



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

International conference ACM SIGIR 2011, Beijing, China

Jan Outrata

Jul 24, 2011 – Jul 28, 2011



The 34th Annual Int. ACM SIGIR Conference on Research and Development in Informational Retrieval



- main forum dedicated to representation, retrieving, searching and analysis of information from large electronic sources (mainly internet and WWW)
- main and most cited people from this area of computer science:
 - Ricardo Baeza-Yates** (Yahoo! Research, Spain)
 - Ravi Kumar** (Yahoo! Research, USA)
 - Berthier Ribeiro-Neto** (Federal University of Minas Gerais, Brazil)
- sponsored by ACM and the biggest companies (Microsoft, Google, Yahoo, IBM, HP) and significant organizations (Baidu, eBay, Yandex) in the field
- among participants there are scientists from both academia and research centers of leading companies (Microsoft Research, IBM Research)

ACM SIGIR 2011 – interesting numbers

Acceptance rate:	19.9 %
Number of participants:	over 800
Number of presented contributions:	231 (108 papers, 89 posters, 15 demonstrations, 8 workshops and 11 tutorials)
Percentages of number of papers by countries:	48 % USA, 17 % China, 6 % Germany, 29 % other in summary
Number of plenary contributions:	2
Number of parallel sections:	3

Plenary contributions (1/2)

Qi Lu (Microsoft): Future of the Web and Search.

- presented thoughts and visions of the speaker about directions of present Web and what will be the directions of search technologies for new Web
 - Web presented as reorganized for search based on knowledge of identity, social interactions, location and previous behaviour of users (not only on the web) and understanding what users want
 - presented some specifics of Microsoft, including utilization of information retrieval methods, and calls and possibilities for academic research in this area
- materials from presentations are available at <http://www.sigir2011.org/keynotes.htm>

Plenary contributions (2/2)

ChengXiang Zhai (University of Illinois at Urbana-Champaign): Beyond Search: Statistical Topic Models for Text Analysis.

- utilization of statistical models for analysis of huge amount of information acknowledged from searching and founding relevant information
 - the models included methods of analysis of extraction of themes and their patterns in text (particularly several variants of probabilistic latent semantic analysis)
 - also presented demonstration applications of models in several managerial areas (comparative summarization, contextual topic trend analysis, event impact analysis)
- materials from presentations are available at <http://www.sigir2011.org/keynotes.htm>

Selected regular contributions (1/3)

L. Wang, J. Lin, D. Metzler: A Cascade Ranking Model for Efficient Ranked Retrieval.

Ranking model utilising a cascade of ranking functions presented. The model leads to better decision efficiency of top-k queries as well as search speed. Also the algorithm for learning the model was presented.

H. Ma, Ch. Liu, I. King, M. R. Lyu: Probabilistic Factor Models for Web Site Recommendation.

Probabilistic model for recommending personalized web content presented. The model is based on dimension reduction techniques utilizing real matrix factorization. The model should be linearly scalable.

Ch. Jin, Ch. Yang: Integrating Hierarchical Feature Selection and Classifier Training for Multi-Label Image Annotation.

Two-level algorithm of feature subset selection presented. The algorithm is used for reduction of dimensionality of feature space in high dimensional data like picture representations (the algorithm was demonstrated on such data)

Selected regular contributions (2/3)

S. Rendle, Z. Gantner, Ch. Freudenthaler, L. Schmidt-Thieme: Fast Context-aware Recommendations with Factorization Machines.

Utilization of the so-called factorization machines (which cover matrix factorization or SVD methods) for rating predictions in recommendation systems with linear complexity presented. The systems comprise context information (so-called context-aware rating).

Y. Kim, J. Seo, B. W. Croft: Automatic Boolean Query Suggestion for Professional Search.

The talk was about generating Boolean queries from decision trees created from top-k documents retrieved using basic search, in the field of professional search engines (patent, medicine).

H. Wang, H. Huang, F. Nie, Ch. Ding: Cross-Language Web Page Classification via Dual Knowledge Transfer Using Nonnegative Matrix Tri-Factorization.

