Search Result Diversification

Rodrygo L. T. Santos
Department of Computer Science
Universidade Federal de Minas Gerais
rodrygo@dcc.ufmg.br

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

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

now
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

R. L. T. Santos, C. Macdonald and I. Ounis. *Search Result Diversification*. Foundations and Trends® in Information Retrieval, vol. 9, no. 1, pp. 1–90, 2015.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-001-9

Foundations and Trends® in Information Retrieval
Volume 9, Issue 1, 2015

Editorial Board

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, 2015, Volume 9, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Foundations and Trends® in Information Retrieval
Vol. 9, No. 1 (2015) 1–90

DOI: 10.1561/1500000040

now
the essence of knowledge

Search Result Diversification

Rodrygo L. T. Santos
Department of Computer Science
Universidade Federal de Minas Gerais
rodrygo@dcc.ufmg.br

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

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

Contents

Notations

elements

q	A query
a	A relevant query aspect
s	A mined query aspect
d	A document
f	A function (e.g., a ranking function)
r	The rank position of a retrieved document
g_i	The relevance label of the i-th retrieved document

sets

$\mathcal{A}_q$	A set of aspects relevant to a query q
$\mathcal{S}_q$	A set of aspects mined for a query q
$\mathcal{G}_q$	A set of documents relevant for a query q
$\mathcal{R}_q$	A set of documents retrieved for a query q
$\mathcal{D}_q$	A set of documents diversified for a query q

parameters

n	The total number of documents in the corpus
n_q	The number of documents retrieved for the query q
v	The number of unique terms in the corpus
k	The number of aspects underlying a query
κ	An evaluation cutoff
τ	The diversification cutoff
λ	The diversification trade-off

Abstract

Ranking in information retrieval has been traditionally approached as a pursuit of relevant information, under the assumption that the users' information needs are unambiguously conveyed by their submitted queries. Nevertheless, as an inherently limited representation of a more complex information need, every query can arguably be considered ambiguous to some extent. In order to tackle query ambiguity, search result diversification approaches have recently been proposed to produce rankings aimed to satisfy the multiple possible information needs underlying a query. In this survey, we review the published literature on search result diversification. In particular, we discuss the motivations for diversifying the search results for an ambiguous query and provide a formal definition of the search result diversification problem. In addition, we describe the most successful approaches in the literature for producing and evaluating diversity in multiple search domains. Finally, we also discuss recent advances as well as open research directions in the field of search result diversification.

R. L. T. Santos, C. Macdonald and I. Ounis. *Search Result Diversification*. Foundations and Trends® in Information Retrieval, vol. 9, no. 1, pp. 1–90, 2015. DOI: 10.1561/1500000040.

1

Introduction

Queries submitted to an information retrieval (IR) system are often ambiguous to some extent. For instance, a user issuing the query "bond" to an IR system could mean the financial instrument for debt security, the classical crossover string quartet "Bond", or Ian Fleming's secret agent character "James Bond". At the same time, the documents retrieved by an IR system for a given query may convey redundant information. Indeed, a user looking for the IMDb page of the James Bond film "Spectre" may be satisfied after observing just one relevant result. Ambiguity and redundancy have been traditionally ruled out by simplifying modelling assumptions underlying most ranking approaches in IR. Nevertheless, in a realistic search scenario, ambiguity and redundancy may render a traditional relevance-oriented ranking approach suboptimal, in terms of subjecting the user to non-relevant results. In this situation, alternative ranking policies should be considered. In this chapter, we provide a historical perspective of relevance-oriented ranking in IR and discuss the challenges posed by ambiguity and redundancy as a motivation for diversifying the search results.

1.1 The Holy Grail of IR

The key challenge faced by an IR system is to determine the *relevance* of a document given a user's query [Goffman, 1964]. The concept of relevance, the holy grail of IR, has been discussed in the fields of information science and retrieval since the 1950s. Despite the rich literature on the subject, relevance per se is still an ill-understood concept [Mizzaro, 1997]. In a practical environment, relevance can span multiple dimensions, related to the topicality and usefulness of the retrieved documents as they are perceived by the target user [Borlund, 2003]. Indeed, relevance is ultimately a prerogative of the user, in which case an IR system can at best estimate it [Baeza-Yates and Ribeiro-Neto, 2011].

Estimating relevance is a challenging task. Indeed, while current search users may have high expectations regarding the quality of the documents returned by a modern web search engine, they often provide the search engine with a rather limited representation of their information need, in the form of a short keyword-based query [Jansen et al., 2000]. Besides understanding the information needs of a mass of users with varying interests and backgrounds, web search engines must also strive to understand the information available on the Web. In particular, the decentralised nature of content publishing on the Web has led to an unprecedentedly large and heterogeneous repository of information, comprising over 30 trillion uniquely addressable documents [Cutts, 2012] in different languages, writing styles, and with varying degrees of authoritativeness and trustworthiness [Arasu et al., 2001].

The enormous size of the Web most often results in an amount of documents matching a user's query that by far exceeds the very few top ranking positions that the user is normally willing to inspect for relevance [Silverstein et al., 1999]. In such a challenging environment, effectively ranking the returned documents, so that the most relevant documents are presented ahead of less relevant ones, becomes of utmost importance for satisfying the information needs of search users [Baeza-Yates and Ribeiro-Neto, 2011]. A standard boolean retrieval is typically insufficient in a web search scenario, in which case more sophisticated approaches can be deployed to produce a ranking of documents likely to be relevant to the user's information need.

1.2 Relevance-oriented Ranking

Probabilistic ranking approaches have been extensively studied in IR as a mechanism to surface relevant information. Although relevance is an unknown variable to an IR system, properties of a query and of a given document may provide evidence to estimate the probability that the document is relevant to the information need expressed by the query. The probability of relevance of a document to a query is central to the well-known probability ranking principle (PRP) in IR [Cooper, 1971, Robertson, 1977, Robertson and Zaragoza, 2009]:

> *"If a reference retrieval system's response to each request is a ranking of the documents in the collection in order of decreasing probability of relevance to the user who submitted the request, where the probabilities are estimated as accurately as possible on the basis of whatever data have been made available to the system for this purpose, the overall effectiveness of the system to its user will be the best that is obtainable on the basis of those data".*

In practice, as an abstract ranking policy, the PRP does not prescribe how the probability of relevance of a given query-document pair should be estimated. Nonetheless, several probabilistic ranking models have been proposed throughout the years, inspired by the principle. In particular, the literature on probabilistic ranking dates back to 1960, with the seminal work by Maron and Kuhns [1960] on probabilistic indexing and retrieval in a library setting. The field experienced intensive development in the 1970s and 1980s [Cooper, 1971, Harter, 1975a,b, Robertson and Spärck Jones, 1976, Robertson, 1977, Robertson et al., 1981], culminating in some of the most effective ranking functions used by current IR systems [Robertson et al., 1994, 2004, Zaragoza et al., 2004]. Later developments in the field led to effective alternative probabilistic formulations, including statistical language models [Ponte and Croft, 1998, Hiemstra, 1998, Zhai, 2008] and divergence from randomness models [Amati, 2003, 2006].

