An Axiomatic Approach for Result Diversification

Sreenivas Gollapudi
Microsoft Search Labs
Microsoft Research
sreenig@microsoft.com

Aneesh Sharma*
Institute for Computational and Mathematical Engineering, Stanford University
aneeshs@stanford.edu

ABSTRACT

Understanding user intent is key to designing an effective ranking system in a search engine. In the absence of any explicit knowledge of user intent, search engines want to diversify results to improve user satisfaction. In such a setting, the probability ranking principle-based approach of presenting the most relevant results on top can be sub-optimal, and hence the search engine would like to trade-off relevance for diversity in the results.

In analogy to prior work on ranking and clustering systems, we use the axiomatic approach to characterize and design diversification systems. We develop a set of natural axioms that a diversification system is expected to satisfy, and show that no diversification function can satisfy all the axioms simultaneously. We illustrate the use of the axiomatic framework by providing three example diversification objectives that satisfy different subsets of the axioms. We also uncover a rich link to the facility dispersion problem that results in algorithms for a number of diversification objectives. Finally, we propose an evaluation methodology to characterize the objectives and the underlying axioms. We conduct a large scale evaluation of our objectives based on two data sets: a data set derived from the Wikipedia disambiguation pages and a product database.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and RetrievalInformation Search and Retrieval

General Terms

Algorithms, Theory, Performance

Keywords

Search engine, Diversification, Approximation Algorithms, Axiomatic framework, Facility dispersion, Wikipedia

1. INTRODUCTION

In the current search model, the user expresses her information need with the use of a few query terms. In such a scenario, the small number of terms often specify the intent implicitly. In the absence of explicit information representing user intent, the search engine needs to "guess" the results that are most likely to satisfy different intents. In particular, for an ambiguous query such as *eclipse*, the search engine could either take the probability ranking principle approach of taking the "best guess" intent and showing the results, or it could choose to present search results that maximize the probability of a user with a random intent finding *at least one* relevant document on the results page. This problem of the user not finding any *any* relevant document in her scanned set of documents is defined as *query abandonment*. Result diversification lends itself as an effective solution to minimizing query abandonment [1, 9, 18].

Intuitively, diversification implies a trade-off between having more relevant results of the "correct" intent and having diverse results in the top positions for a given query [6, 8]. Hence the twin objectives of being diverse and being relevant compete with each other, and any diversification system must figure out how to trade-off these objectives appropriately. This often results in the diversification problem being characterized as a bi-criteria optimization problem. Diversification can be viewed as combining both ranking (presenting more relevant results in the higher positions) and clustering (grouping documents satisfying similar intents) and therefore addresses a loosely defined goal of picking a set of most relevant but novel documents. This has resulted in the development of a set of very different objective functions and algorithms ranging from combinatorial optimizations [6, 18, 1] to those based on probabilistic language models [8, 22]. The underlying principles supporting these techniques are often different and therefore admit different trade-off criteria. Given the importance of the problem there has been relatively little work aimed at understanding result diversification independent of the objective functions or the algorithms used to solve the problem.

In this work, we initiate an axiomatic study of result diversification. We propose a set of simple properties that any diversification system ought to satisfy and these properties help serve as a *basis* for the space of objective functions for result diversification. Generally, a diversification function can be thought of as taking two application specific inputs *viz,* a relevance function that specifies the relevance of document for a given query, and a distance function that captures the pairwise similarity between any pair of documents in the set of relevant results for a given query. In the context of web search, one can use the search engine's ranking

*Work done while author was an intern at Microsoft Search Labs

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

function[1] as the relevance function. The characterization of the distance function is not that clear. In fact, designing the right distance function is key for having effective result diversification. For example, by restricting the distance function to be a *metric* by imposing the triangle inequality $d(u,w) \leq d(u,v) + d(v,w)$ for all $u, v, w \in U$, we can exploit efficient approximation algorithms to solve certain class of diversification objectives (see Section 3).

1.1 Contributions of this study

In this work, we propose a set of natural axioms for result diversification that aid in the choice of an objective function and therefore help in constraining the resulting solution. Our work is similar in spirit to recent work on axiomatization of ranking and clustering systems [2, 12]. We study the functions that arise out of the requirement of satisfying a set of simple properties and show an *impossibility result* which states that there exists no diversification function f that satisfies all the properties. We state the properties in Section 2.

Although we do not aim to completely map the space of objective functions in this study, we show that some diversification objectives reduce to different versions of the well-studied *facility dispersion* problem. Specifically, we pick three functions that satisfy different subsets of the properties and characterize the solutions obtained by well-known approximation algorithms for each of these functions. We also characterize some of the objective functions defined in earlier works [1, 18, 8] using the axioms.

