Efficient Query Processing for Scalable Web Search

Other titles in Foundations and Trends® in Information Retrieval

Web Forum Retrieval and Text Analytics: A Survey
Doris Hoogeveen, Li Wang, Timothy Baldwin and Karin M. Verspoor
ISBN: 978-1-68083-350-8

Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting
Jun Wang, Weinan Zhang and Shuai Yuan
ISBN: 978-1-68083-310-2

Applications of Topic Models
Jordan Boyd-Graber, Yuening Hu and David Mimno
ISBN: 978-1-68083-308-9

Searching the Enterprise
Udo Kruschwitz and Charlie Hull
ISBN: 978-1-68083-304-1

Full text available at: http://dx.doi.org/10.1561/1500000057

Efficient Query Processing for Scalable Web Search

Nicola Tonellotto
National Research Council of Italy
nicola.tonellotto@isti.cnr.it

Craig Macdonald
University of Glasgow
craig.macdonald@glasgow.ac.uk

Iadh Ounis
University of Glasgow
iadh.ounis@glasgow.ac.uk

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

N. Tonellotto, C. Macdonald and I. Ounis. *Efficient Query Processing for Scalable Web Search*. Foundations and Trends® in Information Retrieval, vol. 12, no. 4-5, pp. 319–500, 2018.

ISBN: 978-1-68083-543-4

Foundations and Trends® in Information Retrieval

Volume 12, Issue 4-5, 2018

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2018, Volume 12, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

1 Introduction **6**

2 Modern Infrastructure Foundations **13**
2.1 Data Structures 13
2.2 Query Processing 28
2.3 Summary 43

3 Dynamic Pruning Query Processing **44**
3.1 Introduction to Dynamic Pruning 45
3.2 TAAT Optimisations 49
3.3 DAAT Optimisations 56
3.4 Block-based Dynamic Pruning 69
3.5 Limitations of Dynamic Pruning 80
3.6 Summary 81

4 Query Efficiency Prediction for Dynamic Pruning **82**
4.1 Implementations of Query Efficiency Prediction 84
4.2 Delayed Query Efficiency Prediction 86
4.3 Query Efficiency Prediction Applications 87
4.4 Summary 96

5 Impact-Sorted Indexes **98**
5.1 Data Structures . . . 100
5.2 Query Processing . . . 103
5.3 Summary . . . 108

6 Learning-to-Rank and Cascades **109**
6.1 Candidate Generation for Learning-to-Rank . . . 113
6.2 Feature Calculation in Learning-to-Rank . . . 115
6.3 Application of Learning-to-Rank Models . . . 121
6.4 Summary . . . 135

7 Open Directions and Conclusions **136**
7.1 Summary . . . 136
7.2 Signatures . . . 138
7.3 Energy Efficiency . . . 139
7.4 Modern Hardware Architectures . . . 142
7.5 Clouds and New Software Paradigms . . . 144
7.6 Real-Time and Social Media Search . . . 145

Acknowledgements **147**

References **148**

Efficient Query Processing for Scalable Web Search

Nicola Tonellotto[1], Craig Macdonald[2] and Iadh Ounis[3]

[1] *National Research Council of Italy; nicola.tonellotto@isti.cnr.it*
[2] *University of Glasgow; craig.macdonald@glasgow.ac.uk*
[3] *University of Glasgow; iadh.ounis@glasgow.ac.uk*

ABSTRACT

Search engines are exceptionally important tools for accessing information in today's world. In satisfying the information needs of millions of users, the effectiveness (the quality of the search results) and the efficiency (the speed at which the results are returned to the users) of a search engine are two goals that form a natural trade-off, as techniques that improve the effectiveness of the search engine can also make it less efficient. Meanwhile, search engines continue to rapidly evolve, with larger indexes, more complex retrieval strategies and growing query volumes. Hence, there is a need for the development of efficient query processing infrastructures that make appropriate sacrifices in effectiveness in order to make gains in efficiency. This survey comprehensively reviews the foundations of search engines, from index layouts to basic term-at-a-time (TAAT) and document-at-a-time (DAAT) query processing strategies, while also providing the latest trends in the literature in efficient query processing, including the coherent and systematic reviews of techniques such as dynamic pruning and impact-sorted posting lists as well as their variants and optimisations. Our explanations of query processing strategies, for instance the WAND and

Nicola Tonellotto, Craig Macdonald and Iadh Ounis (2018), "Efficient Query Processing for Scalable Web Search", Foundations and Trends® in Information Retrieval: Vol. 12, No. 4-5, pp 319–500. DOI: 10.1561/1500000057.

BMW dynamic pruning algorithms, are presented with illustrative figures showing how the processing state changes as the algorithms progress. Moreover, acknowledging the recent trends in applying a cascading infrastructure within search systems, this survey describes techniques for efficiently integrating effective learned models, such as those obtained from learning-to-rank techniques. The survey also covers the selective application of query processing techniques, often achieved by predicting the response times of the search engine (known as query efficiency prediction), and making per-query tradeoffs between efficiency and effectiveness to ensure that the required retrieval speed targets can be met. Finally, the survey concludes with a summary of open directions in efficient search infrastructures, namely the use of signatures, real-time, energy-efficient and modern hardware and software architectures.

Acronyms

Here we report the main acronyms used in this survey. Acronyms typeset in Sans-serif pertain directly to information retrieval concepts that we explain in this survey.

NDCG	Normalised Discounted Cumulative Gain
MAP	Mean Average Precision
ERR	Expected Reciprocal Rank
MED	Maximised Effectiveness Difference
RBP	Rank Biased Precision
IR	Information Retrieval
QPS	Queries per second
IDF	Inverse Document Frequency
FOR	Frame-Of-Reference
PFOR	Patched FOR
Vbyte	Variable Byte
EF	Elias-Fano
PEF	Paritioned EF
QMX	Quantities, Multiplier and eXtractor

SIMD	Single Instruction Multiple Data
TAAT	Term-At-A-Time
DAAT	Document-At-A-Time
WAND	Weighted AND or Weak AND
BMW	Block-Max WAND
BMM	Block-Max MaxScore
BMA	Block-Max AND
LBMW	Local BMW
VBMW	Variable BMW
QEP	Query Efficiency Prediction/Predictor
QPP	Query Performance Prediction/Predictor
SLA	Service Level Agreement
PESOS	Predictive Energy Saving Online Scheduling
DVFS	Dynamic Voltage and Frequency Scaling
TFIDF	Term Frequency - Inverse Document Frequency
SAAT	Score-At-A-Time
LTR	learning-to-rank
FPGA	Field Programmable Gate Array
IoT	Internet-of-Things
ISN	index serving node

Notations

Here we only report the recurrent notation symbols used in this survey. `Fixed size` text is used for pseudocode-related symbols.

R	the number of replicas of a shard.
S	the number of shards of an index.
K	the number of top results returned by a search engine.
s	speedup, used as a performance measure.
r	reduction, used as a performance measure.
d	a document, as indexed by an IR system.
q	a query, as processed by an IR system, i.e., a set of terms.
N	the number of documents indexed by the IR system.
t, t_i	a term, as may exist within a query.
$\textsc{Score}_q(d)$	a generic query-document ranking function.
$s_t(q, d)$	a generic term-document similarity function.
f_t	the document frequency of a term.
IDF_t	the inverse document frequency of a term.
$f_{d,t}$	the number of occurrences of a term in a document.
$\perp$	special symbol to denote the end of a posting list.

n	number of terms in a query.
`p`, `I`, `O`	an array of posting lists.
`q`	a priority queue of docids or ⟨`docid`, `score`⟩ pairs.
`A`	an accumulators map from docids to scores.
λ	the size of an unordered window complex operator.
p	parallelism degree, i.e., number of threads.
$t(p)$	expected query processing time with p threads.
$\sigma_t(q)$	the term upper bound, a.k.a. its max score.
$\sigma_d(q)$, `σd`	the document upper bound computed with term upper bounds.
$\hat{q}$	a set of terms from query q already processed.
θ, `θ`	a threshold, i.e., the smallest (partial) score of the current top K documents.
L	parameter of the Quit and Continue strategies.
N_t	the number of documents indexed in a top candidates list.
`σ`	an array of term upper bounds.
`ub`	an array of document upper bounds.
`pivot`	a index of a posting list in `p`.
`pivot_id`	the docid of the `pivot` posting list iterator.
F	the aggressiveness tradeoff of a dynamic pruning strategy.
`b`	an array of block lists.
`σb`	a document upper bound computed with block upper bounds.
$s_j(t)$	a QEP term statistic.
A_i	a QEP aggregation function, e.g., max, sum, variance.
$f_{ij}(q)$	a QEP feature defined for query q.
τ	a posting list score threshold.
$A(t_1, t_2)$	size of the intersection of the posting lists of terms t_1 and t_2.
δ, ϵ, β	small positive constants.
f_{ins}, f_{add}	filtering thresholds.
w	a weight.
b	the number of bits used to represent impacts.
g, h	fitting parameters for estimation of b.
L, U	global lower and upper bounds on term scores.
Q	a fidelity control knob.
ρ	number of postings to process.
M	a standard evaluation metric.
f_i, `f`	a feature id.
$\mathcal{F}$	a feature id set.
`x`	a feature vector.
T_i	a regression/decision tree.

$\mathcal{T}$	a set of regression/decision trees.
w_i	the weight of a regression/decision tree.
d_T	the depth of tree T.
s_i	a tree score contribution.
n_i	a branching node of a regression/decision tree.
$e_i(\mathbf{x})$	the exit leaf of a tree for a given feature vector.
$\mathcal{N}_i$	the set of branching nodes of a tree.
$\mathcal{L}_i$	the set of leaf nodes of a tree.
γ	a threshold value.
a	a array of two elements.
`tid`	an array of tree ids.
`mask`	an array of mask bitvectors.
`th`	an array of threshold values.
`exit`	an array of leaf bitvectors.
`scores`	a lookup table of scores.
$\mathcal{Q}$	a set of queries.

1

Introduction

Search engines are exceptionally important tools for accessing information in today's increasingly digital world. Classical commercial Web search engines, such as those maintained by Google, Bing, Yandex, Baidu, have processed billions if not trillions of Web documents, and have kept maintaining these in continuously updated index data structures[1] requiring petabytes of storage space,[2] to ensure *satisfying* the users of the search engine through billions of user queries received every month[3] (Bosch *et al.*, 2016).

Satisfaction of the search engine users is a key metric for search engine providers. Without drawing too broad a sweeping generalisation, one of the fundamental goals of a search engine is to derive income from advertising traffic, for instance from the ads that are often presented next to the organic search results. Users that are not satisfied with the search engine results may switch to a different engine (White, 2016), and may not return. This is a loss of advertising revenue for the search

[1] https://www.google.com/insidesearch/howsearchworks/thestory/index.html

[2] https://www.google.com/insidesearch/howsearchworks/crawling-indexing.html

[3] https://googleblog.blogspot.it/2010/09/google-instant-behind-scenes.html

engine. As a consequence, ensuring that their users are satisfied with the results is of utmost importance to search engines.

There are various reasons why the result page for a search does not satisfy a user (Diriye *et al.*, 2012), but the primary causes are the *effectiveness* – the quality of the returned results – and the *efficiency* – the speed at which the results were returned. Indeed, search engines that are slow to return results to users can negatively damage the user's perception of the quality of the results (Brutlag and Schuman, 2009). Hence, a search engine needs to be both effective (deploying advanced ranking mechanisms), while ensuring that its results are efficiently returned to the user. A key contribution of this survey is to review both the foundational background and the recent advances in search engine infrastructures.

Fortunately, search is a parallelisable problem, and scaling can be applied to the search engine computing infrastructure. Indeed, as shown in Figure 1.1, a large index can be partitioned across multiple *shards*, allowing each single search engine server to service a small portion of the index in order to ensure fast retrieval. Each of the S index shards can be replicated R times, allowing both resilience and scaling. When user queries arrive, the *broker* routes queries to the less loaded replica of each shard for processing (Freire *et al.*, 2013; Freire *et al.*, 2012).

