence retrieval methods, Latent Semantic Indexing retrieval and a topic identification method based on Singular Value Decomposition. Experiments on the TREC novelty track data show these techniques as valid alternative approaches to other more ad-hoc methods devised for this task. Smith and Rodríguez [20] present an algorithm for topic detection that considers the temporal evolution of news and the structure of Web documents, the result of which is used for searching and navigating in an online news source. Yao *et al.* [21] present a novel method to identify important news in the Web environment that consists of diversified online news sites. Their method uses a tripartite graph to capture the facts that a piece of important news generally occupies a visually significant place in some homepage of a news site and that important news events will be reported by many news sites.

User interests and workspaces. Bogers and van den Bosch [22] present a novel method of re-ranking search results within closed-community search environments; it utilises information related to topical expertise of workgroup members. Boydell and Smyth [23] describe how snippet-text and title similarities can be used to promote documents without selection history in collaborative retrieval environments. Freschi *et al.* [24] propose a technique for filtering obfuscated spam e-mail using approximate pattern matching performed on the original message and on its phonetic transcription. Tamine-Lechani and Boughanem [25] present a retrieval model based on influence diagrams to incorporate long-term interests of the users into the retrieval process. Vildjiounaite and Kyllönen [26] deploy a Support Vector Machine classifier to learn how to associate user information needs with the contents of their electronic calendar to facilitate proactive information collection and presentation.

Clustering and classification. Bouma and de Rijke [27] investigate the impact on classification accuracy of broadness and narrowness of categories in terms of their distance to the root of a hierarchically organised thesaurus. Carpineto *et al.* [28] present Credino, a clustering engine for PDAs based on the theory of concept lattices that can help overcome some specific challenges posed by small-screen, narrow-band devices. Chakraborti *et al.* [29] adapt Latent Semantic Indexing for document classification by treating class labels as additional terms. Ke *et al.* [30] present an improved procedure for automatically categorising e-mails into user-

defined folders that have few example messages. Naughton *et al.* [31] cluster text spans in a news article that refer to the same event and exploit the order in which events are described for better clustering. Osinski [32] shows how approximate matrix factorisations can be used to organise document summaries returned by a search engine into meaningful thematic categories. SanJuan and Ibekwe-SanJuan [33] present a new method for clustering multi-word terms based on general lexico-syntactic relations that does not require prior domain knowledge or the existence of a training set. Yin and Power [34] present a machine-learning approach for ranking Web documents according to the proportion of procedural text they contain, where “procedural text” refers to ordered lists of steps, which are very common in some instructional genres such as online manuals.

Refinement and feedback. Clinchant *et al.* [35] investigate various lexical entailment models in information retrieval, using the language modelling framework. They show that lexical entailment potentially provides a significant boost in performance, similar to pseudo-relevance feedback, but at a lower computational cost. Keskustalo *et al.* [36] define a user model, which helps to quantify some interaction decisions involved in simulated relevance feedback. They use the model to construct several simulated relevance feedback scenarios in a laboratory setting. Rode and Hiemstra [37] propose a new kind of relevance feedback that shows how so-called query profiles can be employed for disambiguation and clarification. Yamout *et al.* [38] demonstrate a new relevance feedback technique which propagates relevance information of individual documents to unlabelled documents within a given neighbourhood.

Performance and peer-to-peer networks. Büttcher and Clarke [39] present a hybrid approach in which long posting lists are updated in-place, while short lists are updated using a merge strategy. Experimental results show that better indexing performance is obtained with this hybrid approach than either method (in-place, merge-based) alone. Kohlschütter *et al.* [40] introduce a two-dimensional Web model and adapt the PageRank algorithm to efficiently compute an exact rank vector of Web pages, which even for large-scale Web graphs, requires only a few minutes and iteration steps. Nottelmann and Fuhr [41] investigate different building blocks of peer-to-peer architectures, among them the decision-theoretic framework, CORI, hierarchical networks, distributed hash tables and HyperCubes. Siersdorfer and Sizov [42] describe an efficient method to construct reliable machine learning applications in peer-to-peer networks by building ensemble-based meta methods.