Finally, we do a preliminary characterization of the choice of an objective (and its underlying properties) using the natural measures of *relevance* and *novelty*. We posit that different subsets of axioms will exhibit different trade-offs between these measures. Towards this end, we provide an evaluation methodology that computes the measures based on the disambiguation pages in Wikipedia[2], which is the largest public-domain evaluation data set used for testing a diversification system (see Section 5). Further, we consider two distance functions - a semantic distance function and a categorical distance function (see Section 4) - to test the effectiveness of the objectives under two different application contexts.

1.2 Related Work

The early work of Carbonell and Goldstein [6] described the trade-off between relevance and novelty via the choice of a parametrized objective function. Subsequent work on *query abandonment* by Chen and Karger [8] is based on the idea that documents should be selected sequentially according to the probability of the document being relevant *conditioned* on the documents that come before. Das Sarma *et. al.* [18], solved a similar problem by using bypass rates of a document to measure the overall likelihood of a user bypassing all documents in a given set. Thus, the objective in their setting was to produce a set that minimized likelihood of completely getting bypassed.

Agrawal *et. al.,* [1] propose a diversification objective that tries to maximize the likelihood of finding a relevant document in the top-k positions given the categorical information of the queries and documents. Other works on topical diversification include [23]. Zhai and Lafferty [22, 20] propose a risk minimization framework for information retrieval that allows a user to define an arbitrary *loss* function over the set of returned documents. Vee *et. al.,* [19] proposed a method for diversifying query results in online shopping applications wherein the query is presented in a structure form using online forms.

Our work is based on axiomatizations of ranking and clustering systems [3, 12, 2]. Kleinberg [12] proposed a set of three natural axioms for clustering functions and showed that no clustering function satisfies all three axioms. Altman and Tennenholtz [2] study ranking functions that combine individual votes of agents into a social ranking of the agents and compare them to social choice welfare functions which were first proposed in the classical work on social choice theory by Kenneth Arrow [3].

One of the contributions of our work is the mapping of diversification functions to those used in facility dispersion [15, 14]. The reader will find a useful literature in the chapter on facility dispersion in [16, 7].

2. AXIOMATIC FRAMEWORK

This section introduces the axiomatic framework and fixes the notation to be used in the remainder of the paper. We are given a set $U = \{u_1, u_2, \ldots, u_n\}$ of $n \geq 2$ of documents, and a set (we'll assume this to be finite for now) of queries Q. Now, given a query $q \in Q$ and an integer k, we want to output a subset $S_k \subseteq U$ of documents[3] that is simultaneously both relevant and diverse. The relevance of each document is specified by a function $w : U \times Q \rightarrow \mathbf{R}^+$, where a higher value implies that the document is more relevant to a particular query. The diversification objective is intuitively thought of as giving preference to dissimilar documents. To formalize this, we define a distance function $d : U \times U \rightarrow \mathbf{R}^+$ between the documents, where smaller the distance, the more similar the two documents are. We also require the distance function to be discriminative, i.e. for any two documents $u, v \in U$, we have $d(u,v) = 0$ if and only if $u = v$, and symmetric, i.e $d(u,v) = d(v,u)$. Note that the distance function need not be a metric.

We restrict attention to the *set selection* problem instead of the search problem of selecting a *ranked list* as this is clearly a simpler problem. In particular, the approach we will take is to find the best set and then rank it in order of relevance. Formally, the set selection function $f : 2^U \times Q \times w \times d \rightarrow \mathbf{R}$ can be thought of as assigning scores to all possible subsets of U, given a query $q \in Q$, a weight function $w(\cdot)$, a distance function $d(\cdot,\cdot)$. Fixing q, $w(\cdot)$, $d(\cdot,\cdot)$ and a given integer $k \in \mathbf{Z}^+$ $(k \geq 2)$, the objective is to select a set $S_k \subseteq U$ of documents such that the value of the function f is maximized, i.e. the objective is to find

$$S_k^* = \underset{\substack{S_k \subseteq U \\ |S_k| = k}}{\operatorname{argmax}} f(S_k, q, w(\cdot), d(\cdot,\cdot))$$

where all arguments other than the set S_k are fixed inputs to the function.

An important observation is that the diversification framework is underspecified and even if one assumes that the relevance and distance functions are provided, there are many possible choices for the objective function f. These functions could trade-off relevance and similarity in different ways, and

[1]See [17] and references therein.

[2]http://en.wikipedia.org

[3]We will denote the size of the set by the subscript, i.e. $|S_k| = k$

one needs to specify criteria for selection among these functions. A natural mathematical approach in such a situation is to provide axioms that any diversification system should be expected to satisfy and therefore provide *some* basis of comparison between different objective functions.

2.1 Axioms of diversification

We propose that f is such that it satisfy the set of axioms given below, each of which seems intuitive for the setting of diversification. In addition, we show that any proper subset of these axioms is *maximal*, i.e. no diversification function can satisfy all these axioms. This provides a natural method of selecting between various objective functions, as one can choose the essential properties for any particular diversification system. In section 3, we will illustrate the use of the axioms in choosing between different diversification objectives. Before we state the axioms, we state the following notation. Fix any q, $w(\cdot)$, $d(\cdot,\cdot)$, k and f, such that f is maximized by S_k^*, i.e., $S_k^* = \text{argmax}_{S_k \subseteq U} f(S_k, q, w(\cdot), d(\cdot,\cdot))$.

