# Diverse Retrieval via Greedy Optimization of Expected 1-call@k in a Latent Subtopic Relevance Model

Scott Sanner
NICTA & the ANU
Canberra, Australia
first.last@nicta.com.au

Shengbo Guo
Xerox Research Centre Europe
Grenoble, France
first.last@xrce.xerox.com

Thore Graepel
Microsoft Research
Cambridge, UK
thoreg@microsoft.com

Sadegh Kharazmi
RMIT & NICTA
Melbourne, Australia
first.last@rmit.edu.au

Sarvnaz Karimi
NICTA & The University of Melbourne, Dept. CSSE
first.last@nicta.com.au

## ABSTRACT

It has been previously observed that optimization of the 1-call@$k$ relevance objective (i.e., a set-based objective that is 1 if at least one document is relevant, otherwise 0) empirically correlates with diverse retrieval. In this paper, we proceed one step further and show theoretically that greedily optimizing expected 1-call@$k$ w.r.t. a latent subtopic model of binary relevance leads to a diverse retrieval algorithm sharing many features of existing diversification approaches. This new result is complementary to a variety of diverse retrieval algorithms derived from alternate rank-based relevance criteria such as average precision and reciprocal rank. As such, the derivation presented here for expected 1-call@$k$ provides a novel theoretical perspective on the *emergence of diversity via a latent subtopic model of relevance* — an idea underlying both ambiguous and faceted subtopic retrieval that have been used to motivate diverse retrieval.

## Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval Models

## General Terms

Algorithms

## Keywords

diversity, set-level relevance, maximal marginal relevance

## 1. DIVERSE RETRIEVAL AND SUBTOPICS

One of the basic tenets of set-based information retrieval is to minimize redundancy, hence maximize diversity, in the result set to increase the chance that the results will contain items relevant to the user's query [9]. Hence, *diverse retrieval* can be defined as a *set-level* retrieval objective that takes into account inter-document relevance dependences when producing a result set relevant to a query.

*Subtopic retrieval* — "the task of finding documents that cover as many *different* subtopics of a general topic as possible" [19] — has often been noted as a motivating case for diverse retrieval. That is, if a query has multiple facets that should be covered by a result set, or a query has multiple ambiguous interpretations, then a retrieval algorithm should try to "cover" all of these subtopics in its result set. It is this subtopic-based motivation for diverse retrieval — a motivation which also underlies the TREC 6-8 Interactive tracks[1] and TREC 2009-2010 Diversity subset of the Web tracks[2] — that we draw on for the latent subtopic binary relevance model presented in this paper.

If one wants to optimize a result set to cover all possible query subtopics, the question naturally arises as to what set-level relevance objective should be optimized? Wang and Zhu [17] have shown that natural forms of diversification arise via the optimization of *average precision* [3] and *reciprocal rank* [15]. While these results directly motivate diverse retrieval via *rank-based* (ordered set) relevance criteria, they do not use the subtopic motivation for diversity. We use this alternate subtopic motivation in this paper, where we define binary relevance via a *latent subtopic model*. With this definition of relevance, we then optimize the expectation of the *$n$-call@$k$ set-based* relevance criteria (specifically for $n = 1$) that takes the value 1 if at least $n$ of $k$ documents in a result set are relevant and 0 otherwise [6]. We conjecture that an optimal result set w.r.t. this objective and relevance model will attempt to cover all subtopics in order to ensure that at least one document is relevant, hence yielding diversity.

One may ask why we focus on the $n$-call@$k$ metric with $n = 1$ rather than $n > 1$ for diverse retrieval? In [16] (Figure 2c), Wang and Zhu observed that optimizing 1-call@$k$ correlates *most strongly* with diverse retrieval, while as $n \to k$, retrieval becomes less diverse. The reasons for this are simple: as $n \to k$, a higher proportion of documents are required to be relevant; if the top-ranked document is deemed most relevant, similar documents are also likely to be relevant, discouraging diversity. At the other extreme, $n = 1$ encourages diversity since only one relevant document is needed.