Despite the relative success attained by the various ranking approaches inspired by the PRP, the development of the principle has

been permeated by simplifying modelling assumptions that are often inconsistent with the underlying data [Gordon and Lenk, 1992, Cooper, 1995]. In particular, Gordon and Lenk [1991, 1992] analysed the optimality of the PRP under the light of classical decision and utility theories [von Neumann and Morgenstern, 1944], based upon the costs involved in not retrieving a relevant document as well as in retrieving a non-relevant one. While decision-theoretic costs remain the same for each retrieved document, the utility-theoretic benefit of a relevant document retrieved depends on the previously retrieved relevant documents. In their analysis, Gordon and Lenk [1991] discussed two key modelling assumptions underlying probabilistic ranking approaches:

A1. The probability of relevance is well-calibrated[1] and estimated with *certainty*, with no associated measure of dispersion.

A2. The probability of relevance of a document is estimated *independently* of the other retrieved documents.

According to A1, a document with a higher probability of relevance should always be ranked ahead of a document with a lower probability of relevance, regardless of the confidence of such probability estimates. According to A2, the probability of relevance of a document should be estimated regardless of the probability of relevance of the documents ranked ahead of it. As Gordon and Lenk [1991] demonstrated, the PRP attains the greatest expected utility compared to any other ranking policy under these two assumptions. However, when at least one of these assumptions fails to hold, the principle is suboptimal. In this case, a strict ordering of the retrieved documents by decreasing probability of relevance may not be advisable, and alternative ranking policies should be considered [Gordon and Lenk, 1992]. In general, neither A1 nor A2 are realistic assumptions. In practice, while A1 is challenged by the occurrence of *ambiguity* in the user's query, A2 is challenged by the occurrence of *redundancy* among the retrieved documents.

[1] According to the definition of Gordon and Lenk [1991], a well-calibrated IR system is one that predicts an accurate probability of relevance for each document.

1.3 Ambiguity and Redundancy

Relevance-oriented ranking approaches assume that the users' information needs are unambiguously conveyed by their submitted queries, and that the users' assessment of relevance for a document does not depend on their perceived relevance for the other documents. While such assumptions may have held in the library setting where the early studies of relevance-oriented ranking were conducted [Maron and Kuhns, 1960, Cooper, 1971, Harter, 1975a,b, Robertson, 1977], they do not hold in general [Gordon and Lenk, 1992], and are unlikely to hold in a web search setting, which is permeated with ambiguity and redundancy.

Web search queries are typically short, ranging from two to three terms on average [Jansen et al., 2000]. While short queries are more likely to be ambiguous, every query can be arguably considered ambiguous to some extent [Cronen-Townsend and Croft, 2002]. Nevertheless, in the query understanding literature, query ambiguity is typically classified into three broad classes [Clarke et al., 2008, Song et al., 2009]. At one extreme of the ambiguity spectrum, genuinely *ambiguous queries* can have multiple *interpretations*. For instance, it is generally unclear whether the query *"bond"* refers to a debt security certificate or to Ian Fleming's fictional secret agent character.[2] Next, *underspecified queries* have a clearly defined interpretation, but it may be still unclear which particular *aspect* of this interpretation the user is interested in. For instance, while the query *"james bond"* arguably has a clearly defined interpretation (i.e., the secret agent character), it is unclear whether the user's information need is for books, films, games, etc. Finally, at the other extreme, *clear queries* have a generally well understood interpretation. An example of such queries is *"james bond books"*.

Sanderson [2008] investigated the impact of query ambiguity on web search. In particular, he analysed queries from a 2006 query log of a commercial web search engine that exactly matched a Wikipedia disambiguation page[3] or a WordNet[4] entry. Ambiguous queries from

[2] As a matter of fact, Wikipedia's disambiguation page for *"bond"* lists over 100 possible meanings for this particular entry: `http://en.wikipedia.org/wiki/Bond`.

[3] `http://en.wikipedia.org/wiki/Wikipedia:Disambiguation`

[4] `http://wordnet.princeton.edu`

Wikipedia showed a larger number of senses on average than those from WordNet (7.39 vs. 2.96), with the number of senses per ambiguous query following a power law in both cases. The average length of an ambiguous query was also similar across the two sources, with the predominance of single-word queries. In contrast to previous works, which assumed that multi-word queries were relatively unaffected by ambiguity, he found that ambiguous queries with more than one term were also numerous. Importantly, he observed that ambiguous queries comprised over 16% of all queries sampled from the log. Independent investigations based on click log analyses [Clough et al., 2009] and user studies [Song et al., 2009] also reached the consensual figure that around 16% of all user queries are ambiguous, while many more can be underspecified to some degree. As Sanderson [2008] demonstrated through a simulation, current search systems underperform for such queries.

While ambiguity primarily affects retrieval requests, redundancy is a property of the retrieval results. A document may be considered redundant whenever it conveys information already conveyed by the other documents [Bernstein and Zobel, 2005]. The limitation of assuming that documents are conditionally independent given the query was early recognised. In his note on relevance as a measurable quantity, Goffman [1964] pointed out that *"the relationship between a document and a query is necessary but not sufficient to determine relevance"*. Intuitively, once a document satisfying the user's information need has been observed, it is arguable whether other documents satisfying the same need would be deemed relevant. This intuition has been empirically corroborated in recent years with the analysis of users' browsing behaviour from click logs. Indeed, Craswell et al. [2008] observed that the probability of clicking on a given document diminishes as higher ranked documents are clicked. According to this cascade model, once a user has found the desired information, the need for inspecting further documents is reduced. In practice, the amount of information required to satisfy a user's information need may depend on additional factors. For instance, queries with an informational intent [Welch et al., 2011] as well as those of a controversial nature [Demartini, 2011] may require more than just a single relevant document to satisfy the user.

1.4 Diversity-oriented Ranking

Query ambiguity precludes a clear understanding of the user's actual information need. Wrongly guessing this need may compromise the accuracy of estimating the probability of relevance of any retrieved document. Introducing redundancy may further exacerbate the problem, by promoting more documents related to a potentially wrong information need. Indeed, when the user's actual information need is uncertain, relevance estimations may be misguided, leading to a complete retrieval failure and the abandonment of the query [Chen and Karger, 2006]. In this scenario, a standard relevance-oriented ranking approach is clearly suboptimal, and alternative ranking policies should be considered.