1. **Scale invariance**: Informally, this property states that the set selection function should be insensitive to the scaling of the input functions. Consider the set optimal set S_k^*. Now, we require f to be such that we have $S_k^* = \text{argmax}_{S_k \subseteq U} f(S_k, q, \alpha \cdot w(\cdot), \alpha \cdot d(\cdot,\cdot))$ for any fixed positive constant $\alpha \in \mathbf{R}$, $\alpha > 0$, i.e. S_k^* still maximizes f even if all relevance and distance values are scaled by some constant.

2. **Consistency**: Consistency states that making the output documents more relevant and more diverse, and making other documents less relevant and less diverse should not change the output of the ranking. Now, given any two functions $\alpha : U \to \mathbf{R}^+$ and $\beta : U \times U \to \mathbf{R}^+$, we modify the relevance and weight functions as follows:

$$w(u) = \begin{cases} w(u) + \alpha(u) & , \ u \in S_k^* \\ w(u) - \alpha(u) & , \ \text{otherwise} \end{cases}$$

$$d(u,v) = \begin{cases} d(u,v) + \beta(u,v) & , \ u,v \in S_k^* \\ d(u,v) - \beta(u,v) & , \ \text{otherwise} \end{cases}$$

   The ranking function f must be such that it is still maximized by S_k^*.

3. **Richness**: Informally speaking, the richness condition states that we should be able to achieve any possible set as the output, given the right choice of relevance and distance function. Formally, there exists some $w(\cdot)$ and $d(\cdot,\cdot)$ such that for any $k \geq 2$, there is a unique S_k^* which maximizes f.

4. **Stability**: The stability condition seeks to ensure that the output set does not change arbitrarily with the output size, i.e., the function f should be defined such that $S_k^* \subset S_{k+1}^*$.

5. **Independence of Irrelevant Attributes**: This axiom states that the score of a set is not affected by most attributes of documents outside the set. Specifically, given a set S, we require the function f to be such that $f(S)$ is independent of values of:

   - $w(u)$ for all $u \notin S$.
   - $d(u,v)$ for all $u, v \notin S$.

6. **Monotonicity**: Monotonicity simply states that the addition of any document does not decrease the score of the set. Fix any $w(\cdot)$, $d(\cdot,\cdot)$, f and $S \subseteq U$. Now, for any $x \notin S$, we must have

$$f(S \cup \{x\}) \geq f(S)$$

7. **Strength of Relevance**: This property ensures that no function f ignores the relevance function. Formally, we fix some $w(\cdot)$, $d(\cdot,\cdot)$, f and S. Now, the following properties should hold for any $x \in S$:

   (a) There exist some real numbers $\delta_0 > 0$ and $a_0 > 0$, such that the condition stated below is satisfied after the following modification: obtain a new relevance function $w'(\cdot)$ from $w(\cdot)$, where $w'(\cdot)$ is identical to $w(\cdot)$ except that $w'(x) = a_0 > w(x)$. The remaining relevance and distance values could decrease arbitrarily. Now, we must have

$$f(S, w'(\cdot), d(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) + \delta_0$$

   (b) If $f(S \setminus \{x\}) < f(S)$, then there exist some real numbers $\delta_1 > 0$ and $a_1 > 0$ such that the following condition holds: modify the relevance function $w(\cdot)$ to get a new relevance function $w'(\cdot)$ which is identical to $w(\cdot)$ except that $w'(x) = a_1 < w(x)$. Now, we must have

$$f(S, w'(\cdot), d(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) - \delta_1$$

8. **Strength of Similarity**: This property ensures that no function f ignores the similarity function. Formally, we fix some $w(\cdot)$, $d(\cdot,\cdot)$, f and S. Now, the following properties should hold for any $x \in S$:

   (a) There exist some real numbers $\delta_0 > 0$ and $b_0 > 0$, such that the condition stated below is satisfied after the following modification: obtain a new distance function $d'(\cdot,\cdot)$ from $d(\cdot,\cdot)$, where we increase $d(x,u)$ for the required $u \in S$ to ensure that $\min_{u \in S} d(x,u) = b_0$. The remaining relevance and distance values could decrease arbitrarily. Now, we must have

$$f(S, w(\cdot), d'(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) + \delta_0$$

   (b) If $f(S \setminus \{x\}) < f(S)$, then there exist some real numbers $\delta_1 > 0$ and $b_1 > 0$ such that the following condition holds: modify the distance function $d(\cdot,\cdot)$ by decreasing $d(x,u)$ to ensure that $\max_{u \in S} d(x,u) = b_1$. Call this modified distance function $d'(\cdot,\cdot)$. Now, we must have

$$f(S, w(\cdot), d'(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) - \delta_1$$

Given these axioms, a natural question is to characterize the set of functions f that satisfy these axioms. A somewhat surprising observation here is that it is impossible to satisfy all of these axioms simultaneously (proof is in appendix):

Theorem 1. No function f satisfies all 8 axioms stated above.