Using such a distributed setting for a large search engine, $R \times S$ can be very large, covering potentially hundreds of thousands of servers. All of the major search engines run exceedingly large data centres, each often requiring capital investments of billions of dollars,[4] and consuming vast quantities of energy. Data centres use 3% of the global electricity supply and account for about 2% of the total greenhouse gas emissions; this is expected to triple in the next decade, putting an enormous strain on energy supplies and dealing a hefty blow to efforts to contain global warming.[5]

[4] http://www.datacenterknowledge.com/archives/2014/07/23/from-112-servers-to-5b-spent-on-google-data-centers-per-quarter

[5] http://www.independent.co.uk/environment/global-warming-data-centres-to-consume-three-times-as-much-energy-in-next-decade-experts-warn-a6830086.html

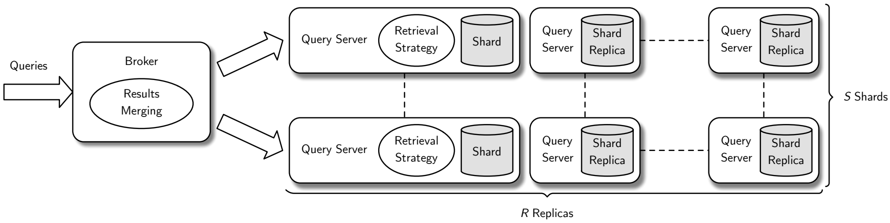

Figure 1.1: Distributed retrieval architecture.

Clearly, at such large scales, the efficiency of the search engine's operating internals are therefore key to the operating costs of such companies. Efficiency improvements of 5% would allow a 5% reduction in R replicas, potentially equating to significant power consumption reductions, and providing a room for further growing the sizes of the search engines' indexes, or servicing growth in the user queries.

The distributed nature of a search engine infrastructure is not within the scope of this survey. The interested reader can find a comprehensive overview of distributed large-scale Web search engines in (Cambazoglu and Baeza-Yates, 2015). Instead, this survey focuses on the general architecture of a search infrastructure as might be deployed within a single server. Our goal is to provide an accurate description of the basic search components involved in the scoring of documents in response to a query, together with a detailed and exhaustive review of the research works aiming at boosting the efficiency of query processing without negatively impacting the effectiveness performance of the system.

A key detail of the manner in which a search engine is designed to operate is the "top-heavy nature" of results: since the users of search engines typically focus on the top-ranked results (as can be measured offline using test collections and metrics such as NDCG (Järvelin and Kekäläinen, 2002) and ERR (Chapelle *et al.*, 2009)), the relevance of those results is key to user satisfaction. This means that the search engine should itself focus on getting the most relevant results at the top of the ranking, at the possible detriment of mis-ranking other results. In his SIGIR 2010 Industry Day talk, Pedersen (2010) described this process as the use of cascading (illustrated in Figure 1.2). In response to a query, each conceptual cascade aims to filter or rank documents, before passing onto the next cascade layer. At the bottom layer, the documents to be retrieved are defined in terms of the subsets of terms present in the query – being able to identify these subsets as quickly as possible, without requiring to scan the contents of each document, is a fundamental architecture decision of an Information Retrieval (IR) system. The bottom layer may filter a collection of billions of documents down to the millions, which should be scored. In the second layer, query processing techniques define how the scoring of document weighting models, such as language modelling or BM25 should be applied. In the final layer (the top layer), various additional ranking features such as PageRank, or URL information may be calculated and used within a learned model to re-rank the documents, before presenting the final top K high-scored documents to the user (usually K is small, e.g., $8 - 20$, as displayed on the first page of the search results).

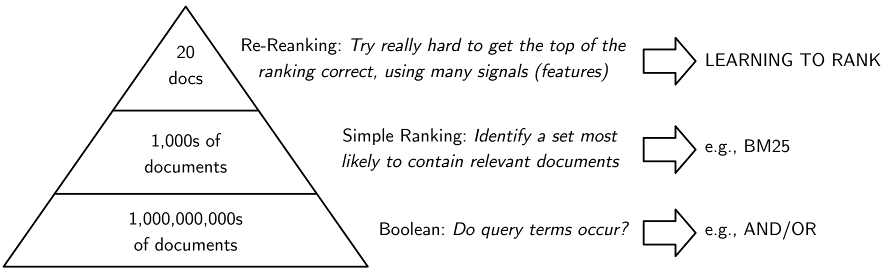

Figure 1.2: Cascading nature of Web search, based on (Pedersen, 2010)

Different techniques are appropriate at different cascade levels, but many are designed to make efficiency savings by avoiding the scoring of documents, which cannot make the top-ranked results that will be returned to the user. In this survey, we cover both the core algorithms and data structures used for retrieval, as well as the optimisations (such as dynamic pruning) that can be applied at a given cascade level. Of course, not all queries are equal – some are easier for the search engine to answer *effectively*, while, orthogonally, some may be less *efficient*, i.e., take longer for the search engine to answer. Being able to know the likely efficiency of a query, as might be obtained from a *query efficiency predictor*, can allow the search engine to make on-the-fly decisions about its configuration.

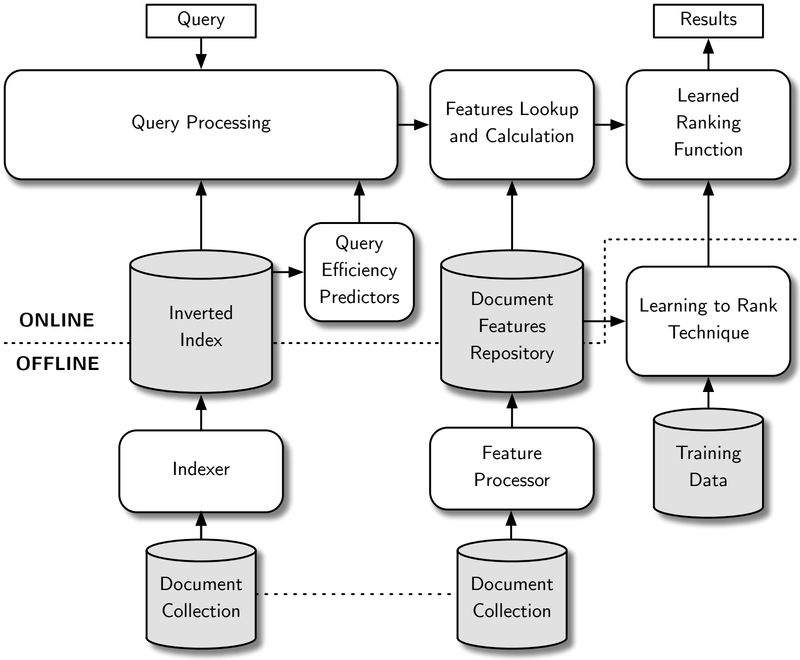

Figure 1.3: A conceptual architecture for a search engine.

Figure 1.3 provides the main infrastructure that is discussed in this survey. We will focus on the "online" components, e.g., those responsible for the cascading components of search, while referring to the "offline" components whenever it is necessary. The remainder of this survey is structured as follows:

- Chapter 2 provides an overview of the modern infrastructure foundations within a search engine, covering the basic form of the inverted index data structure, and the essentials of query processing.
- Chapter 3 provides an introduction to approaches for increasing the efficiency of query processing, namely the dynamic pruning techniques.
- Chapter 4 describes query efficiency predictors – a new technique to estimate the response time of queries – that is gaining attention

for a number of applications involving efficient retrieval on a per-query basis.
- Chapter 5 provides an overview of impact-sorted indexes, which make offline changes to the layout of the inverted index in order to improve the efficiency of query processing.
- Chapter 6 provides an overview of cascading search architectures, and provides insights into how to efficiently deploy learning-to-rank, a retrieval technique known to benefit the search engine's effectiveness by re-ranking a set of K documents.
- Chapter 7 gives an overview of the current open directions in retrieval infrastructures, including the use of signature files instead of inverted indexes, and provides concluding remarks.

Note on Efficiency Performance Measures

In this survey, we illustrate the efficiency measures reported in the cited papers. Since this survey covers papers from over a period of 30 years, comparing the reported results across different papers could lead to the wrong conclusions. Hence, we will only report comparative performance measures derived from single contributions.

The performances of the discussed strategies naturally depend on several factors, such as the index and/or the query characteristics, the inverted index compression, the similarity function adopted, the number of documents returned, the actual underlying implementations, the machine(s) used to perform the experiments and so on. In most papers, when comparing the efficiency of different solution s, two main quantities are typically reported: response times and/or number of processed elements. In order to be as "implementation-independent" as possible, we report the speedup of an optimisation w.r.t. the baseline, in terms of mean response time, and/or its (work) *reduction*, defined as the percentage of postings that are dynamically pruned, i.e., not scored, w.r.t. the baseline.

When comparing two time quantities t_1 and t_2, with $t_1 > t_2$ we will always report their relative *speedup* s, defined as $s = t_1/t_2$ (always greater than 1). For example, if two strategies A and B have an average response time of 20 ms and 8 ms, respectively, their speedup (of B

w.r.t. to A) is $s = t_A/t_B = 20/8 = 2.5\times$. When comparing two numbers of processed elements n_1 and n_2, with $n_1 > n_2$ we will systematically report the percentage *reduction* r, defined as $r = 1 - n_2/n_1$. For example, if strategy A processes 200 elements while strategy B processes just 150 elements, the reduction of B w.r.t. A is $r = 1 - n_B/n_A = 1 - 150/200 = 0.25 = 25\%$.

Finally, the throughput of a query processing node, as well as that of more complex search systems, is measured in *queries per second* (QPS).

Intended Audience

This survey targets readers, researchers and engineers who possess a basic knowledge in Information Retrieval (IR) or in other cognate topics (e.g., databases, data mining). In particular, the survey is of utmost interest to PhD students, researchers and practitioners working on efficiency and system infrastructures in IR and Web search. Indeed, anyone working on search and ranking on big data will benefit from this manuscript. The survey is also particularly of interest to lecturers and tutors looking for a concise and comprehensive textbook on state-of-the-art query processing techniques to support their IR course.

Note on the Origins of the Material

This survey is a new piece of work, but builds upon our research experience in this area. This survey also benefits from the authors' experience acquired from presenting two related tutorials at ECIR 2017 and SIGIR 2018. We would like to thank the attendees of these tutorials for their insightful questions and comments.

References

Amati, G. and C. J. Van Rijsbergen. 2002. “Probabilistic Models of Information Retrieval Based on Measuring the Divergence from Randomness”. *ACM Trans. Inf. Syst.* 20(4): 357–389. ISSN: 1046-8188. DOI: 10.1145/582415.582416.

Anh, V. N., O. de Kretser, and A. Moffat. 2001. “Vector-space Ranking with Effective Early Termination”. In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM. 35–42. ISBN: 1-58113-331-6. DOI: 10.1145/383952.383957. URL: http://doi.acm.org/10.1145/383952.383957.

Anh, V. N. and A. Moffat. 1998. “Compressed Inverted Files with Reduced Decoding Overheads”. In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’98*. Melbourne, Australia: ACM. 290–297. ISBN: 1-58113-015-5. DOI: 10.1145/290941.291011. URL: http://doi.acm.org/10.1145/290941.291011.

Anh, V. N. and A. Moffat. 2002. “Impact transformation: effective and efficient web retrieval”. In: *SIGIR ’02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in Information Retrieval*. Ed. by M. Beaulieu, R. Baeza-Yates, S. H. Myaeng, and K. Jarvelin. New York, NY, USA: ACM. 3–10. ISBN: 1-58113-561-0. DOI: http://doi.acm.org/10.1145/564376.564380.

Anh, V. N. and A. Moffat. 2005a. "Inverted Index Compression Using Word-Aligned Binary Codes". *Inf. Retr.* 8(1): 151–166. ISSN: 1386-4564. DOI: 10.1023/B:INRT.0000048490.99518.5c. URL: http://dx.doi.org/10.1023/B:INRT.0000048490.99518.5c.