Diversity-oriented ranking has been proposed as a means to overcome ambiguity and redundancy during the search process. Diversifying the search results usually involves a departure from the assumptions that the relevance of a document can be estimated with certainty and independently of the other retrieved documents [Gordon and Lenk, 1991]. Indeed, uncertainty arises naturally from the fact that the probability of relevance is estimated based upon limited representations of both information needs and information items [Turtle and Croft, 1996]. Moreover, it is arguable whether users will still find a given document relevant to their information need once other documents satisfying this need have been observed [Bernstein and Zobel, 2005].

In order to account for both ambiguity and redundancy, a diversity-oriented ranking should not consider the relevance of each document in isolation. Instead, it should consider how relevant the document is in light of the multiple possible information needs underlying the query [Spärck-Jones et al., 2007] and in light of the other retrieved documents [Goffman, 1964]. As a result, the retrieved documents should provide the maximum coverage and minimum redundancy with respect to these multiple information needs [Clarke et al., 2008]. Ideally, the covered information needs should also reflect their relative importance, as perceived by the user population [Agrawal et al., 2009]. In its general form, this is an NP-hard problem [Carterette, 2009], for which an extensive body of research has been devoted in recent years. Discussing such a rich literature is the primary goal of this survey.

1.5 Scope of this Survey

This survey describes several approaches in the literature for the search result diversification problem. In particular, we cover approaches aimed to produce diversity-oriented rankings as well as those aimed at evaluating such rankings. Although our primary focus is on web search, this survey also describes diversification approaches that tackle ambiguity and redundancy in other search scenarios, as well as approaches for related tasks, such as query ambiguity detection and query aspect mining. Outside of the scope of this survey are approaches that seek to promote diversity for purposes other than search, such as text summarisation and event detection and tracking. The notations used uniformly throughout this survey are described in the preface.

The remainder of this survey is organised as follows. In Chapter 2, we provide a comprehensive overview of the search result diversification problem, including a discussion of its NP-hardness. We also describe an approximate polynomial-time solution that underlies most diversification approaches in the literature. These approaches are further organised according to a two-dimensional taxonomy, based upon their adopted aspect representation (implicit or explicit) and their diversification strategy (novelty-based, coverage-based, or hybrid).

In Chapters 3 and 4, we thoroughly describe the most prominent implicit and explicit diversification approaches in the literature, respectively. In both chapters, we focus on the diversification strategy and the ranking objective underlying each approach following the uniform notation introduced in the preface. Throughout these two chapters, we highlight the commonalities and differences among these approaches, and contrast their relative effectiveness as reported in the literature.

In Chapter 5, we describe the evaluation methodology most commonly adopted in the field of search result diversification, which builds upon the availability of benchmark test collections. In particular, we show the overall structure and the core components of a typical test collection for diversity evaluation, and provide a summary of salient statistics of the currently available test collections from TREC and NTCIR. Furthermore, we present multiple alternative evaluation frameworks and detail the evaluation metrics derived from each of them. Finally,

we discuss several studies that validate these metrics according to multiple dimensions, including their discriminative power, sensitivity, informativeness, predictive power, optimality, and reusability.

In Chapter 6, we introduce several advanced topics in the field of search result diversification. In particular, we describe approaches proposed for the related tasks of query ambiguity detection and query aspect mining. While the former approaches can be used to selectively adapt the amount of diversification performed for each individual query, the latter can help generate aspect representations that better reflect the possible information needs underlying a user's query. In addition, we describe several diversification approaches introduced for domains other than web search. This includes approaches for diversifying search results in different retrieval domains, such as images, biomedical reports, product reviews and recommendations, as well as for promoting diversity across multiple domains in an aggregated search interface.

Lastly, in Chapter 7, we provide a summary of the materials covered throughout this survey and discuss open research directions in the field of search result diversification. In particular, we highlight open problems related to modelling, estimation, and evaluation of diversification approaches, as a means to foster further research in the field.

References

R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining*, pages 5–14, Barcelona, Spain, 2009. ACM.

G. Amati. Frequentist and Bayesian approach to information retrieval. In *Proceedings of the 28th European Conference on IR Research on Advances in Information Retrieval*, pages 13–24, London, UK, 2006. Springer.

G. Amati. *Probability models for information retrieval based on Divergence From Randomness*. PhD thesis, University of Glasgow, 2003.

A. Arasu, J. Cho, H. Garcia-Molina, A. Paepcke, and S. Raghavan. Searching the Web. *ACM Transactions on Internet Technology*, 1(1):2–43, 2001. ISSN 1533-5399.

A. Ashkan and C. L. A. Clarke. On the informativeness of cascade and intent-aware effectiveness measures. In *Proceedings of the 20th International Conference on World Wide Web*, pages 407–416, Hyderabad, India, 2011. ACM.

J. A. Aslam, E. Yilmaz, and V. Pavlu. The maximum entropy method for analyzing retrieval measures. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 27–34, Salvador, Brazil, 2005. ACM.

R. Baeza-Yates, C. Hurtado, and M. Mendoza. Query recommendation using query logs in search engines. In *Proceedings of the 9th International Conference on Current Trends in Database Technology*, pages 588–596, Heraklion, Greece, 2004. Springer-Verlag.

R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Pearson Education Ltd., Harlow, UK, 2 edition, 2011.

D. Beeferman and A. Berger. Agglomerative clustering of a search engine query log. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 407–416, Boston, MA, USA, 2000. ACM.

F. Belém, R. L. T. Santos, J. Almeida, and M. A. Gonçalves. Topic diversity in tag recommendation. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 141–148, Hong Kong, China, 2013. ACM.

Y. Bernstein and J. Zobel. Redundant documents and search effectiveness. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pages 736–743, Bremen, Germany, 2005. ACM.

D. Berry and B. Fristedt. *Bandit problems: Sequential allocation of experiments*. Chapman and Hall, 1985.

S. Bhatia, D. Majumdar, and P. Mitra. Query suggestions in the absence of query logs. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 795–804, Beijing, China, 2011. ACM.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003. ISSN 1532-4435.

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008 (10):P10008+, 2008. ISSN 1742-5468.

P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna. The query-flow graph: model and applications. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 609–618, Napa Valley, CA, USA, 2008. ACM.

P. Boldi, F. Bonchi, C. Castillo, D. Donato, and S. Vigna. Query suggestions using query-flow graphs. In *Proceedings of the 2009 Workshop on Web Search Click Data*, pages 56–63. ACM, 2009a.