Theorem 1 implies that any subset of the above axioms is maximal. This result allows us to naturally characterize the set of diversification functions, and selection of a particular function reduces to deciding upon the subset of axioms (or properties) that the function is desired to satisfy. The following sections explore this idea further and show that the axiomatic framework could be a powerful tool in choosing between diversification function. Another advantage of the framework is that it allows a theoretical characterization of the function which is independent of the specifics of the diversification system such as the distance and the relevance function.

3. OBJECTIVES AND ALGORITHMS

In light of the impossibility result shown in Theorem 1, we can only hope for diversification functions that satisfy a subset of the axioms. We note that the list of such functions is possibly quite large, and indeed several such functions have been previously explored in the literature (see [8,18,1], for instance). Further, proposing a diversification objective may not be useful in itself unless one can actually find algorithms to optimize the objective. In this section, we aim to address both of the above issues: we demonstrate the power of the axiomatic framework in choosing objectives, and also propose reductions from a number of natural diversification objectives to the well-studied combinatorial optimization problem of facility dispersion [16]. In particular, we propose three diversification objectives in the following sections, and provide algorithms that optimize those objectives. We also present a brief characterization of the objective functions studied in earlier works [1, 18, 8]. We will use the same notation as in the previous section and have the objective as $S_k^* = \operatorname{argmax}_{\substack{S_k \subseteq U \\ |S_k|=k}} f(S_k, q, w(\cdot), d(\cdot,\cdot))$, where f would vary from one function to another. Also, we assume $w(\cdot)$, $d(\cdot,\cdot)$ and k to be fixed here and hence use the shorthand $f(S)$ for the function.

3.1 Max-sum diversification

A natural bi-criteria objective is to maximize the sum of the relevance and dissimilarity of the selected set. This objective can be encoded in terms of our formulation in terms of the function $f(S)$, which is defined as follows:

$$f(S) = (k-1)\sum_{u \in S} w(u) + 2\lambda \sum_{u,v \in S} d(u,v) \qquad (1)$$

where $|S| = k$, and $\lambda > 0$ is a parameter specifying the trade-off between relevance and similarity. Observe that we need to scale up the first sum to balance out the fact that there are $\frac{k(k-1)}{2}$ numbers in the similarity sum, as opposed to k numbers in the relevance sum. We first characterize the objective in terms of the axioms.

Remark 1. The objective function given in equation 1 satisfies all the axioms, except stability.

This objective can be recast in terms of a facility dispersion objective, known as the MaxSumDispersion problem. The MaxSumDispersion problem is a facility dispersion problem having the objective maximizing the sum of all pairwise distances between points in the set S which we show to be equivalent to equation 1. To this end, we define a new distance function $d'(u,v)$ as follows:

$$d'(u,v) = w(u) + w(v) + 2\lambda d(u,v) \qquad (2)$$

Input : Universe U, k
Output: Set S ($|S| = k$) that maximizes $f(S)$

Initialize the set $S = \emptyset$
for $i \leftarrow 1$ **to** $\lfloor \frac{k}{2} \rfloor$ **do**
  Find $(u,v) = \operatorname{argmax}_{x,y \in U} d(x,y)$
  Set $S = S \cup \{u, v\}$
  Delete all edges from E that are incident to u or v
end
If k is odd, add an arbitrary document to S

Algorithm 1: Algorithm for MaxSumDispersion

It is not hard to see the following claim (proof skipped):

Claim 1. *$d'(\cdot,\cdot)$ is a metric if the distance $d(\cdot,\cdot)$ constitutes a metric.*

Further, note that for some $S \subseteq U$ ($|S| = k$), we have:

$$\sum_{u,v \in S} d'(u,v) = (k-1)\sum_{u \in S} w(u) + 2\lambda \sum_{u,v \in S} d(u,v)$$

using the definition of $d'(u,v)$ and the fact that each $w(u)$ is counted exactly $k-1$ times in the sum (as we consider the complete graph on S). Hence, from equation 1 we have that

$$f(S) = \sum_{u,v \in S} d'(u,v)$$

But this is also the objective of the MaxSumDispersion problem described above where the distance metric is given by $d'(\cdot,\cdot)$.

Given this reduction, we can map known results about MaxSumDispersion to the diversification objective. First of all, we observe that maximizing the objective in equation 1 is NP-hard, but there are known approximation algorithms for the problem. In particular, there is a 2-approximation algorithm for the MaxSumDispersion problem [13, 11] (for the metric case) and is given in algorithm 1. Hence, we can use algorithm 1, for the max-sum objective stated in equation 1.