Anh, V. N. and A. Moffat. 2005b. "Simplified Similarity Scoring Using Term Ranks". In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '05*. Salvador, Brazil: ACM. 226–233. ISBN: 1-59593-034-5. DOI: 10.1145/1076034.1076075. URL: http://doi.acm.org/10.1145/1076034.1076075.

Anh, V. N. and A. Moffat. 2006a. "Pruned Query Evaluation Using Pre-computed Impacts". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '06*. Seattle, Washington, USA: ACM. 372–379. ISBN: 1-59593-369-7. DOI: 10.1145/1148170.1148235. URL: http://doi.acm.org/10.1145/1148170.1148235.

Anh, V. N. and A. Moffat. 2006b. "Pruning Strategies for Mixed-mode Querying". In: *Proceedings of the 15th ACM International Conference on Information and Knowledge Management. CIKM '06*. Arlington, Virginia, USA: ACM. 190–197. ISBN: 1-59593-433-2. DOI: 10.1145/1183614.1183645. URL: http://doi.acm.org/10.1145/1183614.1183645.

Anh, V. N. and A. Moffat. 2006c. "Structured Index Organizations for High-Throughput Text Querying". In: *String Processing and Information Retrieval: 13th International Conference, SPIRE 2006, Glasgow, UK, October 11-13, 2006. Proceedings*. Ed. by F. Crestani, P. Ferragina, and M. Sanderson. Berlin, Heidelberg: Springer Berlin Heidelberg. 304–315. ISBN: 978-3-540-45775-6. DOI: 10.1007/11880561_25. URL: http://dx.doi.org/10.1007/11880561_25.

Arapakis, I., X. Bai, and B. B. Cambazoglu. 2014. "Impact of Response Latency on User Behavior in Web Search". In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. SIGIR '14*. Gold Coast, Queensland, Australia: ACM. 103–112. ISBN: 978-1-4503-2257-7. DOI: 10.1145/2600428.2609627. URL: http://doi.acm.org/10.1145/2600428.2609627.

Arroyuelo, D., S. González, M. Marin, M. Oyarzún, and T. Suel. 2012. "To Index or Not to Index: Time-space Trade-offs in Search Engines with Positional Ranking Functions". In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '12*. Portland, Oregon, USA: ACM. 255–264. ISBN: 978-1-4503-1472-5. DOI: 10.1145/2348283.2348320. URL: http://doi.acm.org/10.1145/2348283.2348320.

Asadi, N. and J. Lin. 2012. "Fast Candidate Generation for Two-phase Document Ranking: Postings List Intersection with Bloom Filters". In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management. CIKM '12*. Maui, Hawaii, USA: ACM. 2419–2422. ISBN: 978-1-4503-1156-4. DOI: 10.1145/2396761.2398656. URL: http://doi.acm.org/10.1145/2396761.2398656.

Asadi, N. and J. Lin. 2013. "Document Vector Representations for Feature Extraction in Multi-stage Document Ranking". *Inf. Retr.* 16(6): 747–768. ISSN: 1386-4564. DOI: 10.1007/s10791-012-9217-9. URL: http://dx.doi.org/10.1007/s10791-012-9217-9.

Asadi, N., J. Lin, and M. Busch. 2013. "Dynamic Memory Allocation Policies for Postings in Real-time Twitter Search". In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '13*. Chicago, Illinois, USA: ACM. 1186–1194. ISBN: 978-1-4503-2174-7. DOI: 10.1145/2487575.2488221. URL: http://doi.acm.org/10.1145/2487575.2488221.

Asadi, N., J. Lin, and A. P. de Vries. 2014. "Runtime Optimizations for Tree-Based Machine Learning Models." *IEEE Trans. Knowl. Data Eng.* 26(9): 2281–2292.

Baeza-Yates, R. and B. Ribeiro-Neto. 2008. *Modern Information Retrieval*. 2nd. USA: Addison-Wesley Publishing Company.

Barbay, J., A. López-Ortiz, and T. Lu. 2006. "Faster Adaptive Set Intersections for Text Searching". In: *Experimental Algorithms*. Ed. by C. Àlvarez and M. Serna. Berlin, Heidelberg: Springer Berlin Heidelberg. 146–157. ISBN: 978-3-540-34598-5.

Barroso, L. A., J. Clidaras, and U. Hölzle. 2013. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Second Edition. Synthesis Lectures on Computer Architecture*. Morgan & Claypool Publishers. ISBN: 9781627050098. DOI: 10.2200/S00516ED2V01Y201306CAC024. URL: https://doi.org/10.2200/S00516ED2V01Y201306CAC024.

Barroso, L. A. and U. Hölzle. 2007. "The Case for Energy-Proportional Computing". *Computer*. 40(12): 33–37. ISSN: 0018-9162. DOI: 10.1109/MC.2007.443. URL: https://doi.org/10.1109/MC.2007.443.

Bartell, B. T., G. W. Cottrell, and R. K. Belew. 1994. "Automatic Combination of Multiple Ranked Retrieval Systems". In: *Research and Development in Information Retrieval*. 173–181. URL: citeseer.ist.psu.edu/bartell94automatic.html.

Belkhir, L. and A. Elmeligi. 2018. "Assessing ICT global emissions footprint: Trends to 2040 & recommendations". *Journal of Cleaner Production*. 177: 448–463. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2017.12.239. URL: http://www.sciencedirect.com/science/article/pii/S095965261733233X.

Bendersky, M., W. B. Croft, and Y. Diao. 2011. "Quality-biased ranking of web documents". In: *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM. 95–104.

Bendersky, M., D. Metzler, and W. B. Croft. 2010. "Learning Concept Importance Using a Weighted Dependence Model". In: *WSDM*. 31–40.

Blanco, R., M. Catena, and N. Tonellotto. 2016. "Exploiting Green Energy to Reduce the Operational Costs of Multi-Center Web Search Engines". In: *Proceedings of the 25th International Conference on World Wide Web*. *WWW '16*. Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee. 1237–1247. ISBN: 978-1-4503-4143-1. DOI: 10.1145/2872427.2883021. URL: https://doi.org/10.1145/2872427.2883021.

Boldi, P. and S. Vigna. 2005. "Compressed perfect embedded skip lists for quick inverted-index lookups". In: *Proceedings of SPIRE 2005, Lecture Notes in Computer Science*. Springer. 25–28.

Bonacic, C., C. Garcia, M. Marin, M. Prieto, F. Tirado, and C. Vicente. 2008. "Improving Search Engines Performance on Multithreading Processors". In: *High Performance Computing for Computational Science - VECPAR 2008*. Ed. by J. M. L. M. Palma, P. R. Amestoy, M. Daydé, M. Mattoso, and J. C. Lopes. Berlin, Heidelberg: Springer Berlin Heidelberg. 201–213. ISBN: 978-3-540-92859-1.

Bookstein, A., S. T. Klein, and T. Raita. 1997. "Modeling Word Occurrences for the Compression of Concordances". *ACM Trans. Inf. Syst.* 15(3): 254–290. ISSN: 1046-8188. DOI: 10.1145/256163.256166. URL: http://doi.acm.org/10.1145/256163.256166.

Bortnikov, E., D. Carmel, and G. Golan-Gueta. 2017. "Top-k Query Processing with Conditional Skips". In: *Proceedings of the 26th International Conference on World Wide Web Companion*. Perth, Australia: International World Wide Web Conferences Steering Committee. 653–661. ISBN: 978-1-4503-4914-7. DOI: 10.1145/3041021.3054191.

Bosch, A. van den, T. Bogers, and M. de Kunder. 2016. "Estimating search engine index size variability: a 9-year longitudinal study". *Scientometrics*. 107(2): 839–856. ISSN: 1588-2861. DOI: 10.1007/s11192-016-1863-z. URL: http://dx.doi.org/10.1007/s11192-016-1863-z.

Breiman, L. 2001. "Random Forests". *Machine Learning*. 45(1): 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

Broccolo, D., C. Macdonald, S. Orlando, I. Ounis, R. Perego, F. Silvestri, and N. Tonellotto. 2013. "Load-sensitive Selective Pruning for Distributed Search". In: *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*. *CIKM '13*. San Francisco, California, USA: ACM. 379–388. ISBN: 978-1-4503-2263-8. DOI: 10.1145/2505515.2505699. URL: http://doi.acm.org/10.1145/2505515.2505699.

Broder, A. Z., D. Carmel, M. Herscovici, A. Soffer, and J. Zien. 2003. "Efficient query evaluation using a two-level retrieval process". In: *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*. New York, NY, USA: ACM. 426–434. ISBN: 1-58113-723-0. DOI: http://doi.acm.org/10.1145/956863.956944.

Brown, E. W. 1995. "Fast evaluation of structured queries for information retrieval". In: *SIGIR '95: Proceedings of the 18th annual international ACM SIGIR conference on Research and development in Information Retrieval*. New York, NY, USA: ACM. 30–38. ISBN: 0-89791-714-6. DOI: http://doi.acm.org/10.1145/215206.215329.

Brutlag, J. and E. Schuman. 2009. *Performance Related Changes and their User Impact*. URL: http://goo.gl/hAUENq.

Buckley, C. and A. F. Lewit. 1985. "Optimization of inverted vector searches". In: *SIGIR '85: Proceedings of the 8th annual international ACM SIGIR conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM. 97–110. ISBN: 0-89791-159-8. DOI: http://doi.acm.org/10.1145/253495.253515.

Burges, C., T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. 2005. "Learning to Rank Using Gradient Descent". In: *Proceedings of the 22Nd International Conference on Machine Learning. ICML '05*. Bonn, Germany: ACM. 89–96. ISBN: 1-59593-180-5. DOI: 10.1145/1102351.1102363. URL: http://doi.acm.org/10.1145/1102351.1102363.

Busch, M., K. Gade, B. Larson, P. Lok, S. Luckenbill, and J. Lin. 2012. "Earlybird: Real-Time Search at Twitter". In: *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering. ICDE '12*. Washington, DC, USA: IEEE Computer Society. 1360–1369. ISBN: 978-0-7695-4747-3. DOI: 10.1109/ICDE.2012.149. URL: http://dx.doi.org/10.1109/ICDE.2012.149.

Büttcher, S., C. Clarke, and G. V. Cormack. 2010. *Information Retrieval: Implementing and Evaluating Search Engines*. The MIT Press. ISBN: 0262026511.

Cambazoglu, B. B. and C. Aykanat. 2006. "Performance of Query Processing Implementations in Ranking-based Text Retrieval Systems Using Inverted Indices". *Inf. Process. Manage.* 42(4): 875–898. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2005.06.004. URL: http://dx.doi.org/10.1016/j.ipm.2005.06.004.

Cambazoglu, B. B., H. Zaragoza, O. Chapelle, J. Chen, C. Liao, Z. Zheng, and J. Degenhardt. 2010. "Early exit optimizations for additive machine learned ranking systems". In: *Proceedings of the third ACM international conference on Web search and data mining. WSDM '10*. New York, New York, USA: ACM. 411–420. ISBN: 978-1-60558-889-6. DOI: http://doi.acm.org/10.1145/1718487.1718538. URL: http://doi.acm.org/10.1145/1718487.1718538.

Cambazoglu, B. B. and R. A. Baeza-Yates. 2015. *Scalability Challenges in Web Search Engines. Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan & Claypool Publishers. DOI: 10.2200/S00662ED1V01Y201508ICR045. URL: http://dx.doi.org/10.2200/S00662ED1V01Y201508ICR045.

Cao, Z., T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. 2007. "Learning to Rank: From Pairwise Approach to Listwise Approach". In: *Proceedings of the 24th International Conference on Machine Learning. ICML '07*. Corvalis, Oregon, USA: ACM. 129–136. ISBN: 978-1-59593-793-3. DOI: 10.1145/1273496.1273513. URL: http://doi.acm.org/10.1145/1273496.1273513.