*CIKM'11,* October 24–28, 2011, Glasgow, Scotland, UK.


[1] `http://www-nlpir.nist.gov/projects/t8i/t8i.html`
[2] `http://trec.nist.gov/data/web09.html` (also `web10`)

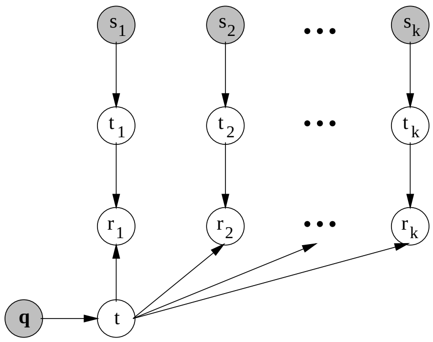


**Figure 1: Latent subtopic binary relevance model.**

In the rest of this paper, we derive a diverse retrieval algorithm via greedy optimization of *expected 1-call@k in a latent subtopic binary relevance model* and compare it to a variety of existing diversification approaches.

## 2. OPTIMIZING EXPECTED 1-CALL@K

Given an *item set* $D$ (e.g., a set of documents) where retrieved items are denoted as $s_i \in D$, we aim to select an optimal subset of items $S_k^* \subset D$ (where $|S_k^*| = k$ and $k < |D|$) relevant to a given query $\mathbf{q}$ (e.g., query terms). For computational efficiency, we will build $S_k^*$ in a greedy manner by choosing the next optimal selection $s_k^*$ given the previous set of optimal selections $S_{k-1}^* = \{s_1^*, \ldots, s_{k-1}^*\}$ and recursively defining $S_k^* = S_{k-1}^* \cup \{s_k^*\}$ with $S_0^* = \emptyset$.

One of the most popular result set diversification methods is Maximal Marginal Relevance (MMR) [4] that chooses $s_k^*$ greedily according to the following criteria:

$$s_k^* = \underset{s_k \in D \setminus S_{k-1}^*}{\arg\max} \; [\lambda(\mathrm{Sim}_1(\mathbf{q}, s_k)) - (1-\lambda) \max_{s_i \in S_{k-1}^*} \mathrm{Sim}_2(s_i, s_k)]. \quad (1)$$

Here, similarity metric $\mathrm{Sim}_1$ measures query-item relevance, metric $\mathrm{Sim}_2$ measures the similarity between two items, and the parameter $\lambda \in [0,1]$ trades off relevance and diversity. In the case of $s_1^*$, the maximization term is vacuous (=0). We take special note of this form for MMR optimization since the results we derive next will bear a close resemblance.

To begin the derivation, we provide a directed graphical model in Figure 1 to formalize the independence assumptions in a probabilistic subtopic model of binary relevance. Shaded nodes represent observed variables while unshaded nodes are latent. The observed variables are the vector of query terms $\mathbf{q}$ and selected items $s_i$ (where for $1 \le i \le k$, $s_i \in D$). For the subtopic variables, let $T$ be a discrete subtopic set. Then variables $t_i \in T$ represent subtopics for respective $s_i$ and $t \in T$ represents a subtopic for query $\mathbf{q}$. The $r_i$ are binary variables indicating whether the respective selected items $s_i$ are relevant (1) or not (0).

The conditional probability tables (CPTs) are as follows: $P(t_i|s_i)$ and $P(t|\mathbf{q})$ respectively represent the subtopic distribution for item $s_i$ and query $\mathbf{q}$. The remaining CPTs are for relevance variables $r_i$, where item $s_i$ is deemed *relevant* $(r_i = 1)$ *iff its subtopic* $t_i$ *matches query subtopic* $t$:

$$P(r_i = 1|t, t_i) \;=\; \mathbb{I}[t_i = t]$$

Here, $\mathbb{I}[\cdot]$ is 1 when its argument is true and 0 otherwise.