3.2 Max-min diversification

The second bi-criteria objective we propose, maximizes the *minimum* relevance and dissimilarity of the selected set. This objective can be encoded in terms of our formulation in terms of the function $f(S)$, which is defined as follows:

$$f(S) = \min_{u \in S} w(u) + \lambda \min_{u,v \in S} d(u,v) \qquad (3)$$

where $|S| = k$, and $\lambda > 0$ is a parameter specifying the trade-off between relevance and similarity. Here is the characterization of the objective in terms of the axioms:

Remark 2. The diversification objective given in equation 3 satisfies all the axioms except consistency and stability.

We proceed as before to link this objective to facility dispersion, and the dispersion objective that is relevant in this case is MaxMinDispersion. The objective for the MaxMinDispersion problem is: $g(P) = \min_{v_i, v_j \in P} d(v_i, v_j)$, which we now show to be equivalent to equation 3. As before, we combine the objective in equation 1 in terms of a single metric, with which we can then solve the MaxMinDispersion problem. To this end, we define a new distance

Input : Universe U, k
Output: Set S ($|S| = k$) that maximizes $f(S)$

Initialize the set $S = \emptyset$; Find $(u, v) = \text{argmax}_{x,y \in U} d(x, y)$ and set $S = \{u, v\}$; For any $x \in U \setminus S$, define $d(x, S) = \min_{u \in S} d(x, u)$;
while $|S| < k$ **do**
  Find $x \in U \setminus S$ such that $x = \text{argmax}_{x \in U \setminus S} d(x, S)$;
  Set $S = S \cup \{x\}$;
end

Algorithm 2: Algorithm for MAXMINDISPERSION

function $d'(u, v)$ as follows: Now, we show how to use this algorithm for the bi-criteria objective given in equation 3. In order to do this, we again need to combine our objective function in terms of a single metric, with which we can then solve the MAXMINDISPERSION problem. Hence, we define a new distance function $d'(u, v)$ as follows:

$$d'(u, v) = \frac{1}{2}(w(u) + w(v)) + \lambda d(u, v) \quad (4)$$

It is not hard to see the following claim (proof skipped):

CLAIM 2. *The distance $d'(\cdot, \cdot)$ forms a metric if the distance $d(\cdot, \cdot)$ forms a metric.*

Further, note that for some $S \subseteq U$ ($|S| = k$), we have:

$$\min_{u,v \in S} d'(u, v) = \min_{u \in S} w(u) + \lambda \min_{u,v \in S} d(u, v) = f(S)$$

from equation 3. This is also the objective from the MAXMINDISPERSION problem where the distance metric is given by $d'(\cdot, \cdot)$. Hence, we can use algorithm 2 to approximately maximize the objective stated in equation 3.

Again, we can map known results about the MAXMINDISPERSION problem to equation 3, such as NP-hardness. We describe a 2-approximation algorithm in algorithm 2 that was proposed in [15], and refer the reader to [15] for further results.

3.3 Mono-objective formulation

The third and final objective we will explore does not relate to facility dispersion as it combines the relevance and the similarity values into a single value for each *document* (as opposed to each edge for the previous two objectives). The objective can be stated in the notation of our framework in terms of the function $f(S)$, which is defined as follows:

$$f(S) = \sum_{u \in S} w'(u) \quad (5)$$

where the new relevance value $w'(\cdot)$ for each document $u \in U$ is computed as follows:

$$w'(u) = w(u) + \frac{\lambda}{|U| - 1} \sum_{v \in U} d(u, v)$$

for some parameter $\lambda > 0$ specifying the trade-off between relevance and similarity. Intuitively, the value $w'(u)$ computes the "global" importance (i.e. not with respect to any particular set S) of each document u. The axiomatic characterization of this objective is as follows:

Remark 3. The objective in equation 5 satisfies all the axioms except consistency.

Also observe that it is possible to exactly optimize objective 5 by computing the value $w'(u)$ for all $u \in U$ and then picking the documents with the top k values of u for the set S of size k.

3.4 Other objective functions

We note that the link to the facility dispersion problem explored above is particularly rich as many dispersion objectives have been studied in the literature (see [16,7]). We only explore two objectives here in order to illustrate the use of the framework, and also because other objectives share common algorithms. For instance, the MAXMSTDISPERSION problem seeks to maximize the weight of the minimum spanning tree of the selected set. It turns out that algorithm 2 is the best known approximation algorithm for this objective as well, although the approximation factor is 4.

The axiomatic framework can also be used to characterize diversification objectives that have been proposed previously (we note that the characterization might be non-trivial to obtain as one needs to cast the objectives in our setting). In particular, we point out that the DIVERSIFY objective function in [1] as well as the MINQQUERYABANDONMENT formulations proposed in [18] violate the stability and the independence of irrelevant attributes axioms.