Web search. Joho and Jose [43] carry out a comparative evaluation of textual and visual forms of document summaries as an additional document surrogate in the search result presentation. Mishne and de Rijke [44] present an analysis of a large blog search engine query log, exploring a number of angles such as query intent, query topics, and user sessions, and show that blog searches have different intents than general Web searches. Song *et al.* [45] suggest the use of the location of query terms occurring in a URL for measuring how well a Web page is matched with a user’s information need in Web search. This is done through

an estimate of URL hit types, i.e. the prior probability of being a good answer given the type of query term hits in the URL.

Structure/XML. Caracciolo and de Rijke [46] examine multiple query-independent ways of segmenting texts into coherent chunks that can be returned in response to a query. They show this approach to be a viable solution for providing a “go-read-here” functionality. Cornacchia and de Vries [47] use array comprehensions as a novel way to bridge the gap between databases and information retrieval. Schenkel and Theobald [48] present a framework that expands a keyword query into a full-fledged content-and-structure query for relevance feedback in XML retrieval. Extensive experiments on INEX benchmark show the feasibility of the approach. Van Zwol *et al.* [49] present a visual query formulation technique for structured document retrieval that aims at reducing the complexity of the query formulation process and required knowledge of the underlying document structure for the user, while maintaining full expression power, as offered by the NEXI query language for XML retrieval. Vittaut and Gallinari [50] use a model trained to optimise a ranking loss criterion to improve the performance of a baseline structured document retrieval system. The model uses a learning ranking algorithm that operates on scores computed from document elements and from their local structural context. Wang and Rölleke [51] investigate a new, parameter free, ranking method for structured documents based on context-specific inverse document frequency.

Multimedia. Chen *et al.* [52] consider episodic memory, based on time and location, for system design in image retrieval. Their user studies shows that the browser that clusters images based on time and location data combined was significantly better than four other more standard browsers. Demuth *et al.* [53] propose an efficient motion retrieval system based on the query-by-example paradigm. This system employs qualitative, geometric similarity measures, which allows for intuitive and interactive browsing in a purely content-based fashion without relying on textual annotations. Gurrin *et al.* [54] investigate the use of relevance feedback in a text-based video retrieval setting and identify an optimal number of terms for composing new queries based on feedback data. Schedl *et al.* [55] investigate approaches for album cover retrieval that use image search functions of popular search engines and complement them with content analysis. Smeaton *et al.* [56] investigate a novel, object-based modality for video retrieval, where objects appearing in the video are segmented from their background and are used for retrieval based on their low-level visual features. Urban and Jose [57] propose a novel image retrieval system that incorporates a workspace where users can organise their search results. A task-oriented and user-centred experiment shows that the proposed approach leads to a more effective and enjoyable search experience. Wilkins *et al.* [58] automatically determine visual feature weights for content-based image retrieval using a subset of top query results. Zhang *et al.* [59] propose decision fusion and hierarchical classifier approaches to combine short and long term audio features for detecting game highlights in TV tennis videos.

Cross-language retrieval. Alberair and Sanderson [60] study how morphological variants of Arabic queries affect retrieval accuracy. Awadallah and Rauber [61] introduce novel techniques for generating answer choices in a multiple choice question answering setting and evaluate it on English and Arabic question answering data. Hoenkamp and van Dijk [62] use the analogy of fingerprinting as employed in forensics to investigate whether Latent Semantic Analysis, and the hyperspace analog to language are directed towards meaning, and this across languages. Koolen *et al.* [63] propose a cross-language approach to historic document retrieval. In particular, they investigate the automatic construction of translation resources for historic languages and the retrieval of historic documents using cross-language information retrieval techniques. Whittaker *et al.* [64] describe how their statistical pattern classification approach can be used for the rapid development of a Question Answering system in a new language. Zhang *et al.* [65] present a system that automatically collects high quality parallel bilingual corpora from the Web. The proposed system use multiple features to identify parallel texts via a k -nearest-neighbour classifier.