We now formally define the *expected 1-call@k* objective:

$$\text{Exp-1-Call@}k(S_k, \mathbf{q}) = \mathbb{E}\left[ \bigvee_{i=1}^{k} r_i = 1 \,\middle|\, s_1, \ldots, s_k, \mathbf{q} \right] \quad (2)$$

Since jointly optimizing Exp-1-Call@$k(S_k, \mathbf{q})$ is NP-hard, we take a greedy approach similar to MMR where we choose the best $s_k^*$ assuming that $S_{k-1}^*$ is given. Then following [6], we can greedily optimize this objective as follows:[3]

$$\begin{aligned}
s_k^* &= \underset{s_k}{\arg\max}\ \text{Exp-1-Call@}k(S_{k-1}^* \cup \{s_k\}, \mathbf{q}) \\
&= \underset{s_k}{\arg\max}\, \mathbb{E}\left[ \bigvee_{i=1}^{k} r_i = 1 \,\middle|\, S_{k-1}^*, s_k, \mathbf{q} \right] \\
&= \underset{s_k}{\arg\max}\, \mathbb{E}\Bigg[ (r_1 = 1) \vee (r_2 = 1 \wedge r_1 = 0) \vee \cdots \vee \\
&\qquad\qquad \left( r_k = 1 \wedge \bigwedge_{i=1}^{k-1} r_i = 0 \right) \,\Bigg|\, S_{k-1}^*, s_k, \mathbf{q} \Bigg] \\
&= \underset{s_k}{\arg\max} \sum_{i=1}^{k} P(r_i = 1, \{r_j = 0\}_{j<i} \,|\, \{s_j^*\}_{j \le i, j<k}, \{s_k\}_{k=i}, \mathbf{q}) \\
&= \underset{s_k}{\arg\max} \sum_{i=1}^{k} P(r_i = 1 \,|\, \{r_j = 0\}_{j<i}, \{s_j^*\}_{j \le i, j<k}, \{s_k\}_{k=i}, \mathbf{q}) \\
&\qquad\qquad P(\{r_j = 0\}_{j<i} \,|\, \{s_j^*\}_{j<i}, \mathbf{q}) \\
&= \underset{s_k}{\arg\max}\, P(r_k = 1 \,|\, \{r_j = 0\}_{j<k}, S_{k-1}^*, s_k, \mathbf{q}) \qquad (3)
\end{aligned}$$

Here, we applied a logical equivalence, exploited additivity of exclusive events, rewrote the expectation of a binary event as its probability, exploited d-separation to remove irrelevant conditions, factorized each joint into a conditional and prior, and removed terms and factors independent of $s_k$. Thus, we need only maximize $s_k$'s probability of relevance conditioned on the query and previous selections (assumed irrelevant).

Next we evaluate the final query from (3) w.r.t. our graphical model of subtopic relevance from Figure 1:

$$\begin{aligned}
s_k^* &= \underset{s_k}{\arg\max}\, P(r_k = 1 \,|\, \{r_j = 0\}_{j<k}, S_{k-1}^*, s_k, \mathbf{q}) \\
&= \underset{s_k}{\arg\max} \sum_{t_1, \cdots, t_k, t} P(t|\mathbf{q}) P(t_k|s_k) \mathbb{I}[t_k = t] \prod_{i=1}^{k-1} P(t_i|s_i^*) \mathbb{I}[t_i \ne t] \\
&= \underset{s_k}{\arg\max} \sum_{t} P(t|\mathbf{q}) \sum_{t_k} P(t_k|s_k) \mathbb{I}[t_k = t] \prod_{i=1}^{k-1} \sum_{t_i} P(t_i|s_i^*) \mathbb{I}[t_i \ne t] \\
&= \underset{s_k}{\arg\max} \sum_{t} P(t|\mathbf{q}) P(t_k = t|s_k) \boxed{\prod_{i=1}^{k-1} (1 - P(t_i = t|s_i^*))}
\end{aligned}$$