4. THE DISTANCE FUNCTION

The diversification algorithm only partially specifies the framework, and to complete the specification, we also need to specify the distance and the relevance functions. We describe the relevance function later in the experiments, and focus on the distance function here as it depends on the content of the data set being used, and might be of independent interest. Specifically, we describe the distance function for web pages and product hierarchies.

4.1 Semantic distance

Semantic distance is based on content similarity between two pages. Instead of using the whole of a page in the similarity computation, we use simple sketching algorithms based on the well known min-hashing scheme [5,10] to compute the sketch of a document and then apply Jaccard similarity between sketches to compute the pairwise semantic distance between the documents. We state this more formally. Fix a hash function h that maps elements from the universe U to a real number uniformly at random in $[0, 1]$. Then the min-hash of a set of elements A is defined as $MH_h(A) = \text{argmin}_x\{h(x)|x \in A\}$. Therefore, $MH_h(A)$ is the element in A whose hash value corresponds to the minimum value among all values hashed into the range $[0, 1]$. This computation can be easily extended to multi-sets wherein the min-hash of a bag A is computed as

$$MH(A) = \underset{x}{\text{argmin}}\{h(x, i)|x \in A, 1 \leq i \leq c_x\},$$

where c_x is the frequency of element x in A. Thus, given k hash functions $h_1, \cdots, h_k$, the sketch of a document d is $S(d) = \{MH_{h_1}(d), MH_{h_2}(d), \ldots, MH_{h_k}(d)\}$. We can now compute the similarity between two documents as

$$sim(u, v) = \frac{|S(u) \cap S(v)|}{|S(u) \cup S(v)|}$$

We note that Jaccard similarity is known to be a metric. However, one issue that makes such a computation of $sim(u, v)$

ineffective is a large difference in lengths of u and v. One simple approach to handle such random documents, would be to discard documents that have a small sketch size.

Thus, one characterization of the semantic distance between two documents u and v could be

$$d(u,v) = 1 - sim(u,v) \quad (6)$$

4.2 Categorical distance

The semantic distance is not applicable in all contexts. One scenario is when two "intuitively" similar documents like `http://www.apache.org/` and `http://www.apache.org/docs` actually might have very different sketches. However, these documents are 'close' to each other with respect to the distance in the underlying web graph. However, computing the pairwise distance of two web pages based on their web graph connectivity can be very expensive. Taxonomies offer a succinct encoding of distances between pages wherein the category of the page can be viewed as its sketch. Therefore, the distance between the same pages on similar topics in the taxonomy is likely to be small. In this context, we use a weighted tree distance [4] as a measure of similarity between two categories in the taxonomy. Distance between two nodes u and v in the tree is computed as

$$d(u,v) = \sum_{i=1}^{l(u)} \frac{1}{2^{e(i-1)}} + \sum_{i=1}^{l(v)} \frac{1}{2^{e(i-1)}} \quad (7)$$

where $e \geq 0$ and $l(\cdot)$ is the depth of the given node in the taxonomy. This definition of a weighted tree distance reduces to the well-known tree distance (measured in path length through the least common ancestor – $lca(u,v)$) when e is set to zero and to the notion of hierarchically separated trees (due to Bartal [4]) for greater values of e. Thus, nodes corresponding to more general categories (e.g.,`/Top/Health` and `/Top/Finance`) are more separated than specific categories (e.g., `/Top/Health/Geriatrics/Osteoporosis` and `/Top/Health/Geriatrics/Mental Health`). We can extend this notion of distance to the case where a document belongs to multiple categories (with different confidences), one cannot equate the categorical distance to the distance between nodes in the taxonomy. Given two documents x and y and their category information $\mathcal{C}_x$ and $\mathcal{C}_y$ respectively, we define their categorical distance as

$$d_c(x,y) = \sum_{u \in \mathcal{C}_x, v \in \mathcal{C}_y} \min(\mathcal{C}_x(u), \mathcal{C}_y(v)) \operatorname*{argmin}_v d(u,v) \quad (8)$$

where $\mathcal{C}_x(u)$ denotes the confidence (or probability) of document x belonging to category u.

5. EXPERIMENTAL EVALUATION

Recall that we used the axiomatic framework to characterize differences between diversification objectives in section 3. We now switch gears to investigate the other method of distinguishing between various objectives, namely through their experimental behavior. In this section, we characterize the choice of the objective function and its underlying axioms using two well-known measures *relevance* and *novelty*. We demonstrate the usefulness of the diversification framework by conducting two sets of experiments.

In the first set of experiments, which we call *semantic disambiguation*, we compare the performance of the three diversification algorithms using the set of Wikipedia disambiguation pages,[4] as the ground truth. For instance, the Wikipedia disambiguation page for *jaguar*[5] lists several different meanings for the word, including jaguar the cat and jaguar cars, along with links to their Wikipedia pages. The titles of the disambiguation pages in Wikipedia (for instance, *jaguar* in the above example) serve as the query set for our evaluation. This is a natural set of queries as they have been "labeled" by human editors as being ambiguous and the search engine would want to cover their different meanings in the search results. The data set also has the advantage of being large scale (about 2.5 million documents) and representative of the words that naturally occur on the Web and in web search (Wikipedia is one of the primary results surfaced for many informational queries). In addition, unlike query data for search engines, the Wikipedia data is in public domain.