P. Boldi, F. Bonchi, C. Castillo, and S. Vigna. From "Dango" to "Japanese cakes": Query reformulation models and patterns. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pages 183–190, Milan, Italy, 2009b. IEEE Computer Society.

P. Borlund. The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10):913–925, 2003. ISSN 1532-2882.

A. Bouchoucha, J. He, and J.-Y. Nie. Diversified query expansion using ConceptNet. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, pages 1861–1864, San Francisco, CA, USA, 2013. ACM.

D. Broccolo, L. Marcon, F. M. Nardini, R. Perego, and F. Silvestri. Generating suggestions for queries in the long tail with an inverted index. *Information Processing and Management*, 48(2):326–339, 2012. ISSN 0306-4573.

A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002. ISSN 0163-5840.

C. Buckley. Why current IR engines fail. In *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval*, pages 584–585, Sheffield, UK, 2004. ACM Press.

G. Capannini, F. M. Nardini, R. Perego, and F. Silvestri. Efficient diversification of web search results. *Proceedings of the VLDB Endowment*, 4(7): 451–459, 2011. ISSN 2150-8097.

J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 335–336, Melbourne, Australia, 1998. ACM.

B. Carterette. Robust test collections for retrieval evaluation. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 55–62, Amsterdam, The Netherlands, 2007. ACM.

B. Carterette. An analysis of NP-completeness in novelty and diversity ranking. In *Proceedings of the 2nd International Conference on Theory of Information Retrieval*, pages 200–211, Cambridge, UK, 2009. Springer-Verlag.

B. Carterette and P. Chandar. Probabilistic models of ranking novel documents for faceted topic retrieval. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 1287–1296, Hong Kong, China, 2009. ACM.

P. Chandar and B. Carterette. Preference based evaluation measures for novelty and diversity. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 413–422, Dublin, Ireland, 2013. ACM.

O. Chapelle, D. Metlzer, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 621–630, Hong Kong, China, 2009. ACM.

O. Chapelle, Y. Chang, and T.-Y. Liu. Future directions in learning to rank. *Journal of Machine Learning Research, Proceedings Track*, pages 91–100, 2011a.

O. Chapelle, S. Ji, C. Liao, E. Velipasaoglu, L. Lai, and S.-L. Wu. Intent-based diversification of web search results: Metrics and algorithms. *Information Retrieval*, 14(6):572–592, 2011b.

H. Chen and D. R. Karger. Less is more: Probabilistic models for retrieving fewer relevant documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 429–436, Seattle, WA, USA, 2006. ACM.

C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 659–666, Singapore, Singapore, 2008. ACM.

C. L. A. Clarke, N. Craswell, and I. Soboroff. Overview of the TREC 2009 Web track. In *Proceedings of the 18th Text REtrieval Conference*, Gaithersburg, MD, USA, 2009a.

C. L. A. Clarke, M. Kolla, and O. Vechtomova. An effectiveness measure for ambiguous and underspecified queries. In *Proceedings of the 2nd International Conference on Theory of Information Retrieval*, pages 188–199, Cambridge, UK, 2009b. Springer-Verlag.

C. L. A. Clarke, N. Craswell, I. Soboroff, and G. V. Cormack. Overview of the TREC 2010 Web track. In *Proceedings of the 19th Text REtrieval Conference*, Gaithersburg, MD, USA, 2010.

C. L. A. Clarke, N. Craswell, I. Soboroff, and A. Ashkan. A comparative analysis of cascade measures for novelty and diversity. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, pages 75–84, Hong Kong, China, 2011a. ACM.

C. L. A. Clarke, N. Craswell, I. Soboroff, and E. M. Voorhees. Overview of the TREC 2011 Web track. In *Proceedings of the 20th Text REtrieval Conference*, Gaithersburg, MD, USA, 2011b.

C. L. A. Clarke, N. Craswell, and E. M. Voorhees. Overview of the TREC 2012 Web track. In *Proceedings of the 21st Text REtrieval Conference*, Gaithersburg, MD, USA, 2012.

C. Cleverdon. The Cranfield tests on index language devices. *Aslib Proceedings*, 19(6):173–194, 1967.

C. W. Cleverdon. The significance of the Cranfield tests on index languages. In *Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3–12, Chicago, IL, USA, 1991. ACM.

P. Clough, M. Sanderson, M. Abouammoh, S. Navarro, and M. Paramita. Multiple approaches to analysing query diversity. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 734–735, Boston, MA, USA, 2009. ACM.

K. Collins-Thompson, P. Bennett, F. Diaz, C. L. A. Clarke, and E. M. Voorhees. TREC 2013 Web track overview. In *Proceedings of the 22nd Text REtrieval Conference*, Gaithersburg, MD, USA, 2013.

W. S. Cooper. The inadequacy of probability of usefulness as a ranking criterion for retrieval system output. Technical report, University of California, Berkeley, Berkeley, CA, USA, 1971.

W. S. Cooper. Some inconsistencies and misidentified modeling assumptions in probabilistic information retrieval. *ACM Transactions on Information Systems*, 13(1):100–111, 1995.

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2nd edition, 2001.

N. Craswell and M. Szummer. Random walks on the click graph. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 239–246, Amsterdam, The Netherlands, 2007. ACM.

N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 1st International Conference on Web Search and Data Mining*, pages 87–94. ACM, 2008.

S. Cronen-Townsend and W. B. Croft. Quantifying query ambiguity. In *Proceedings of the 2nd International Conference on Human Language Technology Research*, pages 104–109, San Diego, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

M. Cutts. Spotlight keynote. In *Proceedings of Search Engine Strategies*, San Francisco, CA, USA, 2012.

V. Dang and B. W. Croft. Term level search result diversification. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 603–612, Dublin, Ireland, 2013. ACM.

V. Dang and W. B. Croft. Query reformulation using anchor text. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, pages 41–50, New York, NY, USA, 2010. ACM.

V. Dang and W. B. Croft. Diversity by proportionality: an election-based approach to search result diversification. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 65–74, Portland, OR, USA, 2012. ACM.

V. Dang, X. Xue, and W. B. Croft. Inferring query aspects from reformulations using clustering. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 2117–2120, Glasgow, UK, 2011. ACM.

G. Demartini. ARES: a retrieval engine based on sentiments sentiment-based search result annotation and diversification. In *Proceedings of the 33rd European Conference on IR Research on Advances in Information Retrieval*, pages 772–775, Dublin, Ireland, 2011. Springer-Verlag.

E. Demidova, P. Fankhauser, X. Zhou, and W. Nejdl. Divq: Diversification for keyword search over structured databases. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 331–338, Geneva, Switzerland, 2010. ACM.