Defining $\tilde{P}(t|S_{k-1}^*) = 1 - \square = 1 - \prod_{i=1}^{k-1}(1 - P(t_i = t|s_i^*))$, this is the probability that set $S_{k-1}^*$ already *covers* topic $t$ w.r.t. a *noisy-or* interpretation. Substituting $(1 - \tilde{P}(t|S_{k-1}^*))$ for $\square$ since $(1 - \tilde{P}(t|S_{k-1}^*)) = 1 - (1 - \square) = \square$, we obtain

$$\begin{aligned}
s_k^* &= \underset{s_k}{\arg\max} \sum_{t} P(t|\mathbf{q}) P(t_k = t|s_k) \left(1 - \tilde{P}(t|S_{k-1}^*)\right) \\
&= \underset{s_k}{\arg\max} \sum_{t} \underbrace{P(t|\mathbf{q}) P(t_k = t|s_k)}_{\text{query similarity}} \\
&\qquad - \sum_{t} \underbrace{P(t|\mathbf{q}) P(t_k = t|s_k) \tilde{P}(t|S_{k-1}^*)}_{\text{query-reweighted diversity}}. \qquad (4)
\end{aligned}$$

This final result in (4) has a clear interpretation as a diverse information retrieval algorithm where $D$ consists of

[3] The notation $\{\cdot\}_C$ refers to a (possibly empty) set of variables (or variable assignments) $\cdot$ that meet constraints $C$.

**Table 1: MMR vs Exp-1-call@$k$ on various ranking measures of diversity.**

| Testbed | Algorithm | ERR-IA@5 | ERR-IA@10 | ERR-IA@20 | $\alpha$-nDCG@5 | $\alpha$-nDCG@10 | $\alpha$-nDCG@20 | MAP-IA |
|---|---|---|---|---|---|---|---|---|
| TREC6-8 | MMR ($\lambda = .5$) | 0.0433 | 0.0548 | 0.0607 | 0.2310 | 0.2590 | 0.2728 | 0.0361 |
| | Exp-1-call@$k$ | 0.0456 | 0.0561 | 0.0621 | 0.2332 | 0.2602 | 0.2750 | 0.0365 |
| ClueWeb 2009 | MMR ($\lambda = .5$) | 0.0984 | 0.1085 | 0.1174 | 0.1500 | 0.1696 | 0.2045 | 0.0095 |
| | Exp-1-call@$k$ | 0.0972 | 0.1084 | 0.1158 | 0.1435 | 0.1698 | 0.1997 | 0.0123 |
| ClueWeb 2010 | MMR ($\lambda = .5$) | 0.1198 | 0.1422 | 0.1503 | 0.1516 | 0.2066 | 0.2339 | 0.0085 |
| | Exp-1-call@$k$ | 0.1211 | 0.1408 | 0.1550 | 0.1527 | 0.1984 | 0.2476 | 0.0093 |

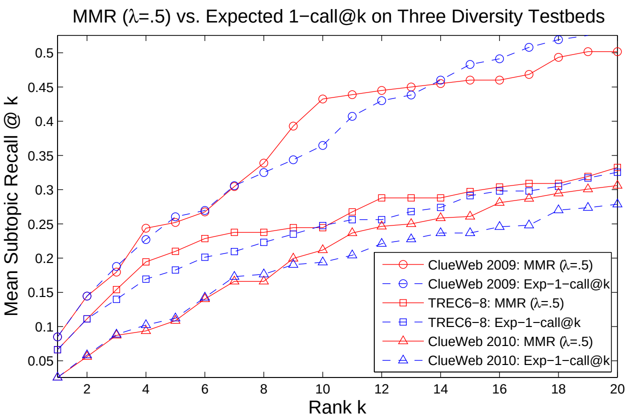


**Figure 2: MMR vs Exp-1-call@$k$ on subtopic recall.**

documents: at each step, $s_k$ is chosen so as to maximize a similarity function while minimizing a diversity penalty that increases as $S^*_{k-1}$'s coverage of *query-relevant* subtopics in document $s_k$ increases. Thus we have achieved our goal of *deriving* a diverse retrieval algorithm via optimization of Exp-1-call@$k$ in a latent subtopic model of binary relevance.