Capannini, G., C. Lucchese, F. M. Nardini, S. Orlando, R. Perego, and N. Tonellotto. 2016. "Quality versus efficiency in document scoring with learning-to-rank models". *Information Processing & Management*. 52(6): 1161–1177. ISSN: 0306-4573. DOI: https://doi.org/10.1016/j.ipm.2016.05.004. URL: http://www.sciencedirect.com/science/article/pii/S0306457316301248.

Carmel, D., D. Cohen, R. Fagin, E. Farchi, M. Herscovici, Y. S. Maarek, and A. Soffer. 2001. "Static Index Pruning for Information Retrieval Systems". In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '01*. New Orleans, Louisiana, USA: ACM. 43–50. ISBN: 1-58113-331-6. DOI: 10.1145/383952.383958. URL: http://doi.acm.org/10.1145/383952.383958.

Carmel, D. and E. Yom-Tov. 2010. *Estimating the Query Difficulty for Information Retrieval. Synthesis Lectures on Information Concepts, Retrieval, and Services*. Morgan & Claypool Publishers. DOI: 10.2200/S00235ED1V01Y201004ICR015. URL: https://doi.org/10.2200/S00235ED1V01Y201004ICR015.

Catena, M. and N. Tonellotto. 2017. "Energy-Efficient Query Processing in Web Search Engines". *IEEE Transactions on Knowledge and Data Engineering*. 29(7): 1412–1425. ISSN: 1041-4347. DOI: 10.1109/TKDE.2017.2681279.

Catena, M., O. Frieder, and N. Tonellotto. 2018. "Efficient Energy Management in Distributed Web Search". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM. 1555–1558. DOI: 10.1145/3269206.3269263.

Catena, M., C. Macdonald, and N. Tonellotto. 2015. "Load-sensitive CPU Power Management for Web Search Engines". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 751–754.

Chakrabarti, K., S. Chaudhuri, and V. Ganti. 2011. "Interval-based Pruning for Top-k Processing over Compressed Lists". In: *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering. ICDE '11*. Washington, DC, USA: IEEE Computer Society. 709–720. ISBN: 978-1-4244-8959-6. DOI: 10.1109/ICDE.2011.5767855. URL: http://dx.doi.org/10.1109/ICDE.2011.5767855.

Chapelle, O. and Y. Chang. 2011. "Yahoo! Learning to Rank Challenge Overview." *Journal of Machine Learning Research-Proceedings Track*. 14: 1–24.

Chapelle, O., D. Metlzer, Y. Zhang, and P. Grinspan. 2009. "Expected Reciprocal Rank for Graded Relevance". In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management. CIKM '09*. Hong Kong, China: ACM. 621–630. ISBN: 978-1-60558-512-3. DOI: 10.1145/1645953.1646033. URL: http://doi.acm.org/10.1145/1645953.1646033.

Chen, R.-C., L. Gallagher, R. Blanco, and J. S. Culpepper. 2017. "Efficient Cost-Aware Cascade Ranking in Multi-Stage Retrieval". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17.* Shinjuku, Tokyo, Japan: ACM. 445–454. ISBN: 978-1-4503-5022-8. DOI: 10.1145/3077136.3080819. URL: http://doi.acm.org/10.1145/3077136.3080819.

Chierichetti, F., S. Lattanzi, F. Mari, and A. Panconesi. 2008. "On Placing Skips Optimally in Expectation". In: *Proceedings of the 2008 International Conference on Web Search and Data Mining. WSDM '08.* Palo Alto, California, USA: ACM. 15–24. ISBN: 978-1-59593-927-2. DOI: 10.1145/1341531.1341537. URL: http://doi.acm.org/10.1145/1341531.1341537.

Clarke, C. L., J. S. Culpepper, and A. Moffat. 2016. "Assessing Efficiency—effectiveness Tradeoffs in Multi-stage Retrieval Systems Without Using Relevance Judgments". *Inf. Retr.* 19(4): 351–377. ISSN: 1386-4564. DOI: 10.1007/s10791-016-9279-1. URL: http://dx.doi.org/10.1007/s10791-016-9279-1.

Crane, M., J. S. Culpepper, J. Lin, J. Mackenzie, and A. Trotman. 2017. "A Comparison of Document-at-a-Time and Score-at-a-Time Query Evaluation". In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. WSDM '17.* Cambridge, United Kingdom: ACM. 201–210. ISBN: 978-1-4503-4675-7. DOI: 10.1145/3018661.3018726. URL: http://doi.acm.org/10.1145/3018661.3018726.

Crane, M. and J. Lin. 2017. "An Exploration of Serverless Architectures for Information Retrieval". In: *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval. ICTIR '17.* Amsterdam, The Netherlands: ACM. 241–244. ISBN: 978-1-4503-4490-6. DOI: 10.1145/3121050.3121086. URL: http://doi.acm.org/10.1145/3121050.3121086.

Crane, M., A. Trotman, and R. O'Keefe. 2013. "Maintaining discriminatory power in quantized indexes". In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. CIKM '13*. San Francisco, California, USA: ACM. 1221–1224. ISBN: 978-1-4503-2263-8. DOI: 10.1145/2505515.2507860. URL: http://doi.acm.org/10.1145/2505515.2507860.

Croft, B., D. Metzler, and T. Strohman. 2009. *Search Engines: Information Retrieval in Practice*. 1st. USA: Addison-Wesley Publishing Company. ISBN: 0136072240.

Cronen-Townsend, S., Y. Zhou, and W. B. Croft. 2002. "Predicting Query Performance". In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '02*. Tampere, Finland: ACM. 299–306. ISBN: 1-58113-561-0. DOI: 10.1145/564376.564429. URL: http://doi.acm.org/10.1145/564376.564429.

Culpepper, J. S., C. L. A. Clarke, and J. Lin. 2016. "Dynamic Cutoff Prediction in Multi-Stage Retrieval Systems". In: *Proceedings of the 21st Australasian Document Computing Symposium. ADCS '16*. Caulfield, VIC, Australia: ACM. 17–24. ISBN: 978-1-4503-4865-2. DOI: 10.1145/3015022.3015026. URL: http://doi.acm.org/10.1145/3015022.3015026.

Culpepper, J. S., F. Diaz, and M. D. Smucker. 2018. "Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018)". *SIGIR Forum*. 52(1): 34–90. ISSN: 0163-5840. DOI: 10.1145/3274784.3274788. URL: http://doi.acm.org/10.1145/3274784.3274788.

Culpepper, J. S. and A. Moffat. 2010. "Efficient Set Intersection for Inverted Indexing". *ACM Trans. Inf. Syst.* 29(1): 1:1–1:25. ISSN: 1046-8188. DOI: 10.1145/1877766.1877767. URL: http://doi.acm.org/10.1145/1877766.1877767.

Dang, V., M. Bendersky, and W. B. Croft. 2013. "Two-Stage Learning to Rank for Information Retrieval". In: *Proceedings of the 35th European Conference on Advances in Information Retrieval. ECIR'13*. Moscow, Russia: Springer-Verlag. 423–434. ISBN: 978-3-642-36972-8. DOI: 10.1007/978-3-642-36973-5_36. URL: http://dx.doi.org/10.1007/978-3-642-36973-5_36.

Daoud, C. M., E. S. de Moura, A. Carvalho, A. S. da Silva, D. Fernandes, and C. Rossi. 2016. "Fast top-k preserving query processing using two-tier indexes". *Information Processing & Management*. 52(5): 855–872. ISSN: 0306-4573. DOI: http://dx.doi.org/10.1016/j.ipm.2016.03.005. URL: http://www.sciencedirect.com/science/article/pii/S0306457316300371.

Dato, D., C. Lucchese, F. M. Nardini, S. Orlando, R. Perego, N. Tonellotto, and R. Venturini. 2016. "Fast Ranking with Additive Ensembles of Oblivious and Non-Oblivious Regression Trees". *ACM Trans. Inf. Syst.* 35(2): 15:1–15:31. ISSN: 1046-8188. DOI: 10.1145/2987380. URL: http://doi.acm.org/10.1145/2987380.

Dean, J. 2009. "Keynote: Challenges in building large-scale information retrieval systems". In: *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining (WSDM)*.

Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. 1990. "Indexing by latent semantic analysis". *Journal of the American Society for Information Science*. 41(6): 391–407. ISSN: 1097-4571. DOI: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9. URL: http://dx.doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASI1%3E3.0.CO;2-9.

Dehghani, M., H. Zamani, A. Severyn, J. Kamps, and W. B. Croft. 2017. "Neural Ranking Models with Weak Supervision". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17.* Shinjuku, Tokyo, Japan: ACM. 65–74. ISBN: 978-1-4503-5022-8. DOI: 10.1145/3077136.3080832. URL: http://doi.acm.org/10.1145/3077136.3080832.

Demaine, E. D., A. López-Ortiz, and J. I. Munro. 2000. "Adaptive Set Intersections, Unions, and Differences". In: *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '00.* San Francisco, California, USA: Society for Industrial and Applied Mathematics. 743–752. ISBN: 0-89871-453-2. URL: http://dl.acm.org/citation.cfm?id=338219.338634.

Deveaud, R., M.-D. Albakour, C. Macdonald, and I. Ounis. 2015. "Experiments with a Venue-Centric Model for Personalisedand Time-Aware Venue Suggestion". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. CIKM '15*. Melbourne, Australia: ACM. 53–62. ISBN: 978-1-4503-3794-6. DOI: 10.1145/2806416.2806484. URL: http://doi.acm.org/10.1145/2806416.2806484.

Dimopoulos, C., S. Nepomnyachiy, and T. Suel. 2013a. "A Candidate Filtering Mechanism for Fast Top-k Query Processing on Modern Cpus". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '13*. Dublin, Ireland: ACM. 723–732. ISBN: 978-1-4503-2034-4. DOI: 10.1145/2484028.2484087. URL: http://doi.acm.org/10.1145/2484028.2484087.

Dimopoulos, C., S. Nepomnyachiy, and T. Suel. 2013b. "Optimizing Top-k Document Retrieval Strategies for Block-max Indexes". In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. WSDM '13*. Rome, Italy: ACM. 113–122. ISBN: 978-1-4503-1869-3. DOI: 10.1145/2433396.2433412. URL: http://doi.acm.org/10.1145/2433396.2433412.

Ding, S. and T. Suel. 2011. "Faster Top-k Document Retrieval Using Block-max Indexes". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11*. Beijing, China: ACM. 993–1002. ISBN: 978-1-4503-0757-4. DOI: 10.1145/2009916.2010048. URL: http://doi.acm.org/10.1145/2009916.2010048.

Diriye, A., R. White, G. Buscher, and S. Dumais. 2012. "Leaving So Soon?: Understanding and Predicting Web Search Abandonment Rationales". In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management. CIKM '12*. Maui, Hawaii, USA: ACM. 1025–1034. ISBN: 978-1-4503-1156-4. DOI: 10.1145/2396761.2398399. URL: http://doi.acm.org/10.1145/2396761.2398399.

Doszkocs, T. E. 1982. "From Research to Application: The Cite Natural Language Information Retrieval System". In: *Proceedings of the 5th Annual ACM Conference on Research and Development in Information Retrieval. SIGIR '82*. West Berlin, Germany: Springer-Verlag New York, Inc. 251–262. ISBN: 0-387-11978-7. URL: http://dl.acm.org/citation.cfm?id=636713.636731.

Elias, P. 1974. "Efficient Storage and Retrieval by Content and Address of Static Files". *J. ACM*. 21(2): 246–260.

Elias, P. 1975. "Universal codeword sets and representations of the integers". *IEEE Transactions on Information Theory*. 21(2): 194–203. ISSN: 0018-9448. DOI: 10.1109/TIT.1975.1055349.

Fano, R. M. 1971. "On the number of bits required to implement an associative memory". *Memorandum 61, Computer Structures Group, MIT, Cambridge, MA*.