Genomic Information Retrieval. Bernstein and Cameron [66] present an approach based on document fingerprinting for identifying highly similar sequences in large genomic collections. Their approach is shown to use a modest amount of memory and to execute in a time roughly proportional to the size of the collection. Zhou *et al.* [67] focused on addressing the synonym and polysemy issue within the language model framework. A comparative experiment on the TREC 2004 Genomics Track data shows that significant improvements are obtained by incorporating concept-based indexing into a basic language model for this task.

References

1. Giambattista Amati. Frequentist and Bayesian approach to Information Retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 13–24. Springer-Verlag, 2006.
2. Leif Azzopardi and David E. Losada. An efficient computation of the multiple-Bernoulli language model. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 478–481. Springer-Verlag, 2006.
3. Miriam Fernández, David Vallet, and Pablo Castells. Probabilistic score normalization for rank aggregation. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 546–549. Springer-Verlag, 2006.
4. Xiaojun Wan and Jianwu Yang. Using proportional transportation distances for measuring document similarity. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 25–36. Springer-Verlag, 2006.
5. Jun Wang, Arjen P. de Vries, and Marcel J. T. Reinders. A user-item relevance model for log-based collaborative filtering. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 37–48. Springer-Verlag, 2006.

6. Matthew Chang and Chung Keung Poon. Efficient phrase querying with common phrase index. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 61–72. Springer-Verlag, 2006.
7. Andrea Ernst-Gerlach and Norbert Fuhr. Generating search term variants for text collections with historic spellings. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 49–60. Springer-Verlag, 2006.
8. Soheila Karbasi and Mohand Boughanem. Document length normalization using effective level of term frequency in large collections. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 73–84. Springer-Verlag, 2006.
9. Lorna Kane, Joseph Carthy, and John Dunnion. Readability applied to Information Retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 518–521. Springer-Verlag, 2006.
10. Sven Meyer zu Eissen and Benno Stein. Intrinsic plagiarism detection. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 558–561. Springer-Verlag, 2006.
11. Paul D. Clough, Hideo Joho, and Ross Purves. Judging the spatial relevance of documents for GIR. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 542–545. Springer-Verlag, 2006.
12. Gianluca Demartini and Stefano Mizzaro. A classification of IR effectiveness metrics. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 486–489. Springer-Verlag, 2006.
13. Sebastian Marius Kirsch, Melanie Gnasa, and Armin B. Cremers. Beyond the Web: Retrieval in social information spaces. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 85–96. Springer-Verlag, 2006.
14. Shao Fen Liang, Siobhan Devlin, and John Tait. Evaluating Web search result summaries. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 97–108. Springer-Verlag, 2006.
15. Vincenzo della Mea, Gianluca Demartini, Luca di Gaspero, and Stefano Mizzaro. Experiments on average distance measure. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 490–493. Springer-Verlag, 2006.
16. Colum Mooney, Micheál Scully, Gareth J. F. Jones, and Alan F. Smeaton. Investigating biometric response for Information Retrieval applications. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 562–565. Springer-Verlag, 2006.
17. Vishwa Vinay, Ingemar J. Cox, Natasa Milic-Frayling, and Ken Wood. Measuring the complexity of a collection of documents. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 109–120. Springer-Verlag, 2006.
18. Lei Wen, Ian Ruthven, and Pia Borlund. The effects on topic familiarity on online search behaviour and use of relevance criteria. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 454–457. Springer-Verlag, 2006.