# 3. RELATED WORK

## 3.1 MMR

The result in (4) is strikingly similar to MMR — it contains two terms, one for query similarity and the other for result set diversification, where each term represents a similarity kernel — more specifically a *probability product kernel* (PPK) [11] that is an inner product of probability vectors (or more generally, functions). More formally, let $\mathbf{T}'$, $\mathbf{T}_k$, and $\mathbf{T}_{S^*_{k-1}}$ be respective topic probability vectors $P(t' = t|\mathbf{q})$, $P(t_k = t|s_k)$ and $\tilde{P}(t_k = t|S^*_{k-1})$ with vector indices for each topic $t \in T$. Then the similarity and diversity terms from (4) can be respectively written as

$$\sum_{t \in T} P(t' = t|\mathbf{q})P(t_k = t|s_k) \;=\; \langle \mathbf{T}', \mathbf{T}_k \rangle \text{ and} \tag{5}$$

$$\sum_{t \in T} P(t|\mathbf{q})P(t_k = t|s_k)\tilde{P}(t|S^*_{k-1}) \;=\; \langle \mathbf{T}_k, \mathbf{T}_{S^*_{k-1}} \rangle_{\mathbf{T}'}. \tag{6}$$

Here, we let $\langle \cdot, \cdot \rangle$ denote an inner product of two vectors and $\langle \cdot, \cdot \rangle_{\mathbf{v}}$ a $\mathbf{v}$-*reweighted* inner product, defined as in (6).

While having similarity and diversity terms similar to MMR, Exp-1-call@$k$ in (4) clearly differs from MMR:

1. While MMR's definition allows for any similarity function, not just PPKs, we note that *equating words to subtopics*, popular kernels like TF and TFIDF [13] can be viewed directly as PPKs if the TF and TFIDF vectors are $L_1$ normalized to represent probability vectors.
2. MMR uses a maximization term for diversity, whereas optimization of Exp-1-call@$k$ instead calls for a product (noisy-or) diversity term $\tilde{P}(t|S^*_{k-1})$. We note that a noisy-or reduces to a max when the subtopic probabilities are deterministic (0 or 1).
3. While MMR proposes a $\lambda$ term to explicitly trade off the similarity and diversity terms, the greedy optimization of Exp-1-call@$k$ in (4) yields no such trade-off term (or alternately, an implicit $\lambda = .5$). Although it seems a tunable $\lambda$ is not needed for maximizing Exp-1-call@$k$, it may be desirable when maximizing surrogate retrieval objectives (e.g., ranking objectives).
4. Optimizing Exp-1-call@$k$ introduces query-specific relevance into the diversification term as shown by the query topic ($\mathbf{T}'$) reweighted diversity function in (6).

To verify whether the differences between MMR and Exp-1-call@$k$ matter empirically, we compare the two algorithms across a number of metrics on three diversity testbeds: the TREC 6-8 Interactive Track[1] (17 queries) and 2009 and 2010 ClueWeb Diversity tasks of the TREC Web Track[2] (50 queries each). On these testbeds, we evaluate *mean subtopic recall@k* [19] (fraction of total annotated aspects/subtopics covered by a result set at rank $k$, averaged over queries), which is an appropriate loss function for the *set-level* metric (2) [6]. We also evaluate a variety of more recent *rank-based* diversity evaluation metrics such as intent-aware expected reciprocal rank (ERR-IA@$k$) [5], $\alpha$-nDCG@$k$ [7], and intent-aware mean average precision (MAP-IA) [1].