T. Deselaers, T. Gass, P. Dreuw, and H. Ney. Jointly optimising relevance and diversity in image retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pages 1–8, Santorini, Greece, 2009. ACM.

F. Diaz, M. Lalmas, and M. Shokouhi. From federated to aggregated search. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 910, 2010.

Z. Dou, S. Hu, K. Chen, R. Song, and J.-R. Wen. Multi-dimensional search result diversification. In *Proceedings of the fourth ACM international Conference on Web Search and Data Mining*, pages 475–484, Hong Kong, China, 2011. ACM.

D. Downey, S. Dumais, and E. Horvitz. Heads and tails: studies of web search with common and rare queries. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 847–848, Amsterdam, The Netherlands, 2007. ACM.

M. Dundar, B. Krishnapuram, J. Bi, and R. B. Rao. Learning classifiers when the training data is not IID. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, pages 756–761, Hyderabad, India, 2007. Morgan Kaufmann Publishers Inc.

N. Eiron and K. S. McCurley. Analysis of anchor text for web search. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 459–460, Toronto, Canada, 2003. ACM.

U. Feige. A threshold of $\ln(n)$ for approximating set cover. *Journal of the ACM*, 45:634–652, 1998. ISSN 0004-5411.

B. M. Fonseca, P. B. Golgher, E. S. De Moura, B. Pôssas, and N. Ziviani. Discovering search engine related queries using association rules. *Journal of Web Engineering*, 2(4):215–227, October 2003. ISSN 1540-9589.

E. A. Fox and J. A. Shaw. Combination of multiple searches. In *Proceedings of the 2nd Text REtrieval Conference*, pages 243–252, Gaithersburg, MD, USA, 1993.

J. H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

X. Geng, T.-Y. Liu, T. Qin, A. Arnold, H. Li, and H.-Y. Shum. Query dependent ranking using k-nearest neighbor. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 115–122, Singapore, Singapore, 2008. ACM.

V. Gil-Costa, R. L. T. Santos, C. Macdonald, and I. Ounis. Sparse spatial selection for novelty-based search result diversification. In *Proceedings of the 18th International Symposium on String Processing and Information Retrieval*, pages 344–355, Pisa, Italy, 2011. Springer.

V. Gil-Costa, R. L. T. Santos, C. Macdonald, and I. Ounis. Modelling efficient novelty-based search result diversification in metric spaces. *Journal of Discrete Algorithms*, 18:75–88, 2013. ISSN 1570-8667.

W. Goffman. On relevance as a measure. *Information Storage and Retrieval*, 2(3):201–203, 1964.

P. B. Golbus, J. A. Aslam, and C. L. Clarke. Increasing evaluation sensitivity to diversity. *Information Retrieval*, 16(4):530–555, 2013. ISSN 1386-4564.

S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *Proceedings of the 18th International Conference on World Wide Web*, pages 381–390, Madrid, Spain, 2009. ACM.

M. D. Gordon and P. Lenk. A utility theoretic examination of the probability ranking principle in information retrieval. *Journal of the American Society for Information Science and Technology*, 42(10):703–714, 1991.

M. D. Gordon and P. Lenk. When is the probability ranking principle suboptimal? *Journal of the American Society for Information Science and Technology*, 43(1):1–14, 1992.

D. Harman. Overview of the second Text REtrieval Conference (TREC-2). In *Proceedings of the 2nd Text REtrieval Conference*, Gaithersburg, MD, USA, 1993.

S. P. Harter. A probabilistic approach to automatic keyword indexing. Part I: On the distribution of specialty words in a technical literature. *Journal of the American Society for Information Science*, 26(4):197–206, 1975a.

S. P. Harter. A probabilistic approach to automatic keyword indexing. Part II: An algorithm for probabilistc indexing. *Journal of the American Society for Informaiton Science*, 26(4):280–289, 1975b.

J. He, E. Meij, and M. de Rijke. Result diversification based on query-specific cluster ranking. *Journal of the American Society for Information Science and Technology*, 62(3):550–571, 2011. ISSN 1532-2882.

J. He, V. Hollink, and A. de Vries. Combining implicit and explicit topic representations for result diversification. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 851–860, Portland, OR, USA, 2012. ACM.

W. R. Hersh and P. Over. TREC-8 Interactive track report. In *Proceedings of the 8th Text REtrieval Conference*, Gaithersburg, MD, USA, 1999.

D. Hiemstra. A linguistically motivated probabilistic model of information retrieval. In *Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries*, pages 569–584, Heraklion, Greece, 1998. Springer.

D. S. Hochbaum, editor. *Approximation algorithms for NP-hard problems*. PWS Publishing Co., Boston, MA, USA, 1997.

D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, New York, NY, USA, 1st edition, 2010.

B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval: A study of user queries on the Web. *SIGIR Forum*, 32(1):5–17, 1998. ISSN 0163-5840.

B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management*, 36(2):207–227, 2000. ISSN 0306-4573.

K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002. ISSN 1046-8188.

R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *Proceedings of the 15th international conference on World Wide Web*, pages 387–396, Edinburgh, UK, 2006. ACM.

I.-H. Kang and G. Kim. Query type classification for web document retrieval. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 64–71, Toronto, Canada, 2003. ACM.

J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Springer, 1960.

S. Kharazmi, M. Sanderson, F. Scholer, and D. Vallet. Using score differences for search result diversification. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1143–1146. ACM, 2014.

E. Kharitonov, C. Macdonald, P. Serdyukov, and I. Ounis. Intent models for contextualising and diversifying query suggestions. In *Proceedings of the 22nd ACM International Conference on Conference on information and Knowledge Management*, pages 2303–2308, San Francisco, CA, USA, 2013. ACM.

S. Khuller, A. Moss, and J. S. Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70:39–45, 1999. ISSN 0020-0190.

Y. Kim and W. B. Croft. Diversifying query suggestions based on query documents. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 891–894, Gold Coast, QLD, Australia, 2014. ACM.

R. Kraft and J. Zien. Mining anchor text for query refinement. In *Proceedings of the 13th International Conference on World Wide Web*, pages 666–674, New York, NY, USA, 2004. ACM.

R. Krestel and N. Dokoohaki. Diversifying product review rankings: Getting the full picture. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pages 138–145, Washington, DC, USA, 2011. IEEE Computer Society.

U. Kruschwitz, D. Lungley, M.-D. Albakour, and D. Song. Deriving query suggestions for site search. *Journal of the American Society for Information Science and Technology*, 64(10):1975–1994, 2013. ISSN 1532-2890.