Fontoura, M., V. Josifovski, J. Liu, S. Venkatesan, X. Zhu, and J. Zien. 2011. "Evaluation Strategies for Top-k Queries over Memory-Resident Inverted Indexes". *Proceedings of the VLDB Endowment*. 4(12): 1213–1224.

Fox, E. A. and J. A. Shaw. 1994. "Combination of Multiple Searches". In: *Proceedings of the 2nd Text REtrieval Conference (TREC-2)*. NIST Special Publication 500-215. 243–252.

Frachtenberg, E. 2009. "Reducing Query Latencies in Web Search Using Fine-Grained Parallelism". *World Wide Web*. 12(4): 441. ISSN: 1573-1413. DOI: 10.1007/s11280-009-0066-4. URL: https://doi.org/10.1007/s11280-009-0066-4.

Freire, A., C. Macdonald, N. Tonellotto, I. Ounis, and F. Cacheda. 2012. "Scheduling Queries Across Replicas". In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 1139–1140. DOI: 10.1145/2348283.2348508.

Freire, A., C. Macdonald, N. Tonellotto, I. Ounis, and F. Cacheda. 2013. "Hybrid Query Scheduling for a Replicated Search Engine". In: *Proceedings of the 35th European Conference on IR Research*. Springer. 435–446. DOI: 10.1007/978-3-642-36973-5_37.

Freire, A., C. Macdonald, N. Tonellotto, I. Ounis, and F. Cacheda. 2014. "A Self-adapting Latency/Power Tradeoff Model for Replicated Search Engines". In: *Proceedings of the 7th ACM International Conference on Web Search and Data Mining. WSDM '14*. New York, New York, USA: ACM. 13–22. ISBN: 978-1-4503-2351-2. DOI: 10.1145/2556195.2556246. URL: http://doi.acm.org/10.1145/2556195.2556246.

Freire, A., C. Macdonald, N. Tonellotto, I. Ounis, and F. Cacheda. 2015. "Queuing Theory-based Latency/Power Tradeoff Models for Replicated Search Engines". *J. UCS*. 21(13): 1790–1809. DOI: 10.3217/jucs-021-13-1790.

Freund, Y., R. Iyer, R. E. Schapire, and Y. Singer. 2003. "An Efficient Boosting Algorithm for Combining Preferences". *J. Mach. Learn. Res.* 4(Dec.): 933–969. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=945365.964285.

Friedman, J. H. 2001. "Greedy function approximation: A gradient boosting machine." *Ann. Statist.* 29(5): 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451.

Goldstein, J., R. Ramakrishnan, and U. Shaft. 1998. "Compressing Relations and Indexes". In: *Proceedings of the Fourteenth International Conference on Data Engineering. ICDE '98*. Washington, DC, USA: IEEE Computer Society. 370–379. ISBN: 0-8186-8289-2. URL: http://dl.acm.org/citation.cfm?id=645483.656226.

Golomb, S. 1966. "Run-length Encodings (Corresp.)" *IEEE Trans. Inf. Theor.* 12(3).

Goodwin, B., M. Hopcroft, D. Luu, A. Clemmer, M. Curmei, S. Elnikety, and Y. He. 2017. "BitFunnel: Revisiting Signatures for Search". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17*. Shinjuku, Tokyo, Japan: ACM. 605–614. ISBN: 978-1-4503-5022-8. DOI: 10.1145/3077136.3080789. URL: http://doi.acm.org/10.1145/3077136.3080789.

Greenberg, A., J. Hamilton, D. A. Maltz, and P. Patel. 2008. "The Cost of a Cloud: Research Problems in Data Center Networks". *SIGCOMM Comput. Commun. Rev.* 39(1): 68–73. ISSN: 0146-4833. DOI: 10.1145/1496091.1496103. URL: http://doi.acm.org/10.1145/1496091.1496103.

Harman, D. and G. Candela. 1990. "Retrieving records from a gigabyte of text on a minicomputer using statistical ranking". *Journal of the American Society for Information Science.* 41(8): 581–589. ISSN: 1097-4571. DOI: 10.1002/(SICI)1097-4571(199012)41:8<581::AID-ASI4>3.0.CO;2-U. URL: http://dx.doi.org/10.1002/(SICI)1097-4571(199012)41:8%3C581::AID-ASI4%3E3.0.CO;2-U.

Hazelwood, K., S. Bird, D. Brooks, S. Chintala, U. Diril, D. Dzhulgakov, M. Fawzy, B. Jia, Y. Jia, A. Kalro, J. Law, K. Lee, J. Lu, P. Noordhuis, M. Smelyanskiy, L. Xiong, and X. Wang. 2018. "Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective". In: *2018 IEEE International Symposium on High Performance Computer Architecture, HPCA 2018, Vienna, Austria, February 24-88, 2018.* 10.

He, B. and I. Ounis. 2006. "Query Performance Prediction". *Inf. Syst.* 31(7): 585–594. ISSN: 0306-4379. DOI: 10.1016/j.is.2005.11.003. URL: http://dx.doi.org/10.1016/j.is.2005.11.003.

He, X., J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, and J. Q. Candela. 2014. "Practical Lessons from Predicting Clicks on Ads at Facebook". In: *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. ADKDD'14.* New York, NY, USA: ACM. 5:1–5:9. ISBN: 978-1-4503-2999-6. DOI: 10.1145/2648584.2648589. URL: http://doi.acm.org/10.1145/2648584.2648589.

Heaps, H. S. 1978. *Information Retrieval: Computational and Theoretical Aspects.* Orlando, FL, USA: Academic Press, Inc. ISBN: 0123357500.

Hennessy, J. L. and D. A. Patterson. 2011. *Computer Architecture, Fifth Edition: A Quantitative Approach.* 5th. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN: 012383872X.

Järvelin, K. and J. Kekäläinen. 2002. "Cumulated Gain-based Evaluation of IR Techniques". *ACM Trans. Inf. Syst.* 20(4): 422–446. ISSN: 1046-8188. DOI: 10.1145/582415.582418. URL: http://doi.acm.org/10.1145/582415.582418.

Jeon, M., Y. He, S. Elnikety, A. L. Cox, and S. Rixner. 2013. "Adaptive Parallelism for Web Search". In: *Proceedings of the 8th ACM European Conference on Computer Systems. EuroSys '13*. Prague, Czech Republic: ACM. 155–168. ISBN: 978-1-4503-1994-2. DOI: 10.1145/2465351.2465367. URL: http://doi.acm.org/10.1145/2465351.2465367.

Jeon, M., S. Kim, S.-w. Hwang, Y. He, S. Elnikety, A. L. Cox, and S. Rixner. 2014. "Predictive Parallelization: Taming Tail Latencies in Web Search". In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. SIGIR '14*. Gold Coast, Queensland, Australia: ACM. 253–262. ISBN: 978-1-4503-2257-7. DOI: 10.1145/2600428.2609572. URL: http://doi.acm.org/10.1145/2600428.2609572.

Jia, X.-F., A. Trotman, and R. O'Keefe. 2010. "Efficient Accumulator initialisation". In: *ADCS*. 44–51.

Jin, X., T. Yang, and X. Tang. 2016. "A Comparison of Cache Blocking Methods for Fast Execution of Ensemble-based Score Computation". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '16*. Pisa, Italy: ACM. 629–638. ISBN: 978-1-4503-4069-4. DOI: 10.1145/2911451.2911520. URL: http://doi.acm.org/10.1145/2911451.2911520.

Joachims, T. 2002. "Optimizing Search Engines Using Clickthrough Data". In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '02*. Edmonton, Alberta, Canada: ACM. 133–142. ISBN: 1-58113-567-X. DOI: 10.1145/775047.775067. URL: http://doi.acm.org/10.1145/775047.775067.

Jonassen, S. and S. E. Bratsberg. 2011. "Efficient Compressed Inverted Index Skipping for Disjunctive Text-Queries". In: *Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings*. Ed. by P. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, and V. Mudoch. Berlin, Heidelberg: Springer Berlin Heidelberg. 530–542. ISBN: 978-3-642-20161-5. DOI: 10.1007/978-3-642-20161-5_53. URL: http://dx.doi.org/10.1007/978-3-642-20161-5_53.

Jones, R., B. Rey, O. Madani, and W. Greiner. 2006. "Generating Query Substitutions". In: *WWW*. 387–396.

Kane, A. and F. W. Tompa. 2018. "Split-Lists and Initial Thresholds for WAND-based Search". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. SIGIR '18*. Ann Arbor, MI, USA: ACM. 877–880. ISBN: 978-1-4503-5657-2. DOI: 10.1145/3209978.3210066. URL: http://doi.acm.org/10.1145/3209978.3210066.

Kaszkiel, M. and J. Zobel. 1998. "Term-ordered Query Evaluation Versus Document-ordered Query Evaluation for Large Document Databases". In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '98*. Melbourne, Australia: ACM. 343–344. ISBN: 1-58113-015-5. DOI: 10.1145/290941.291031. URL: http://doi.acm.org/10.1145/290941.291031.

Kaszkiel, M., J. Zobel, and R. Sacks-Davis. 1999. "Efficient Passage Ranking for Document Databases". *ACM Trans. Inf. Syst.* 17(4): 406–439. ISSN: 1046-8188. DOI: 10.1145/326440.326445. URL: http://doi.acm.org/10.1145/326440.326445.

Kayaaslan, E., B. B. Cambazoglu, R. Blanco, F. P. Junqueira, and C. Aykanat. 2011. "Energy-price-driven Query Processing in Multi-center Web Search Engines". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11*. Beijing, China: ACM. 983–992. ISBN: 978-1-4503-0757-4. DOI: 10.1145/2009916.2010047. URL: http://doi.acm.org/10.1145/2009916.2010047.

Kim, S., Y. He, S.-w. Hwang, S. Elnikety, and S. Choi. 2015. "Delayed-Dynamic-Selective (DDS) Prediction for Reducing Extreme Tail Latency in Web Search". In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. WSDM '15*. Shanghai, China: ACM. 7–16. ISBN: 978-1-4503-3317-7. DOI: 10.1145/2684822.2685289. URL: http://doi.acm.org/10.1145/2684822.2685289.

Kim, Y. and J. Lin. 2018. "Serverless Data Analytics with Flint". In: *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*. 451–455. DOI: 10.1109/CLOUD.2018.00063.

Kohavi, R. 1994. "Bottom-Up Induction of Oblivious Read-Once Decision Graphs: Strengths and Limitations". In: *Proceedings of the 12th National Conference on Artificial Intelligence, (AAAI)*. 613–618. URL: http://www.aaai.org/Library/AAAI/1994/aaai94-093.php.

Kyriazis, D., C. Doulkeridis, P. Gouvas, R. Jimenez-Peris, A. J. Ferrer, L. Kallipolitis, P. Kranas, G. Kousiouris, C. Macdonald, R. McCreadie, *et al.* 2018. "BigDataStack: A Holistic Data-Driven Stack for Big Data Applications and Operations". In: *2018 IEEE International Congress on Big Data (BigData Congress)*. IEEE. 237–241.

Lacour, P., C. Macdonald, and I. Ounis. 2008. "Efficiency Comparison of Document Matching Techniques". In: *Proceedings of the Efficiency Issues in Information Retrieval Workshop at ECIR 2008*. Ed. by R. Blanco and F. Silvestri. Glasgow, UK: Dept of Computing Science, University of Glasgow.

Lafferty, J. and C. Zhai. 2001. "Document Language Models, Query Models, and Risk Minimization for Information Retrieval". In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '01*. New Orleans, Louisiana, USA: ACM. 111–119. ISBN: 1-58113-331-6. DOI: 10.1145/383952.383970. URL: http://doi.acm.org/10.1145/383952.383970.

Langley, P. and S. Sage. 1994. "Oblivious decision trees and abstract cases". In: *Working Notes of the AAAI-94 Workshop on Case-Based Reasoning*. AAAI Press. 113–117.