We use MMR with $\lambda = 0.5$ to match the equal weighting of similarity and diversity in Exp-1-call@$k$. An LDA [2] topic model is trained on the top-100 OKAPI BM25 [12] results for each query (on its respective collection) and these subtopic distributions are used for the similarity and diversity kernels in both algorithms: for MMR we choose $\mathrm{Sim}_1$ and $\mathrm{Sim}_2$ kernels as in (5) — effectively LDA variants of latent semantic indexing (LSI) [8] kernels; for Exp-1-call@$k$, we use the similarity and diversity kernels respectively defined in (5) and (6). Both MMR and Exp-1-call@$k$ are used to rank the top-20 documents from the top-100 OKAPI BM25 results.

Results in Table 1 and Figure 2 show the performances of MMR and Exp-1-call@$k$ on the three diversity testbeds across various diversity measures; although there are minor performance differences, we note that these differences are not statistically significant w.r.t. 95% confidence intervals. Nonetheless, the results appear to indicate that the structural similarities in the use of MMR and the optimization of Exp-1-call@$k$ outweigh the differences in this evaluation.

## 3.2 Other Diversification Approaches

Recent years have seen numerous proposals for diversification approaches and here we summarize the relationship between optimization of Exp-1-call@$k$ and representatives of these alternative approaches:

**Portfolio Theory:** [16] motivates diversification in set-based information retrieval by a risk-minimizing portfolio selection approach. Viewing a result set as an investment portfolio with the objective to maximize return while minimizing risk, the derived result of [16] mimics both MMR and Exp-1-call@$k$ in that the similarity term may be viewed as *expected portfolio payoff* (relevance) and the diversity term may be viewed as *expected portfolio risk*, which increases as the correlations between documents in the result set increase. One major difference in this framework is that rather than computing the diversity term via a max (MMR) or product (Exp-1-call@$k$) the portfolio theory derivation uses a summation — we examine the implications of this next.

**Set Covering:** Yue and Joachims [18] propose a set covering approach for training SVMs to predict diverse result sets for information retrieval. In their work, they equate subtopics with words and build a loss function for SVM training that penalizes result sets according to the sum of weights of query-relevant words *not* covered by the result set. While their approach provides a "hard" set-covering view of diversity, we note that an expansion of $\tilde{P}(t|S^*_{k-1})$ used in the diversity term of (4) provides a "soft" latent set-covering interpretation; that is, $s_k$ is chosen so as to best cover (in a probabilistic sense) the latent topic space not already covered by $\{s^*_1, \ldots, s^*_{k-1}\}$. Formally, expanding the product in $\tilde{P}(t|S^*_{k-1}) = \prod_{i=1}^{k-1}(1 - P(t_i = t|s^*_i))$, collecting terms and writing it as a series, we arrive at a form that reflects the inclusion-exclusion principle applied to the calculation of probability that topic $t$ is covered by $\{s^*_1, \ldots, s^*_{k-1}\}$:

$$\prod_{i=1}^{k-1}(1 - P(t_i = t|s^*_i))$$
$$= 1 - \left[\sum_{i=1}^{k-1} P(t_i = t|s^*_i) - \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} P(t_i = t|s^*_i)P(t_j = t|s^*_j) + \cdots - (-1)^{k-1}\prod_{i=1}^{k-1} P(t_i = t|s^*_i)\right] \quad (7)$$

This result has a natural interpretation: the first summation term determines the coverage of topic $t$ by each document $s_i$ $(1 \leq i \leq k-1)$ currently in the result set, the second double summation term corrects the first term by removing the joint probability mass from all pairs of documents that was double counted, and so on according to the principle of inclusion-exclusion. (7) not only provides a probabilistic set covering view of Exp-1-call@$k$, but it also suggests that a portfolio approach to diversity using only the first summation would overcount each document's contribution to the diversity metric according to this set covering perspective.