E. Lagergren and P. Over. Comparing interactive information retrieval systems across sites: The TREC-6 Interactive track matrix experiment. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 164–172, Melbourne, Australia, 1998. ACM.

N. Lathia, S. Hailes, L. Capra, and X. Amatriain. Temporal diversity in recommender systems. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 210–217, Geneva, Switzerland, 2010. ACM.

V. Lavrenko and W. B. Croft. Relevance based language models. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 120–127, New Orleans, LA, USA, 2001. ACM.

T. Leelanupab, G. Zuccon, and J. M. Jose. A comprehensive analysis of parameter settings for novelty-biased cumulative gain. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 1950–1954, Maui, HI, USA, 2012. ACM.

S. Liang, Z. Ren, and M. de Rijke. Fusion helps diversification. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 303–312, Gold Coast, QLD, Australia, 2014. ACM.

N. Limsopatham, C. Macdonald, and I. Ounis. Modelling relevance towards multiple inclusion criteria when ranking patients. In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management*, pages 1639–1648, Shanghai, China, 2014. ACM.

H. Liu and P. Singh. ConceptNet—a practical commonsense reasoning toolkit. *BT Technology Journal*, 22(4):211–226, 2004. ISSN 1358-3948.

H. Ma, M. R. Lyu, and I. King. Diversifying query suggestion results. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, Atlanta, GA, USA, 2010. AAAI Press.

J. I. Marden. *Analyzing and modeling rank data*. Taylor & Francis, 1996.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 00221082.

M. E. Maron and J. L. Kuhns. On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 7(3):216–244, 1960. ISSN 0004-5411.

Q. Mei, D. Zhou, and K. Church. Query suggestion using hitting time. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 469–478, Napa Valley, CA, USA, 2008. ACM.

M. Melucci. Contextual search: A computational framework. *Foundations and Trends in Information Retrieval*, 6(4-5):257–405, 2012.

S. Mizzaro. Relevance: The whole history. *Journal of the American Society for Information Science*, 48(9):810–832, 1997. ISSN 0002-8231.

A. Moffat and J. Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Transactions on Information Systems*, 27(1):1–27, 2008. ISSN 1046-8188.

V. Murdock and M. Lalmas. Workshop on aggregated search. *SIGIR Forum*, 42:80–83, 2008. ISSN 0163-5840.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14:265–294, 1978. ISSN 0025-5610.

T. N. Nguyen and N. Kanhabua. Leveraging dynamic query subtopics for time-aware search result diversification. In *Proceedings of the 36th European Conference on Information Retrieval*, pages 222–234, Amsterdam, The Netherlands, 2014. Springer.

P. Over. TREC-6 Interactive report. In *Proceedings of the 6th Text REtrieval Conference*, pages 73–81, Gaithersburg, MD, USA, 1997.

P. Over. TREC-7 Interactive track report. In *Proceedings of the 7th Text REtrieval Conference*, pages 33–39, Gaithersburg, MD, USA, 1998.

A. M. Ozdemiray and I. S. Altingovde. Score and rank aggregation methods for explicit search result diversification. Technical Report METU-CENG-2013-01, Middle East Technical University, Ankara, Turkey, 2013.

M. L. Paramita, J. Tang, and M. Sanderson. Generic and spatial approaches to image search results diversification. In *Proceedings of the 31st European Conference on IR Research on Advances in Information Retrieval*, pages 603–610, Toulouse, France, 2009. Springer.

J. Peng, C. Macdonald, and I. Ounis. Learning to select a ranking function. In *Proceedings of the 31st European Conference on IR Research on Advances in Information Retrieval*, pages 114–126, Milton Keynes, UK, 2010. Springer.

V. Plachouras. Diversity in expert search. In *Proceedings of the 1st International Workshop on Diversity in Document Retrieval*, pages 63–67, Dublin, Ireland, 2011.

A. Plakhov. Entity-oriented search result diversification. In *Proceedings of the 1st International Workshop on Entity-Oriented Search*, Beijing, China, 2011.

J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 275–281, Melbourne, Australia, 1998. ACM.

F. Radlinski and S. Dumais. Improving personalized web search using result diversification. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 691–692, Seattle, WA, USA, 2006. ACM.

F. Radlinski, R. Kleinberg, and T. Joachims. Learning diverse rankings with multi-armed bandits. In *Proceedings of the 25th International Conference on Machine Learning*, pages 784–791, Helsinki, Finland, 2008. ACM.

F. Radlinski, P. N. Bennett, B. Carterette, and T. Joachims. Redundancy, diversity and interdependent document relevance. *SIGIR Forum*, 43(2): 46–52, 2009. ISSN 0163-5840.

F. Radlinski, M. Szummer, and N. Craswell. Metrics for assessing sets of subtopics. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 853–854, Geneva, Switzerland, 2010a. ACM.

F. Radlinski, M. Szummer, and N. Craswell. Inferring query intent from reformulations and clicks. In *Proceedings of the 19th International Conference on World Wide Web*, pages 1171–1172, Raleigh, NC, USA, 2010b.

D. Rafiei, K. Bharat, and A. Shukla. Diversifying web search results. In *Proceedings of the 19th International Conference on World Wide Web*, pages 781–790, Raleigh, NC, USA, 2010.

K. Raman, P. Shivaswamy, and T. Joachims. Online learning to diversify from implicit feedback. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, China, 2012. ACM.

S. E. Robertson and K. Spärck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27:129–146, 1976.

S. E. Robertson, C. J. van Rijsbergen, and M. F. Porter. Probabilistic models of indexing and searching. In *Proceedings of the 3rd Annual ACM Conference on Research and Development in Information Retrieval*, pages 35–56. Butterworth & Co., 1981.

S. Robertson and H. Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389, 2009. ISSN 1554-0669.

S. Robertson, H. Zaragoza, and M. Taylor. Simple bm25 extension to multiple weighted fields. In *Proceedings of the 13th ACM International Conference on Information and Knowledge Management*, pages 42–49, Washington, DC, USA, 2004. ACM.

S. E. Robertson. The probability ranking principle in IR. *Journal of Documentation*, 33(4):294–304, 1977.

S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *Proceedings of the 3rd Text REtrieval Conference*, Gaithersburg, MD, USA, 1994.

D. E. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of the 13th International Conference on World Wide Web*, pages 13–19, New York, NY, USA, 2004. ACM.

B. R. Rowe, D. W. Wood, A. N. Link, and D. A. Simoni. Economic impact assessment of NIST's Text REtrieval Conference (TREC) program. Technical Report 0211875, RTI International, 2010.

T. Sakai. Evaluating evaluation metrics based on the bootstrap. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 525–532, Seattle, WA, USA, 2006. ACM.