In the second set of experiments, which we call *product disambiguation*, we demonstrate the efficacy of our algorithms in diversifying product searches using the the categorical distance function in Section 4.2. We use a product catalog of about $41,799,440$ products and 6808 product categories. We use a logistic regression classifier[6] to classify the queries into the product taxonomy.

5.1 Semantic disambiguation

We will refer to the queries drawn from the title of Wikipedia disambiguation pages as *ambiguous queries*. Let us denote the set of these titles by $\mathcal{Q}$ and the set of meanings or topics (i.e. the different Wikipedia pages) associated with each disambiguation title q by $\mathcal{S}_q$. Now, we posit that an ideal set of diverse search results for query q would represent a large selection of the topics in S_q (and not necessarily the Wikipedia pages) within its top set of results.

To associate any search result page with a Wikipedia topic, we compute the semantic distance between the web page and the Wikipedia topic page using the distance function described in Section 4.1. Thus, for a given query q, we compute the topical distribution of a result page by computing its distance to all pages in $\mathcal{S}_q$. Let us denote the probability (distance normalized by the sum) of a document d representing a particular topic $s \in S_q$ by $p_q(x,s)$. We will use this idea of coverage of a give topic to use the Wikipedia data set for evaluating the effectiveness of the diversification algorithm.

Recall from the framework that we view diversification as a re-ranking process for the search results, and we use the search results for baseline comparison here. Thus, given an ambiguous query $q \in \mathcal{Q}$, we first retrieve its top n results $R(q)$ using a commercial search engine. Then we run our diversification algorithm to choose the top k diversified results (the ordered list of diversified results is denote by $D(q)$) from the set of n results and compare the set of top k results, $D(q)$ and $R_k(q)$, in terms of relevance and novelty. The details of the performance measurement for each of the two measures are described next.

5.1.1 Novelty Evaluation

The idea behind the evaluation of novelty for a list is to compute the number of categories represented in the list

[4] `http://en.wikipedia.org/wiki/Disambiguation_page`

[5] `http://en.wikipedia.org/wiki/Jaguar_(disambiguation)`

[6] `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`

L of top-k results for a given query q which we denote by $\text{Novelty}_q(L)$. We note that this measure is same as the S-recall measure proposed in [21]. The list of topics for the query q is given by the set S_q of Wikipedia disambiguation pages for q.

We compute the probability distribution over S_q for all documents in the list L. To compute the representation or coverage of a topic $s \in S_q$ in the list L of search results, we aggregate the confidence on the topic over all the results in L, i.e. $\sum_{x \in L} p_q(x, s)$. If this sum is above a threshold $\theta \in \mathbf{R}$ we conclude that the topic s is covered by the S. The fraction of covered topics gives us a measure of novelty of the set S:

$$\text{Novelty}_q(L) = \frac{1}{|S_q|} \sum_{s \in S_q} \mathbf{I}\left(\sum_{x \in L} p_q(x, s) > \theta \right)$$

where $\mathbf{I}(\cdot)$ is the indicator function for an expression, evaluating to 1 if the expression is true, and 0 otherwise. Since we are only interested in the difference in this value between the two lists $D(q)$ and $R_k(q)$, we define fractional novelty:

$$\text{FN}_q = \frac{\text{Novelty}_q(D(q)) - \text{Novelty}_q(R_k(q))}{\max\left(\text{Novelty}_q(D(q)), \text{Novelty}_q(R_k(q))\right)}$$

We note that the value FN_q could also be negative, though we would expect it to be positive for a diversified set of results.

5.1.2 Relevance Evaluation

The relevance of a given document is often measured by the likelihood of the document satisfying the information need of the user expressed in terms of the search query. Therefore, it could be viewed as a measure of how close a document is to the query in the high-dimensional embedding. To compute the overall effectiveness of an ordering S, we compute its relevance based on its relative distance to the ideal ordering $S^{'}$ as

$$R(S, q) = \sum_{s \in S^{'}} \left| \frac{1}{r_s} - \frac{1}{r_s^{'}} \right| \quad (9)$$

where r_s is the rank of the document s in S and $r_s^{'}$ is the rank of s in $S^{'}$[7]