**Subtopic Relevance Models:** We use a subtopic relevance model that is a simplified version of the model in [10] with fewer dependence assumptions. In other work, Zhai *et al* [19] present an empirical risk minimization view of dependent document retrieval from a subtopic perspective, where they derive a formalization of the *greedy* selection step that is similar to MMR and to a lesser extent, Exp-1-call@$k$.

**Set-based Relevance Objectives:** Chen and Karger [6], whose derivation we extended, directly optimize 1-call@$k$, but their intention is not to formalize MMR and instead use naïve Bayes to directly evaluate (3). Agrawal et al [1] and Santos et al (xQuad) [14] both specify set-based diversity metrics *very* similar to Exp-1-call@$k$ but do not provide formal derivations as we have done in this work.

**Ranking Based Objectives:** Finally, returning to our introductory motivation, Wang and Zhu [17] have shown that natural forms of result set diversification arise via the optimization of average precision [3] and reciprocal rank [15]. Both of these methods share the view of directly optimizing a *ranking-based* objective, whereas this paper proposes a novel derivation from the alternate view of optimizing a *set-based* objective w.r.t. a subtopic model of relevance. However, even though Exp-1-call@$k$ is a set-based objective, an indirect consequence of (and motivation for) greedily optimizing it is that documents added earlier yield a greater increase in objective than those added later; this yields a natural rank ordering on the greedy Exp-1-call@$k$ result set.

# 4. CONCLUSION

This paper presented a new derivation of diverse retrieval by directly optimizing the expected 1-call@$k$ set-based retrieval objective w.r.t. a latent subtopic model of binary relevance. This result both motivates and contrasts with various related diversification approaches, providing a new theoretical basis for the investigation of diverse retrieval.


## Acknowledgements

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.


# 5. REFERENCES


[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *WSDM-09*, 5–14, ACM, 2009.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research (JLMR)*, 3:993–1022, 2003.

[3] C. Buckley and E. M. Voorhees. Evaluating evaluation measure stability. In *SIGIR-00*, 33–40, ACM, 2000.

[4] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reording documents and producing summaries. In *SIGIR-98*, 335–336, ACM, 1998.

[5] O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *CIKM-09*, 621–630, ACM, 2009.

[6] H. Chen and D. R. Karger. Less is more: Probabilistic models for retrieving fewer relevant documents. In *SIGIR-06*, 429–436, ACM, 2006.

[7] C. Clarke, M. Kolla, G. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR-08*, 659–666, ACM, 2008.

[8] S. Deerwester, S. T. Dumaisand, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

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

[10] S. Guo and S. Sanner. Probabilistic latent maximal marginal relevance. In *SIGIR-10*, 833–834, ACM, 2010.

[11] T. Jebara, R. Kondor, and A. Howard. Probability product kernels. *Journal of Machine Learning Research (JLMR)*, 5:819–844, 2004.

[12] S. Robertson and S. Walker. Some simple approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR-94*, 345–354, ACM, 1994.

[13] G. Salton and M. McGill. *Introduction to modern information retrieval*. McGraw-Hill, 1983.

[14] R. L. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *WWW-10*, 881–890, ACM, 2010.

[15] E. M. Voorhees. TREC-8 question answering track report. In *Proceedings of the 8th Text Retrieval Conference*, pages 77–82, 1999.

[16] J. Wang and J. Zhu. Portfolio theory of information retrieval. In *SIGIR-09*, 115–122. ACM, 2009.

[17] J. Wang and J. Zhu. On statistical analysis and optimization of information retrieval effectiveness metrics. In *SIGIR-10*, 226–233, ACM, 2010.

[18] Y. Yue and T. Joachims. Predicting diverse subsets using structural SVMs. In *ICML-08*, 1224–1231, ACM, 2008.

[19] C. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In *SIGIR-03*, pages 10–17. ACM, 2003.