Method of classification of multilingual web pages on the same theme presented. The method utilizes a similarity of pages and is based on training from several different sources (so-called transfer learning) and matrix factorization.

Selected regular contributions (3/3)

S. Tatikonda, B. Barla Cambazoglu, F. P. Junqueira: Posting List Intersection on Multicore Architectures.

The talk was about parallelization of intersection of lists of ids of documents containing given expression (so-called posting list), which is the basic operation of search engines. The parallelization is done by decomposing the computation into disjoint independent parts processed in parallel with the use of the finest-grained parallelization on multicore processors.

S. Ding, T. Suel: Faster Top-k Document Retrieval Using Block-Max Indexes.

Improvement of algorithms for the top-k operation presented. The improvement involves new index block structure (block-max index) which allows faster operations with posting lists.

- summaries of other visited presentations are in the Report from conference attendance (20 in total, in Czech)
- abstracts of presentations are available at <http://www.sigir2011.org/papers.asp>

Selected posters and demonstrations (1/2)

Ch. Hoobin, S. Puglisi, J. Zobel: Sample Selection for Dictionary-Based Corpus Compression.

Reduction of vocabular redundancy in an efficient LZ-like compression method presented.

S. Wang, B.J. Gao, K. Wang, H.W. Lauw: Parallel Learning to Rank for Information Retrieval.

Learning of a ranking function which can be used for ordering of results of future queries presented.

M. Huggett, E. Rasmussen: The Meta-Dex Suite: Generating and Analyzing Indexes and Meta-Indexes.

Software for extraction and indexation of text from collection of PDF files and generation of so-called meta-index presented.

Selected posters and demonstrations (2/2)

- the contributions were across the whole spectrum of conference areas, summaries of other which interested me are in the Report from conference attendance (5 in total, in Czech)
- I did not attend workshops and tutorials since my registration did not include them (they had their own additional fees)

Conclusions

- presented influential contributions from areas of: representation, analysis and indexing of data (documents) and their models, classification and clustering, ranking, interactive IR (WWW, communities), query analysis and content search, filtration and, in a large amount, content recommendation (personalization)
- presumed perspective directions of future research are: representation, analysis and indexing of documents, ranking, query analysis and in particular (mainly web) content search and recommendation
- demonstrations of software tools for content search and indexing and processing search results, and of real applications supporting document searching and management from 5 companies (Google, eBay, Facebook, Microsoft, Elsevier)



- Topic analysis on social networks [Mei et al. 08]
- Opinion Integration [Lu & Zhai 08]
- Latent Aspect Rating Analysis [Wang et al. 10]

Qiaozhu Mei, Deng Cai, Duo Zhang, ChengXiang Zhai. Topic Modeling with Network Regularization, Proceedings of the World Wide Conference 2008 (WWW'08), pages 103-110.

Yue Lu, ChengXiang Zhai. Opinion Integration Through Semi-supervised Topic Modeling, Proceedings of the World Wide Conference 2008 (WWW'08), pages 123-130.

Hongning Wang, Yue Lu, ChengXiang Zhai. Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach, Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08), pages 125-134, 2008.

ACM SIGIR 2011
July 24-28, 2011, Beijing, China
The 34th Annual International ACM SIGIR Conference

Microsoft
IBM EMC Google SAP ORACLE Yahoo! eBay HP Labs IBM Intel

- Topic analysis on social networks [Mei et al. 08]
- Opinion Integration [Lu & Zhai 08]
- Latent Aspect Rating Analysis [Wang et al. 10]

Qiaozhu Mei, Deng Cai, Duo Zhang, ChengXiang Zhai. Topic Modeling with Network Regularization, Proceedings of the World Wide Conference 2008 (WWW'08), pages 103-110.

Yue Lu, ChengXiang Zhai. Opinion Integration Through Semi-supervised Topic Modeling, Proceedings of the World Wide Conference 2008 (WWW'08), pages 123-130.

Hongning Wang, Yue Lu, ChengXiang Zhai. Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach, Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08), pages 125-134, 2008.