Lemire, D. and L. Boytsov. 2015. "Decoding Billions of Integers Per Second Through Vectorization". *Softw. Pract. Exper.* 45(1): 1–29. ISSN: 0038-0644. DOI: 10.1002/spe.2203. URL: http://dx.doi.org/10.1002/spe.2203.

Lester, N., A. Moffat, W. Webber, and J. Zobel. 2005. "Space-Limited Ranked Query Evaluation Using Adaptive Pruning". In: *Proceedings of the 6th International Conference on Web Information Systems Engineering (WISE 05)*. Ed. by A. H. H. Ngu, M. Kitsuregawa, E. J. Neuhold, J.-Y. Chung, and Q. Z. Sheng. Vol. 3806. *Lecture Notes in Computer Science*. New York, NY, USA: Springer. 470–477.

Lettich, F., C. Lucchese, F. M. Nardini, S. Orlando, R. Perego, N. Tonellotto, and R. Venturini. 2018. "Parallel Traversal of Large Ensembles of Decision Trees". In: 14. DOI: 10.1109/TPDS.2018.2860982.

Li, P., C. J. C. Burges, and Q. Wu. 2007. "McRank: Learning to Rank Using Multiple Classification and Gradient Boosting". In: *Proceedings of the 20th International Conference on Neural Information Processing Systems. NIPS'07*. Vancouver, British Columbia, Canada: Curran Associates Inc. 897–904. ISBN: 978-1-60560-352-0. URL: http://dl.acm.org/citation.cfm?id=2981562.2981675.

Lillis, D., F. Toolan, R. Collier, and J. Dunnion. 2006. "ProbFuse: A Probabilistic Approach to Data Fusion". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '06*. Seattle, Washington, USA: ACM. 139–146. ISBN: 1-59593-369-7. DOI: 10.1145/1148170.1148197. URL: http://doi.acm.org/10.1145/1148170.1148197.

Lin, J. 2015. "Building a Self-Contained Search Engine in the Browser". In: *Proceedings of the 2015 International Conference on The Theory of Information Retrieval. ICTIR '15*. Northampton, Massachusetts, USA: ACM. 309–312. ISBN: 978-1-4503-3833-2. DOI: 10.1145/2808194.2809478. URL: http://doi.acm.org/10.1145/2808194.2809478.

Lin, J. and A. Trotman. 2015. "Anytime Ranking for Impact-Ordered Indexes". In: *Proceedings of the 2015 International Conference on The Theory of Information Retrieval. ICTIR '15*. Northampton, Massachusetts, USA: ACM. 301–304. ISBN: 978-1-4503-3833-2. DOI: 10.1145/2808194.2809477. URL: http://doi.acm.org/10.1145/2808194.2809477.

Lin, J. and A. Trotman. 2017. "The role of index compression in score-at-a-time query evaluation". *Information Retrieval Journal*: 1–22. ISSN: 1573-7659. DOI: 10.1007/s10791-016-9291-5. URL: http://dx.doi.org/10.1007/s10791-016-9291-5.

Ling, X., W. Deng, C. Gu, H. Zhou, C. Li, and F. Sun. 2017. "Model Ensemble for Click Prediction in Bing Search Ads". In: *Proceedings of the 26th International Conference on World Wide Web Companion. WWW '17 Companion*. Perth, Australia: International World Wide Web Conferences Steering Committee. 689–698. ISBN: 978-1-4503-4914-7. DOI: 10.1145/3041021.3054192. URL: https://doi.org/10.1145/3041021.3054192.

Liu, T.-Y. 2009. "Learning to Rank for Information Retrieval". *Foundations and Trends in Information Retrieval*. 3(3): 225–331.

Liu, X., Z. Zhang, B. Hou, R. J. Stones, G. Wang, and X. Liu. 2018. "Index Compression for BitFunnel Query Processing". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. SIGIR '18*. Ann Arbor, MI, USA: ACM. 921–924. ISBN: 978-1-4503-5657-2. DOI: 10.1145/3209978.3210086. URL: http://doi.acm.org/10.1145/3209978.3210086.

Lo, D., L. Cheng, R. Govindaraju, L. A. Barroso, and C. Kozyrakis. 2014. "Towards Energy Proportionality for Large-scale Latency-critical Workloads". In: *Proc. ISCA*. 301–312.

Lu, X., A. Moffat, and J. S. Culpepper. 2015. "On the Cost of Extracting Proximity Features for Term-Dependency Models". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. CIKM '15*. Melbourne, Australia: ACM. 293–302. ISBN: 978-1-4503-3794-6. DOI: 10.1145/2806416.2806467. URL: http://doi.acm.org/10.1145/2806416.2806467.

Lucarella, D. 1988. "A document retrieval system based on nearest neighbour searching". *Journal of Information Science*. 14(1): 25–33.

Lucchese, C., F. M. Nardini, S. Orlando, R. Perego, and N. Tonellotto. 2015a. "Speeding Up Document Ranking with Rank-based Features". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '15*. Santiago, Chile: ACM. 895–898. ISBN: 978-1-4503-3621-5. DOI: 10.1145/2766462.2767776. URL: http://doi.acm.org/10.1145/2766462.2767776.

Lucchese, C., F. M. Nardini, S. Orlando, R. Perego, N. Tonellotto, and R. Venturini. 2015b. "QuickScorer: A Fast Algorithm to Rank Documents with Additive Ensembles of Regression Trees". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '15*. Santiago, Chile: ACM. 73–82. ISBN: 978-1-4503-3621-5. DOI: 10.1145/2766462.2767733. URL: http://doi.acm.org/10.1145/2766462.2767733.

Lucchese, C., F. M. Nardini, S. Orlando, R. Perego, N. Tonellotto, and R. Venturini. 2016. "Exploiting CPU SIMD Extensions to Speed-up Document Scoring with Tree Ensembles". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '16*. Pisa, Italy: ACM. 833–836. ISBN: 978-1-4503-4069-4. DOI: 10.1145/2911451.2914758. URL: http://doi.acm.org/10.1145/2911451.2914758.

Macdonald, C., R. McCreadie, and I. Ounis. 2018. "Agile Information Retrieval Experimentation with Terrier Notebooks". In: *Proceedings of DESIRES 2018*. URL: http://eprints.gla.ac.uk/165871/.

Macdonald, C., R. McCreadie, R. Santos, and I. Ounis. 2012a. "From Puppy to Maturity: Experiences in Developing Terrier". In: *Proceedings of the SIGIR 2012 Workshop in Open Source Information Retrieval*. Portland, OR.

Macdonald, C., I. Ounis, and N. Tonellotto. 2011. "Upper-bound Approximations for Dynamic Pruning". *ACM Trans. Inf. Syst.* 29(4): 17:1–17:28. ISSN: 1046-8188. DOI: 10.1145/2037661.2037662. URL: http://doi.acm.org/10.1145/2037661.2037662.

Macdonald, C., R. L. Santos, and I. Ounis. 2013a. "The Whens and Hows of Learning to Rank for Web Search". *Inf. Retr.* 16(5): 584–628. ISSN: 1386-4564. DOI: 10.1007/s10791-012-9209-9. URL: http://dx.doi.org/10.1007/s10791-012-9209-9.

Macdonald, C., R. L. Santos, and I. Ounis. 2012b. "On the Usefulness of Query Features for Learning to Rank". In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management. CIKM '12*. Maui, Hawaii, USA: ACM. 2559–2562. ISBN: 978-1-4503-1156-4. DOI: 10.1145/2396761.2398691. URL: http://doi.acm.org/10.1145/2396761.2398691.

Macdonald, C., R. L. Santos, I. Ounis, and B. He. 2013b. "About Learning Models with Multiple Query-dependent Features". *ACM Trans. Inf. Syst.* 31(3): 11:1–11:39. ISSN: 1046-8188. DOI: 10.1145/2493175.2493176. URL: http://doi.acm.org/10.1145/2493175.2493176.

Macdonald, C. and N. Tonellotto. 2017. "Upper Bound Approximations for BlockMaxWand". In: *Proceedings of the* 3^{rd} *ACM International Conference on the Theory of Information Retrieval (ICTIR 2017)*. Tokyo, Japan: ACM. 10. ISBN: 9781450344906. DOI: 10.1145/3121050.3121094. URL: https://doi.org/10.1145/3121050.3121094.

Macdonald, C., N. Tonellotto, and I. Ounis. 2012c. "Effect of Dynamic Pruning Safety on Learning to Rank Effectiveness". In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 1051–1052. DOI: 10.1145/2348283.2348464.

Macdonald, C., N. Tonellotto, and I. Ounis. 2012d. "Learning to Predict Response Times for Online Query Scheduling". In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '12*. Portland, Oregon, USA: ACM. 621–630. ISBN: 978-1-4503-1472-5. DOI: 10.1145/2348283.2348367. URL: http://doi.acm.org/10.1145/2348283.2348367.

Macdonald, C., N. Tonellotto, and I. Ounis. 2017. "Efficient & Effective Selective Query Rewriting with Efficiency Predictions". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17*. Shinjuku, Tokyo, Japan: ACM. 495–504. ISBN: 978-1-4503-5022-8. DOI: 10.1145/3077136.3080827. URL: http://doi.acm.org/10.1145/3077136.3080827.

Mackenzie, J., J. S. Culpepper, R. Blanco, M. Crane, C. L. A. Clarke, and J. Lin. 2018. "Query Driven Algorithm Selection in Early Stage Retrieval". In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. WSDM '18*. Marina Del Rey, CA, USA: ACM. 396–404. ISBN: 978-1-4503-5581-0. DOI: 10.1145/3159652.3159676. URL: http://doi.acm.org/10.1145/3159652.3159676.

Mackenzie, J., F. Scholer, and J. S. Culpepper. 2017. "Early Termination Heuristics for Score-at-a-Time Index Traversal". In: *Proceedings of the 22Nd Australasian Document Computing Symposium. ADCS 2017*. Brisbane, QLD, Australia: ACM. 8:1–8:8. ISBN: 978-1-4503-6391-4. DOI: 10.1145/3166072.3166073. URL: http://doi.acm.org/10.1145/3166072.3166073.

Mallia, A., G. Ottaviano, E. Porciani, N. Tonellotto, and R. Venturini. 2017. "Faster BlockMax WAND with Variable-sized Blocks". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17*. Shinjuku, Tokyo, Japan: ACM. 625–634. ISBN: 978-1-4503-5022-8. DOI: 10.1145/3077136.3080780. URL: http://doi.acm.org/10.1145/3077136.3080780.

Manmatha, R., T. Rath, and F. Feng. 2001. "Modeling Score Distributions for Combining the Outputs of Search Engines". In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '01*. New Orleans, Louisiana, USA: ACM. 267–275. ISBN: 1-58113-331-6. DOI: 10.1145/383952.384005. URL: http://doi.acm.org/10.1145/383952.384005.

Manning, C. D., P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press. ISBN: 0521865719.

Maron, M. E. and J. L. Kuhns. 1960. "On Relevance, Probabilistic Indexing and Information Retrieval". *J. ACM*. 7(3): 216–244. ISSN: 0004-5411. DOI: 10.1145/321033.321035. URL: http://doi.acm.org/10.1145/321033.321035.

McCreadie, R., D. Albakour, J. Manotumruksa, C. Macdonald, and I. Ounis. 2016. "Searching the Internet of Things". *Building Blocks for IoT Analytics*: 39–80.

Metzler, D. and W. Bruce Croft. 2007. "Linear feature-based models for information retrieval". *Information Retrieval*. 10(3): 257–274. ISSN: 1573-7659. DOI: 10.1007/s10791-006-9019-z. URL: https://doi.org/10.1007/s10791-006-9019-z.

Metzler, D. and W. B. Croft. 2005. "A Markov Random Field Model for Term Dependencies". In: *SIGIR*. 472–479.