19. David Parapar and Álvaro Barreiro. Sentence retrieval with LSI and topic identification. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 121–132. Springer-Verlag, 2006.
20. Simón C. Smith and M. Andrea Rodríguez. Clustering-based searching and navigation in an online news source. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 145–156. Springer-Verlag, 2006.
21. Jinyi Yao, Jue Wang, Zhiwei Li, Mingjing Li, and Wei-Ying Ma. Ranking Web news via homepage visual layout and cross-site voting. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 133–144. Springer-Verlag, 2006.
22. Toine Bogers and Antal van den Bosch. Authoritative re-ranking of search results. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 514–517. Springer-Verlag, 2006.
23. Oisín Boydell and Barry Smyth. Title and snippet based result re-ranking in collaborative Web search. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 482–485. Springer-Verlag, 2006.
24. Valerio Freschi, Andrea Seraghi, and Alessandro Bogliolo. Filtering obfuscated email spam by means of phonetic string matching. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 502–505. Springer-Verlag, 2006.
25. Lynda Tamine-Lechani and Mohand Boughanem. Influence diagrams for contextual Information Retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 462–465. Springer-Verlag, 2006.
26. Elena Vildjiounaite and Kyllönen. Learning links between a user’s calendar and information needs. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 550–553. Springer-Verlag, 2006.
27. Lucas Bouma and Maarten de Rijke. Specificity helps text classification. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 534–537. Springer-Verlag, 2006.
28. Claudio Carpineto, Andrea della Pietra, Stefano Mizzaro, and Giovanni Romano. Mobile clustering engine. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 157–168. Springer-Verlag, 2006.
29. Sutanu Chakraborti, Robert Lothian, Nirmalie Wiratunga, and Stuart Watt. Sprinkling: Supervised Latent Semantic Indexing. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 506–509. Springer-Verlag, 2006.
30. ShihWen Ke, Chris Bowerman, and Michael Oakes. PERC: A personal email classifier. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 458–461. Springer-Verlag, 2006.
31. Martina Naughton, Nicholas Kushmerick, and Joe Carthy. Clustering sentences for discovering events in news articles. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 530–533. Springer-Verlag, 2006.

32. Stanislaw Osinski. Improving quality of search results clustering with approximate matrix factorisations. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 169–180. Springer-Verlag, 2006.
33. Eric SanJuan and Fidelia Ibekwe-SanJuan. Phrase clustering without document context. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 494–497. Springer-Verlag, 2006.
34. Ling Yin and Richard Power. Adapting the naive Bayes classifier to rank procedural texts. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 181–192. Springer-Verlag, 2006.
35. Stéphane Clinchant, Cyril Goutte, and Eric Gaussier. Lexical entailment for Information Retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 216–227. Springer-Verlag, 2006.
36. Heikki Keskustalo, Kalervo Järvelin, and Ari Pirkola. The effects of relevance feedback quality and quantity in interactive relevance feedback: A simulation based on user modeling. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 193–203. Springer-Verlag, 2006.
37. Henning Rode and Djoerd Hiemstra. Using query profiles for clarification. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 204–215. Springer-Verlag, 2006.
38. Fadi Yamout, Michael Oakes, and John Tait. Relevance feedback using weight propagation. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 566–569. Springer-Verlag, 2006.
39. Stefan Büttcher and Charles L. A. Clarke. A hybrid approach to index maintenance in dynamic text retrieval systems. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 228–239. Springer-Verlag, 2006.
40. Christian Kohlschütter, Paul-Alexandru Chirita, and Wolfgang Nejdl. Efficient parallel computation of PageRank. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 240–251. Springer-Verlag, 2006.
41. Henrik Nottelmann and Norbert Fuhr. Comparing different architectures for query routing in peer-to-peer networks. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 252–263. Springer-Verlag, 2006.
42. Stefan Siersdorfer and Sergej Sizov. Automatic document organization in a P2P environment. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 264–275. Springer-Verlag, 2006.
43. Hideo Joho and Joemon M. Jose. A comparative study of the effectiveness of search result presentation on the Web. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 300–311. Springer-Verlag, 2006.
44. Gilad Mishne and Maarten de Rijke. A study of blog search. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 288–299. Springer-Verlag, 2006.