T. Sakai. Alternatives to bpref. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 71–78, Amsterdam, The Netherlands, 2007. ACM.

T. Sakai. Evaluation with informational and navigational intents. In *Proceedings of the 21st International Conference on World Wide Web*, pages 499–508, Lyon, France, 2012. ACM.

T. Sakai. The unreusability of diversified search test collections. In *Proceedings of the 5th International Workshop on Evaluating Information Access*, pages 1–8, Tokyo, Japan, 2013.

T. Sakai and R. Song. Diversified search evaluation: Lessons from the NTCIR-9 Intent task. *Information Retrieval*, 2012. ISSN 1386-4564.

T. Sakai, N. Craswell, R. Song, S. Robertson, Z. Dou, and C.-Y. Lin. Simple evaluation metrics for diversified search results. In *Proceedings of the 3rd International Workshop on Evaluating Information Access*, pages 42–50, Tokyo, Japan, 2010. NII.

T. Sakai, Z. Dou, and C. L. Clarke. The impact of intent selection on diversified search evaluation. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 921–924, Dublin, Ireland, 2013a. ACM.

T. Sakai, Z. Dou, T. Yamamoto, Y. Liu, M. Zhang, and R. Song. Overview of the NTCIR-10 Intent-2 task. In *Proceedings of the 10th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, Tokyo, Japan, 2013b.

P. A. Samuelson and W. D. Nordhaus. *Microeconomics*. McGraw-Hill, 2001.

M. Sanderson. Ambiguous queries: Test collections need more sense. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 499–506, Singapore, Singapore, 2008. ACM.

M. Sanderson. Test collection based evaluation of information retrieval systems. *Foundations and Trends in Information Retrieval*, 4(4):247–375, 2010.

M. Sanderson, M. L. Paramita, P. Clough, and E. Kanoulas. Do user preferences and evaluation measures line up? In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 555–562, Geneva, Switzerland, 2010. ACM.

R. L. T. Santos. *Explicit web search result diversification*. PhD thesis, School of Computing Science, University of Glasgow, Glasgow, UK, 2013.

R. L. T. Santos and I. Ounis. Diversifying for multiple information needs. In *Proceedings of the 1st International Workshop on Diversity in Document Retrieval*, pages 37–41, Dublin, Ireland, 2011.

R. L. T. Santos, C. Macdonald, and I. Ounis. Selectively diversifying web search results. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pages 1179–1188, Toronto, Canada, 2010a. ACM.

R. L. T. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web*, pages 881–890, Raleigh, NC, USA, 2010b. ACM.

R. L. T. Santos, J. Peng, C. Macdonald, and I. Ounis. Explicit search result diversification through sub-queries. In *Proceedings of the 31st European Conference on IR Research on Advances in Information Retrieval*, pages 87–99, Milton Keynes, UK, 2010c. Springer.

R. L. T. Santos, C. Macdonald, and I. Ounis. Aggregated search result diversification. In *Proceedings of the 3rd International Conference on the Theory of Information Retrieval*, pages 250–261, Bertinoro, Italy, 2011a. Springer.

R. L. T. Santos, C. Macdonald, and I. Ounis. Intent-aware search result diversification. In *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 595–604, Beijing, China, 2011b. ACM.

R. L. T. Santos, C. Macdonald, R. McCreadie, I. Ounis, and I. Soboroff. Information retrieval on the blogosphere. *Foundations and Trends in Information Retrieval*, 6(1):1–125, 2012a.

R. L. T. Santos, C. Macdonald, and I. Ounis. On the role of novelty for search result diversification. *Information Retrieval*, 15(5):478–502, 2012b.

R. L. T. Santos, C. Macdonald, and I. Ounis. Learning to rank query suggestions for adhoc and diversity search. *Information Retrieval*, 16(4):429–451, 2013. ISSN 1386-4564.

M. Searcóid. *Metric Spaces*. Springer Undergraduate Mathematics Series. Springer, 2006.

C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1):6–12, 1999. ISSN 0163-5840.

F. Silvestri. Mining query logs: Turning search usage data into knowledge. *Foundations and Trends in Information Retrieval*, 4(1-2):1–174, 2010.

A. Slivkins, F. Radlinski, and S. Gollapudi. Learning optimally diverse rankings over large document collections. In *Proceedings of the 27th Annual International Conference on Machine Learning*, pages 983–990, Haifa, Israel, 2010. Omnipress.

R. Song, Z. Luo, J.-Y. Nie, Y. Yu, and H.-W. Hon. Identification of ambiguous queries in web search. *Information Processing and Management*, 45(2):216–229, 2009. ISSN 0306-4573.

R. Song, M. Zhang, T. Sakai, M. P. Kato, Y. Liu, M. Sugimoto, Q. Wang, and N. Orii. Overview of the NTCIR-9 Intent task. In *Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies*, Tokyo, Japan, 2011a.

Y. Song, D. Zhou, and L. wei He. Post-ranking query suggestion by diversifying search results. In *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 815–824, Beijing, China, 2011b. ACM.

K. Spärck-Jones, S. E. Robertson, and M. Sanderson. Ambiguous requests: Implications for retrieval tests, systems and theories. *SIGIR Forum*, 41(2): 8–17, 2007. ISSN 0163-5840.

I. Szpektor, A. Gionis, and Y. Maarek. Improving recommendation for long-tail queries via templates. In *Proceedings of the 20th international conference on World wide web*, pages 47–56, Hyderabad, India, 2011. ACM.

J. Teevan, S. T. Dumais, and E. Horvitz. Characterizing the value of personalizing search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 757–758, Amsterdam, The Netherlands, 2007. ACM.

I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6:1453–1484, 2005. ISSN 1532-4435.

H. R. Turtle and W. B. Croft. Uncertainty in information retrieval systems. In *Uncertainty Management in Information Systems*, pages 189–224. Kluwer Academic Publishers, Norwell, MA, USA, 1996.

D. Vallet. Crowdsourced evaluation of personalization and diversification techniques in web search. In *Proceedings of the ACM SIGIR Workshop on Crowdsourcing for Information Retrieval*, Beijing, China, 2011. ACM.

D. Vallet and P. Castells. Personalized diversification of search results. In *Proceedings of the 35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 841–850, Portland, OR, USA, 2012. ACM.

R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol. Visual diversification of image search results. In *Proceedings of the 18th International Conference on World Wide Web*, pages 341–350, Madrid, Spain, 2009. ACM.

C. J. van Rijsbergen. *The Geometry of Information Retrieval*. Cambridge University Press, New York, NY, USA, 2004.

S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems*, pages 109–116, Chicago, IL, USA, 2011. ACM.