Mitra, B., F. Diaz, and N. Craswell. 2017. "Learning to Match Using Local and Distributed Representations of Text for Web Search". In: *Proceedings of the 26th International Conference on World Wide Web*. *WWW '17*. Perth, Australia: International World Wide Web Conferences Steering Committee. 1291–1299. ISBN: 978-1-4503-4913-0. DOI: 10.1145/3038912.3052579. URL: https://doi.org/10.1145/3038912.3052579.

Moffat, A. and J. Zobel. 1994. "Fast Ranking in Limited Space". In: *Proc. IEEE Conf. on Data Engineering*. 428–437.

Moffat, A. and J. Zobel. 1996. "Self-Indexing Inverted Files for Fast Text Retrieval". *Transactions on Information Systems*. 14(4): 349–379.

Moffat, A., J. Zobel, and R. Sacks-Davis. 1994. "Memory Efficient Ranking". *Inf. Process. Manage.* 30(6): 733–744. ISSN: 0306-4573. DOI: 10.1016/0306-4573(94)90002-7. URL: http://dx.doi.org/10.1016/0306-4573(94)90002-7.

Montague, M. and J. A. Aslam. 2001. "Relevance Score Normalization for Metasearch". In: *Proceedings of the Tenth International Conference on Information and Knowledge Management*. *CIKM '01*. Atlanta, Georgia, USA: ACM. 427–433. ISBN: 1-58113-436-3. DOI: 10.1145/502585.502657. URL: http://doi.acm.org/10.1145/502585.502657.

Montague, M. and J. A. Aslam. 2002. "Condorcet Fusion for Improved Retrieval". In: *Proceedings of the Eleventh International Conference on Information and Knowledge Management*. *CIKM '02*. McLean, Virginia, USA: ACM. 538–548. ISBN: 1-58113-492-4. DOI: 10.1145/584792.584881. URL: http://doi.acm.org/10.1145/584792.584881.

Nenkova, A. and K. McKeown. 2011. "Automatic Summarization". *Foundations and Trends® in Information Retrieval*. 5(2–3): 103–233. ISSN: 1554-0669. DOI: 10.1561/1500000015. URL: http://dx.doi.org/10.1561/1500000015.

Noreault, T., M. Koll, and M. J. McGill. 1977. "Automatic ranked output from boolean searches in SIRE". *Journal of the American Society for Information Science*. 28(6): 333–339. ISSN: 1097-4571. DOI: 10.1002/asi.4630280605. URL: http://dx.doi.org/10.1002/asi.4630280605.

Ottaviano, G., N. Tonellotto, and R. Venturini. 2015. "Optimal Space-time Tradeoffs for Inverted Indexes". In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. *WSDM '15*. Shanghai, China: ACM. 47–56. ISBN: 978-1-4503-3317-7. DOI: 10.1145/2684822.2685297. URL: http://doi.acm.org/10.1145/2684822.2685297.

Ottaviano, G. and R. Venturini. 2014. "Partitioned Elias-Fano Indexes". In: *SIGIR*. 273–282.

Pang, L., Y. Lan, J. Guo, J. Xu, J. Xu, and X. Cheng. 2017. "DeepRank: A New Deep Architecture for Relevance Ranking in Information Retrieval". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. *CIKM '17*. Singapore, Singapore: ACM. 257–266. ISBN: 978-1-4503-4918-5. DOI: 10.1145/3132847.3132914. URL: http://doi.acm.org/10.1145/3132847.3132914.

Pedersen, J. 2010. "Query understanding at Bing". In: *Proceedings of SIGIR 2010*.

Peng, F., N. Ahmed, X. Li, and Y. Lu. 2007. "Context Sensitive Stemming for Web Search". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. *SIGIR '07*. Amsterdam, The Netherlands: ACM. 639–646. ISBN: 978-1-59593-597-7. DOI: 10.1145/1277741.1277851. URL: http://doi.acm.org/10.1145/1277741.1277851.

Perry, S. A. and P. Willett. 1983. "A Review of the Use of Inverted Files for Best Match Searching in Information Retrieval Systems". *Journal of Information Science*. 6(2–3): 59–66.

Persin, M. 1994. "Document filtering for fast ranking". In: *SIGIR '94: Proceedings of the 17th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Ed. by W. B. Croft and C. J. van Rijsbergen. New York, NY, USA: ACM.

Persin, M., J. Zobel, and R. Sacks-Davis. 1996. "Filtered Document Retrieval with Frequency-sorted Indexes". *J. Am. Soc. Inf. Sci.* 47(10): 749–764. ISSN: 0002-8231. DOI: 10.1002/(SICI)1097-4571(199610)47:10<749::AID-ASI3>3.3.CO;2-U. URL: http://dx.doi.org/10.1002/(SICI)1097-4571(199610)47:10%3C749::AID-ASI3%3E3.3.CO;2-U.

Petri, M., J. S. Culpepper, and A. Moffat. 2013. "Exploring the Magic of WAND". In: *Proceedings of the 18th Australasian Document Computing Symposium. ADCS '13*. Brisbane, Queensland, Australia: ACM. 58–65. ISBN: 978-1-4503-2524-0. DOI: 10.1145/2537734.2537744. URL: http://doi.acm.org/10.1145/2537734.2537744.

Ponte, J. M. and W. B. Croft. 1998. "A Language Modeling Approach to Information Retrieval". In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '98*. Melbourne, Australia: ACM. 275–281. ISBN: 1-58113-015-5. DOI: 10.1145/290941.291008. URL: http://doi.acm.org/10.1145/290941.291008.

Pugh, W. 1990. "Skip Lists: A Probabilistic Alternative to Balanced Trees". *Commun. ACM*. 33(6): 668–676. ISSN: 0001-0782. DOI: 10.1145/78973.78977. URL: http://doi.acm.org/10.1145/78973.78977.

Putnam, A., A. M. Caulfield, E. S. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. P. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J.-Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Y. Xiao, and D. Burger. 2014. "A Reconfigurable Fabric for Accelerating Large-scale Datacenter Services". In: *Proceeding of the 41st Annual International Symposium on Computer Architecuture. ISCA '14*. Minneapolis, Minnesota, USA: IEEE Press. 13–24. ISBN: 978-1-4799-4394-4. URL: http://dl.acm.org/citation.cfm?id=2665671.2665678.

Qin, T., T.-Y. Liu, J. Xu, and H. Li. 2010. "LETOR: A benchmark collection for research on learning to rank for information retrieval". *Information Retrieval*. 13(4): 346–374. ISSN: 1573-7659. DOI: 10.1007/s10791-009-9123-y.

Ramaswamy, V., R. Konow, A. Trotman, J. Degenhardt, and N. Whyte. 2017. "Document Reordering is Good, Especially for e-Commerce". In: *Proceedings of SIGIR 2017 Workshop on eCommerce*.

Rice, R. and J. Plaunt. 1971. "Adaptive Variable-Length Coding for Efficient Compression of Spacecraft Television Data". *Communication Technology, IEEE Transactions on*. 19(6): 889–897. ISSN: 0018-9332. DOI: 10.1109/TCOM.1971.1090789.

Rigutini, L., T. Papini, M. Maggini, and F. Scarselli. 2011. "SortNet: Learning to Rank by a Neural Preference Function". *Trans. Neur. Netw.* 22(9): 1368–1380. ISSN: 1045-9227. DOI: 10.1109/TNN.2011.2160875. URL: http://dx.doi.org/10.1109/TNN.2011.2160875.

Rijsbergen, C. van. 1979. *Information Retrieval, 2nd edition*. Butterworths, London. URL: citeseer.ist.psu.edu/vanrijsbergen79information.html.

Risvik, K. M., Y. Aasheim, and M. Lidal. 2003. "Multi-Tier Architecture for Web Search Engines". In: *LA-WEB*. IEEE Computer Society. 132–143. DOI: 10.1109/LAWEB.2003.1250291.

Risvik, K. M., T. Chilimbi, H. Tan, K. Kalyanaraman, and C. Anderson. 2013. "Maguro, a System for Indexing and Searching over Very Large Text Collections". In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. WSDM '13*. Rome, Italy: ACM. 727–736. ISBN: 978-1-4503-1869-3. DOI: 10.1145/2433396.2433486. URL: http://doi.acm.org/10.1145/2433396.2433486.

Robertson, S. E., S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. 1994. "Okapi at TREC-3". In: *Proceedings of The 3rd Text REtrieval Conference*. National Institute of Standards and Technology (NIST). 109–126.

Rojas, O., V. Gil-Costa, and M. Marin. 2013a. "Distributing Efficiently the Block-Max WAND Algorithm". *Procedia Computer Science*. 18: 120–129. ISSN: 1877-0509. DOI: http://dx.doi.org/10.1016/j.procs.2013.05.175. URL: http://www.sciencedirect.com/science/article/pii/S1877050913003189.

Rojas, O., V. Gil-Costa, and M. Marin. 2013b. "Efficient Parallel Block-Max WAND Algorithm". In: *Euro-Par 2013 Parallel Processing: 19th International Conference, Aachen, Germany, August 26-30, 2013. Proceedings*. Ed. by F. Wolf, B. Mohr, and D. an Mey. Berlin, Heidelberg: Springer Berlin Heidelberg. 394–405. ISBN: 978-3-642-40047-6. DOI: 10.1007/978-3-642-40047-6_41. URL: http://dx.doi.org/10.1007/978-3-642-40047-6_41.

Rossi, C., E. S. de Moura, A. L. Carvalho, and A. S. da Silva. 2013. "Fast Document-at-a-time Query Processing Using Two-tier Indexes". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '13*. Dublin, Ireland: ACM. 183–192. ISBN: 978-1-4503-2034-4. DOI: 10.1145/2484028.2484085. URL: http://doi.acm.org/10.1145/2484028.2484085.

Sadakane, K. and H. Imai. 1999. "Text retrieval by using k-word proximity search". In: *Proceedings 1999 International Symposium on Database Applications in Non-Traditional Environments (DANTE'99) (Cat. No.PR00496)*. 183–188. DOI: 10.1109/DANTE.1999.844958.

Salton, G., A. Wong, and C. S. Yang. 1975. "A Vector Space Model for Automatic Indexing". *Commun. ACM*. 18(11): 613–620. ISSN: 0001-0782. DOI: 10.1145/361219.361220. URL: http://doi.acm.org/10.1145/361219.361220.

Salton, G. 1989. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc. ISBN: 0-201-12227-8.

Scholer, F., H. E. Williams, J. Yiannis, and J. Zobel. 2002. "Compression of Inverted Indexes For Fast Query Evaluation". In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '02*. Tampere, Finland: ACM. 222–229. ISBN: 1-58113-561-0. DOI: 10.1145/564376.564416. URL: http://doi.acm.org/10.1145/564376.564416.

Shan, D., S. Ding, J. He, H. Yan, and X. Li. 2012. "Optimized Top-k Processing with Global Page Scores on Block-max Indexes". In: *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. WSDM '12*. Seattle, Washington, USA: ACM. 423–432. ISBN: 978-1-4503-0747-5. DOI: 10.1145/2124295.2124346. URL: http://doi.acm.org/10.1145/2124295.2124346.

Sharp, T. 2008. "Implementing decision trees and forests on a GPU". In: *Proc. Computer Vision*. Springer. 595–608.

Shieh, W.-Y., T.-F. Chen, J. J.-J. Shann, and C.-P. Chung. 2003. "Inverted File Compression Through Document Identifier Reassignment". *Inf. Process. Manage.* 39(1): 117–131. ISSN: 0306-4573. DOI: 10.1016/S0306-4573(02)00020-1. URL: http://dx.doi.org/10.1016/S0306-4573(02)00020-1.

Silverstein, C., H. Marais, M. Henzinger, and M. Moricz. 1999. "Analysis of a Very Large Web Search Engine Query Log". *SIGIR Forum*. 33(1): 6–12. ISSN: 0163-5840. DOI: 10.1145/331403.331405. URL: http://doi.acm.org/10.1145/331403.331405.