45. Ruihua Song, Guomao Xin, Shuming Shi, Ji-Rong Wen, and Wei-Ying Ma. Exploring URL hit priors for Web search. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 276–287. Springer-Verlag, 2006.
46. Caterina Caracciolo and Maarten de Rijke. Generating and retrieving text segments for focused access to scientific documents. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 348–359. Springer-Verlag, 2006.
47. Roberto Cornacchia and Arjen P. de Vries. A declarative DB-powered approach to IR. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 538–541. Springer-Verlag, 2006.
48. Ralf Schenkel and Martin Theobald. Structural feedback for keyword-based XML retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 324–335. Springer-Verlag, 2006.
49. Roelof van Zwol, Jeroen Baas, Herre van Oostendorp, and Frans Wiering. Bricks: the building blocks to tackle query formulation in structured document retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 312–323. Springer-Verlag, 2006.
50. Jean-Nöel Vittaut and Patrick Gallinari. Machine learning ranking for structured Information Retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 336–347. Springer-Verlag, 2006.
51. Jun Wang and Thomas Rölleke. Context-specific frequencies and discriminativeness for the retrieval of structured documents. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 570–573. Springer-Verlag, 2006.
52. Chufeng Chen, Michael Oakes, and John Tait. Browsing personal images using episodic memory (time + location). In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 360–371. Springer-Verlag, 2006.
53. Bastian Demuth, Tido Röder, Meinard Müller, and Bernhard Eberhardt. An Information Retrieval system for motion capture data. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 372–383. Springer-Verlag, 2006.
54. Cathal Gurrin, Dag Johansen, and Alan F. Smeaton. Supporting relevance feedback in video search. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 554–557. Springer-Verlag, 2006.
55. Markus Schedl, Peter Knees, Tim Pohle, and Gerhard Widmer. Towards automatic retrieval of album covers. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 526–529. Springer-Verlag, 2006.
56. Alan F. Smeaton, Gareth J. F. Jones, Hyowon Lee, Noel E. O’Connor, and Sorin Sav. Object-based access to TV rushes video. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 474–477. Springer-Verlag, 2006.
57. Jana Urban and Joemon M. Jose. Can a workspace help to overcome the query formulation problem in image retrieval? In *Proceedings of the 28th European Confer-*

- ence on *Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 384–395. Springer-Verlag, 2006.
58. Peter Wilkins, Paul Ferguson, Cathal Gurrin, and Alan F. Smeaton. Automatic determination of feature weights for multi-feature CBIR. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 522–525. Springer-Verlag, 2006.
 59. Bin Zhang, Weibei Dou, and Liming Chen. Combining short and long term audio features for TV sports highlight detection. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 470–473. Springer-Verlag, 2006.
 60. Asaad Alberair and Mark Sanderson. Morphological variation of Arabic queries. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 466–469. Springer-Verlag, 2006.
 61. Rawia Awadallah and Andreas Rauber. Web-based multiple choice question answering for English and Arabic questions. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 510–513. Springer-Verlag, 2006.
 62. Eduard Hoenkamp and Sander van Dijk. A fingerprinting technique for evaluating semantics based indexing. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 396–405. Springer-Verlag, 2006.
 63. Marijn Koolen, Frans Adriaans, Jaap Kamps, and Maarten de Rijke. A cross-language approach to historic document retrieval. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 406–417. Springer-Verlag, 2006.
 64. Edward W. D. Whitakker, Julien Hamonic, Dong Yang, Tor Klingberg, and Sadaoki Furui. Rapid development of Web-based monolingual question answering systems. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 498–501. Springer-Verlag, 2006.
 65. Ying Zhang, Ke Wu, Jianfeng Gao, and Phil Vines. Automatic acquisition of Chinese-English parallel corpus from the Web. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 418–429. Springer-Verlag, 2006.
 66. Yaniv Bernstein and Michael Cameron. Fast discovery of similar sequences in large genomic collections. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 430–441. Springer-Verlag, 2006.
 67. Xiaohua Zhou, Xiaodan Zhang, and Xiaohua Hu. Using concept-based indexing to improve language modeling approach to genomic IR. In *Proceedings of the 28th European Conference on Information Retrieval*, volume 3936 of *Lecture Notes in Computer Science*, pages 442–453. Springer-Verlag, 2006.