S. Vargas, P. Castells, and D. Vallet. Intent-oriented diversity in recommender systems. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1211–1212, Beijing, China, 2011. ACM.

S. Vargas, R. L. T. Santos, C. Macdonald, and I. Ounis. Selecting effective expansion terms for diversity. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, pages 69–76, Lisbon, Portugal, 2013. CID.

E. Vee, U. Srivastava, J. Shanmugasundaram, P. Bhat, and S. A. Yahia. Efficient computation of diverse query results. In *Proceedings of the 24th International Conference on Data Engineering*, pages 228–236, Cancún, Mexico, 2008. IEEE Computer Society.

R. V. Vohra and N. G. Hall. A probabilistic analysis of the maximal covering location problem. *Discrete Applied Mathematics*, 43(2):175–183, 1993. ISSN 0166-218X.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

E. M. Voorhees. TREC: Continuing information retrieval's tradition of experimentation. *Communications of the ACM*, 50(11):51–54, 2007. ISSN 0001-0782.

E. M. Voorhees and D. Harman. Overview of the 6th Text REtrieval Conference. In *Proceedings of the 6th Text REtrieval Conference*, Gaithersburg, MD, USA, 1997.

E. M. Voorhees and D. Harman. Overview of the 7th Text REtrieval Conference. In *Proceedings of the 7th Text REtrieval Conference*, Gaithersburg, MD, USA, 1998.

E. M. Voorhees and D. Harman. Overview of the 8th Text REtrieval Conference. In *Proceedings of the 8th Text REtrieval Conference*, Gaithersburg, MD, USA, 1999.

E. M. Voorhees and D. K. Harman. *TREC: Experiment and Evaluation in Information Retrieval*. Digital Libraries and Electronic Publishing. MIT Press, 2005.

J. Wang and J. Zhu. Portfolio theory of information retrieval. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 115–122, Boston, MA, USA, 2009. ACM.

Q. Wang, Y. Qian, R. Song, Z. Dou, F. Zhang, T. Sakai, and Q. Zheng. Mining subtopics from text fragments for a web query. *Information Retrieval*, 16 (4):484–503, 2013. ISSN 1386-4564.

X. Wang and C. Zhai. Mining term association patterns from search logs for effective query reformulation. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 479–488, Napa Valley, CA, USA, 2008. ACM.

X. Wang, H. Fang, and C. Zhai. A study of methods for negative relevance feedback. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 219–226, Singapore, Singapore, 2008. ACM.

M. J. Welch, J. Cho, and C. Olston. Search result diversity for informational queries. In *Proceedings of the 20th International Conference on World Wide Web*, pages 237–246, Hyderabad, India, 2011. ACM.

G. J. Woeginger. Exact algorithms for NP-hard problems: A survey. In *Combinatorial Optimization—Eureka, You Shrink!*, pages 185–207. Springer, 2003.

M. A. Woodbury. Inverting modified matrices. Technical Report MR38136, Statistical Research Group, Princeton University, Princeton, NJ, USA, 1950.

F. Wu, J. Madhavan, and A. Halevy. Identifying aspects for web-search queries. *Journal of Artificial Intelligence Research*, 40(1):677–700, 2011. ISSN 1076-9757.

X. Yin, J. X. Huang, X. Zhou, and Z. Li. A survival modeling approach to biomedical search result diversification using Wikipedia. In *Proceedings of the 33rd international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 901–902, Geneva, Switzerland, 2010. ACM.

C. Yu, L. Lakshmanan, and S. Amer-Yahia. It takes variety to make a world: Diversification in recommender systems. In *Proceedings of the 12th International Conference on Extending Database Technology*, pages 368–378, Saint Petersburg, Russia, 2009. ACM.

Y. Yue and T. Joachims. Predicting diverse subsets using structural svms. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1224–1231, Helsinki, Finland, 2008. ACM.

H. Zaragoza, N. Craswell, M. J. Taylor, S. Saria, and S. E. Robertson. Microsoft Cambridge at TREC 13: Web and Hard tracks. In *Proceedings of the 13th Text REtrieval Conference*, Gaithersburg, MD, USA, 2004.

C. Zhai. Statistical language models for information retrieval: A critical review. *Foundations and Trends in Information Retrieval*, 2(3):137–213, 2008. ISSN 1554-0669.

C. Zhai and J. Lafferty. A risk minimization framework for information retrieval. *Information Processing and Management*, 42(1):31–55, 2006. ISSN 0306-4573.

C. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 10–17, Toronto, Canada, 2003. ACM.

Z. Zhang and O. Nasraoui. Mining search engine query logs for query recommendation. In *Proceedings of the 15th international conference on World Wide Web*, pages 1039–1040, Edinburgh, UK, 2006. ACM.

W. Zheng, H. Fang, C. Yao, and M. Wang. Search result diversification for enterprise data. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 1901–1904, Glasgow, UK, 2011a. ACM.

W. Zheng, X. Wang, H. Fang, and H. Cheng. An exploration of pattern-based subtopic modeling for search result diversification. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, pages 387–388, Ottawa, ON, Canada, 2011b. ACM.

W. Zheng, H. Fang, and C. Yao. Exploiting concept hierarchy for result diversification. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 1844–1848, Maui, HI, USA, 2012. ACM.

W. Zheng, H. Fang, C. Yao, and M. Wang. Leveraging integrated information to extract query subtopics for search result diversification. *Information Retrieval*, 17(1):52–73, 2014. ISSN 1386-4564.

T. Zhou, Z. Kuscsik, J. Liu, M. Medo, J. Wakeling, and Y. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.

X. Zhu, A. B. Goldberg, J. V. Gael, and D. Andrzejewski. Improving diversity in ranking using absorbing randomwalks. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics—Human Language Technologies*, pages 97–104, Rochester, NY, USA, 2007. ACL.

Y. Zhu, Y. Lan, J. Guo, X. Cheng, and S. Niu. Learning for search result diversification. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 293–302, Gold Coast, QLD, Australia, 2014. ACM.

C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, pages 22–32, Chiba, Japan, 2005. ACM.

J. Zobel. How reliable are the results of large-scale information retrieval experiments? In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 307–314, Melbourne, Australia, 1998. ACM.

G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. In *Proceedings of the 32nd European Conference on IR Research on Advances in Information Retrieval*, pages 357–369, Milton Keynes, UK, 2010. Springer.

G. Zuccon, L. Azzopardi, D. Zhang, and J. Wang. Top-k retrieval using facility location analysis. In *Proceedings of the 34th European Conference on Advances in Information Retrieval*, pages 305–316, Barcelona, Spain, 2012. Springer-Verlag.