Silvestri, F. 2007. "Sorting out the Document Identifier Assignment Problem". In: *Proceedings of the 29th European Conference on IR Research. ECIR'07*. Rome, Italy: Springer-Verlag. 101–112. ISBN: 978-3-540-71494-1. URL: http://dl.acm.org/citation.cfm?id=1763653.1763668.

Silvestri, F., S. Orlando, and R. Perego. 2004. "Assigning Identifiers to Documents to Enhance the Clustering Property of Fulltext Indexes". In: *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '04*. Sheffield, United Kingdom: ACM. 305–312. ISBN: 1-58113-881-4. DOI: 10.1145/1008992.1009046. URL: http://doi.acm.org/10.1145/1008992.1009046.

Smeaton, A. F. and C. J. van Rijsbergen. 1981. "The Nearest Neighbour Problem in Information Retrieval: An Algorithm Using Upperbounds". In: *Proceedings of the 4th Annual International ACM SIGIR Conference on Information Storage and Retrieval: Theoretical Issues in Information Retrieval. SIGIR '81*. Oakland, California: ACM. 83–87. ISBN: 0-89791-052-4. DOI: 10.1145/511754.511767. URL: http://doi.acm.org/10.1145/511754.511767.

Snowdon, D. C., S. Ruocco, and G. Heiser. 2005. "Power Management and Dynamic Voltage Scaling: Myths and Facts". In: *Proc. Workshop on Power Aware Real-time Computing*.

Sorokina, D. and E. Cantú-Paz. 2016. "Amazon Search: The Joy of Ranking Products". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*. 459–460. DOI: 10.1145/2911451.2926725. URL: http://doi.acm.org/10.1145/2911451.2926725.

Stepanov, A. A., A. R. Gangolli, D. E. Rose, R. J. Ernst, and P. S. Oberoi. 2011. "SIMD-based Decoding of Posting Lists". In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management. CIKM '11*. Glasgow, Scotland, UK: ACM. 317–326. ISBN: 978-1-4503-0717-8. DOI: 10.1145/2063576.2063627. URL: http://doi.acm.org/10.1145/2063576.2063627.

Strohman, T. 2007. "Efficient Processing of Complex Features for Information Retrieval". *IR*. University of Massachusetts Amherst.

Strohman, T. and W. B. Croft. 2006. "Low Latency Index Maintenance in Indri". In: *Proceedings of the Open Source Information Retrieval Workshop*. 7–11.

Strohman, T. and W. B. Croft. 2007. "Efficient Document Retrieval in Main Memory". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '07*. Amsterdam, The Netherlands: ACM. 175–182. ISBN: 978-1-59593-597-7. DOI: 10.1145/1277741.1277774. URL: http://doi.acm.org/10.1145/1277741.1277774.

Strohman, T., H. Turtle, and W. B. Croft. 2005. "Optimization strategies for complex queries". In: *SIGIR '05: Proceedings of the 28th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM. 219–225. ISBN: 1-59593-034-5. DOI: http://doi.acm.org/10.1145/1076034.1076074.

Tang, X., X. Jin, and T. Yang. 2014. "Cache-conscious runtime optimization for ranking ensembles." In: *Proc. ACM SIGIR*. 1123–1126.

Tatikonda, S., B. B. Cambazoglu, and F. P. Junqueira. 2011. "Posting List Intersection on Multicore Architectures". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11*. Beijing, China: ACM. 963–972. ISBN: 978-1-4503-0757-4. DOI: 10.1145/2009916.2010045. URL: http://doi.acm.org/10.1145/2009916.2010045.

Teymorian, A., O. Frieder, and M. A. Maloof. 2013. "Rank-energy Selective Query Forwarding for Distributed Search Systems". In: *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management. CIKM '13*. San Francisco, California, USA: ACM. 389–398. ISBN: 978-1-4503-2263-8. DOI: 10.1145/2505515.2505710. URL: http://doi.acm.org/10.1145/2505515.2505710.

Tonellotto, N., C. Macdonald, and I. Ounis. 2010. "Efficient dynamic pruning with proximity support". In: *Proceedings of the 8th Workshop on Large-Scale Distributed Systems for Information Retrieval*. 31–35.

Tonellotto, N., C. Macdonald, and I. Ounis. 2011. "Effect of Different Docid Orderings on Dynamic Pruning Retrieval Strategies". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11*. Beijing, China: ACM. 1179–1180. ISBN: 978-1-4503-0757-4. DOI: 10.1145/2009916.2010108. URL: http://doi.acm.org/10.1145/2009916.2010108.

Tonellotto, N., C. Macdonald, and I. Ounis. 2013. "Efficient and Effective Retrieval Using Selective Pruning". In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. WSDM '13*. Rome, Italy: ACM. 63–72. ISBN: 978-1-4503-1869-3. DOI: 10.1145/2433396.2433407. URL: http://doi.acm.org/10.1145/2433396.2433407.

Trofimov, I., A. Kornetova, and V. Topinskiy. 2012. "Using Boosted Trees for Click-through Rate Prediction for Sponsored Search". In: *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy. ADKDD '12*. Beijing, China: ACM. 2:1–2:6. ISBN: 978-1-4503-1545-6. DOI: 10.1145/2351356.2351358. URL: http://doi.acm.org/10.1145/2351356.2351358.

Trotman, A. 2014. "Compression, SIMD, and Postings Lists". In: *Proceedings of the 2014 Australasian Document Computing Symposium. ADCS '14*. Melbourne, VIC, Australia: ACM. 50:50–50:57. ISBN: 978-1-4503-3000-8. DOI: 10.1145/2682862.2682870. URL: http://doi.acm.org/10.1145/2682862.2682870.

Trotman, A. and J. Lin. 2016. "In Vacuo and In Situ Evaluation of SIMD Codecs". In: *Proceedings of the 21st Australasian Document Computing Symposium. ADCS '16*. Caulfield, VIC, Australia: ACM. 1–8. ISBN: 978-1-4503-4865-2. DOI: 10.1145/3015022.3015023. URL: http://doi.acm.org/10.1145/3015022.3015023.

Tu, Z., M. Li, and J. Lin. 2018. "Pay-Per-Request Deployment of Neural Network Models Using Serverless Architectures". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. New Orleans, Louisiana: Association for Computational Linguistics. 6–10. URL: http://aclweb.org/anthology/N18-5002.

Turtle, H. and J. Flood. 1995. "Query evaluation: strategies and optimizations". *Information Processing and Management*. 31(6): 831–850. ISSN: 0306-4573. DOI: http://dx.doi.org/10.1016/0306-4573(95)00020-H.

Van Essen, B., C. Macaraeg, M. Gokhale, and R. Prenger. 2012. "Accelerating a random forest classifier: Multi-core, GP-GPU, or FPGA?" In: *Proc. IEEE FCCM*. IEEE. 232–239.

Vigna, S. 2013. "Quasi-succinct indices". In: *WSDM*.

Vogt, C. C. and G. W. Cottrell. 1998. "Predicting the Performance of Linearly Combined IR Systems". In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '98*. Melbourne, Australia: ACM. 190–196. ISBN: 1-58113-015-5. DOI: 10.1145/290941.290991. URL: http://doi.acm.org/10.1145/290941.290991.

Wang, J., E. Lo, M. L. Yiu, J. Tong, G. Wang, and X. Liu. 2013. "The Impact of Solid State Drive on Search Engine Cache Management". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '13*. Dublin, Ireland: ACM. 693–702. ISBN: 978-1-4503-2034-4. DOI: 10.1145/2484028.2484046. URL: http://doi.acm.org/10.1145/2484028.2484046.

Wang, L., J. Lin, and D. Metzler. 2010. "Learning to Efficiently Rank". In: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '10*. Geneva, Switzerland: ACM. 138–145. ISBN: 978-1-4503-0153-4. DOI: 10.1145/1835449.1835475. URL: http://doi.acm.org/10.1145/1835449.1835475.

Wang, L., J. Lin, and D. Metzler. 2011. "A Cascade Ranking Model for Efficient Ranked Retrieval". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11*. Beijing, China: ACM. 105–114. ISBN: 978-1-4503-0757-4. DOI: 10.1145/2009916.2009934. URL: http://doi.acm.org/10.1145/2009916.2009934.

White, R. W. 2016. *Interactions with Search Systems*. Cambridge University Press. ISBN: 9781139525305. DOI: 10.1017/CBO9781139525305. URL: https://doi.org/10.1017/CBO9781139525305.

Williams, H. E. and J. Zobel. 1999. "Compressing Integers for Fast File Access". *The Computer Journal*. 42: 193–201.

Witten, I. H., A. Moffat, and T. C. Bell. 1999. *Managing Gigabytes (2Nd Ed.): Compressing and Indexing Documents and Images*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN: 1-55860-570-3.

Wong, W. Y. P. and D. L. Lee. 1993. "Implementations of Partial Document Ranking Using Inverted Files". *Information Processing and Management*. 29(5): 647–669.

Wu, D., F. Zhang, N. Ao, G. Wang, X. Liu, and J. Liu. 2010. "Efficient lists intersection by CPU-GPU cooperative computing". In: *2010 IEEE International Symposium on Parallel Distributed Processing, Workshops and Phd Forum (IPDPSW)*. 1–8. DOI: 10.1109/IPDPSW.2010.5470886.

Wu, H. and H. Fang. 2014. "Analytical Performance Modeling for Top-K Query Processing". In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. CIKM '14*. Shanghai, China: ACM. 1619–1628. ISBN: 978-1-4503-2598-1. DOI: 10.1145/2661829.2661931. URL: http://doi.acm.org/10.1145/2661829.2661931.

Xu, J. and H. Li. 2007. "AdaRank: A Boosting Algorithm for Information Retrieval". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '07*. Amsterdam, The Netherlands: ACM. 391–398. ISBN: 978-1-59593-597-7. DOI: 10.1145/1277741.1277809. URL: http://doi.acm.org/10.1145/1277741.1277809.

Yan, H., S. Ding, and T. Suel. 2009a. "Compressing Term Positions in Web Indexes". In: *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '09*. Boston, MA, USA: ACM. 147–154. ISBN: 978-1-60558-483-6. DOI: 10.1145/1571941.1571969. URL: http://doi.acm.org/10.1145/1571941.1571969.

Yan, H., S. Ding, and T. Suel. 2009b. "Inverted Index Compression and Query Processing with Optimized Document Ordering". In: *Proceedings of the 18th International Conference on World Wide Web. WWW '09*. Madrid, Spain: ACM. 401–410. ISBN: 978-1-60558-487-4. DOI: 10.1145/1526709.1526764. URL: http://doi.acm.org/10.1145/1526709.1526764.

Yin, D., Y. Hu, J. Tang, T. Daly, M. Zhou, H. Ouyang, J. Chen, C. Kang, H. Deng, C. Nobata, J.-M. Langlois, and Y. Chang. 2016. "Ranking Relevance in Yahoo Search". In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16*. San Francisco, California, USA: ACM. 323–332. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939677. URL: http://doi.acm.org/10.1145/2939672.2939677.

Zobel, J. and A. Moffat. 2006. "Inverted files for text search engines". *ACM Computing Surveys*. 38(2). ISSN: 0360-0300. DOI: 10.1145/1132956.1132959. URL: http://dx.doi.org/10.1145/1132956.1132959.

Zobel, J., A. Moffat, and K. Ramamohanarao. 1998. "Inverted Files Versus Signature Files for Text Indexing". *ACM Trans. Database Syst.* 23(4): 453–490. ISSN: 0362-5915. DOI: 10.1145/296854.277632. URL: http://doi.acm.org/10.1145/296854.277632.

Zukowski, M., S. Heman, N. Nes, and P. Boncz. 2006. "Super-Scalar RAM-CPU Cache Compression". In: *Proceedings of the 22Nd International Conference on Data Engineering. ICDE '06.* Washington, DC, USA: IEEE Computer Society. 59–70. ISBN: 0-7695-2570-9. DOI: 10.1109/ICDE.2006.150. URL: http://dx.doi.org/10.1109/ICDE.2006.150.