The aim of this evaluation is to compare the relevance of the diversified list $D(q)$ with the original list $R_k(q)$. To be able to compare these lists, we need to know the relative importance of each topic $s \in \mathcal{S}_q$. For the list $R_k(q)$ (or L), we achieve this by using the search engine to perform a site restricted search using only Wikipedia sites. From this search, we produce a relevance ordering $\mathcal{O}$ on S_q by noting the position of each $s \in S_q$. In the case of the diversified list $D(q)$, we compute the relevance of a topic $s \in \mathcal{S}_q$ as

$$Rel(s, q) = \sum_{d \in D(q)} \frac{1}{pos(d)} p_q(d, s),$$

where $pos(d)$ is the 1-based rank of d in the list $D(q)$. We compute a relevance ordering $\mathcal{O}^{'}$ for $D(q)$ and use the function in Equation 9 to compute the relevance distance between the two lists. The value computed from applying Equation 9 to these lists is the relevance score $\text{Relevance}_q(L)$. As before, we can compute the fractional difference in the relevance score:

$$\text{FR}_q = \frac{\text{Relevance}_q(D(q)) - \text{Relevance}_q(R_k(q))}{\max\left(\text{Relevance}_q(D(q)), \text{Relevance}_q(R_k(q))\right)}$$

[7]Note that since we formulate the result diversification as a re-ranking problem, we assume that both S and $S^{'}$ contain the same documents albeit in a different order. One simple characterization of relevance could set the rank of each document to its position in the ordered set of results.

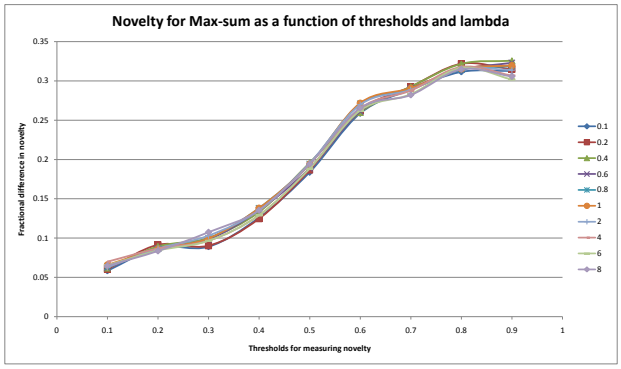


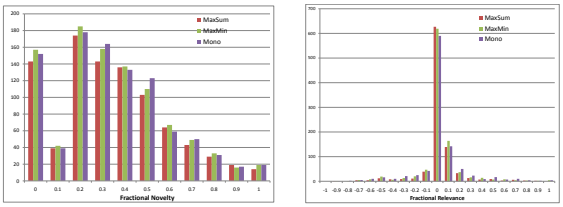
Figure 1: [Best viewed in color] The effect of varying the value of the trade-off parameter λ, and the threshold for measuring novelty on the output of the search results from MaxSumDispersion.

5.1.3 Results

The parameters used for the experiments in this work as: $n = 30$, $k = 10$. We choose $n = 30$ as relevance decreases rapidly beyond these results. The other parameters for the experiments are λ and θ, and the effect of these parameters is shown in Figure 1. Each point in this plot represents the average fractional difference in novelty given a value of λ and θ. The average is computed over a 100 queries drawn randomly from the set of the Wikipedia disambiguation pages. First of all, note that the fractional difference is always positive indicating that the diversification algorithm does succeed in increasing the novelty of the search results. Further, observe that increasing the confidence threshold θ has the effect of increasing the fractional difference between the search results. This indicates that the diversified search results have a higher confidence on the coverage of each category, and consequently the upward trend is a further vindication of increase in novelty of the diversified search results. Recall that the λ value governs the trade-off between relevance and diversity, and hence one would expect the novelty to increase with λ. This trend is observed in the plot, although the increase is marginal above a certain value of λ when the trade-off is already heavily biased towards novelty.

Figure 2(a) plots the histogram of the fractional difference in novelty as obtained over a 1000 queries drawn randomly from the set of Wikipedia disambiguation pages. It is worth noting that from the definition of fractional novelty that a fractional difference value of 0.1, with 10 covered categories, implies that the diversified search results covered one more category than the vanilla search results. Hence, on an average, the diversified search results cover as many as 4 more categories out of every 10 as compared to the original set of search results. In fact, we can say about 75% of the queries

produced more diverse results than the search results[8]. We note that on the overall, MAXMINDISPERSION outperforms the other two objectives.



(a) Novelty (b) Relevance

Figure 2: [Best viewed in color] The histogram of fractional difference in novelty (a) and relevance (b) plotted over a 1000 ambiguous queries.

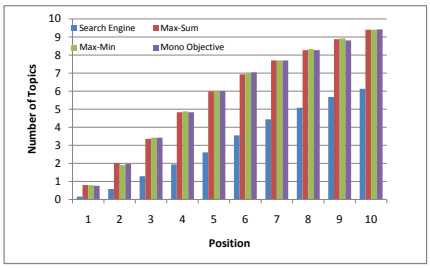
Figure 2(b) plots the histogram of the fractional difference in relevance as obtained over the same sample of 1000 queries. We note that the diversified set of search results does as well as the search engine ranking in the majority of the queries. Thus, the diversification algorithm does not suffer much in terms of relevance, while gaining a lot in novelty. Within the three algorithms, MONOOBJECTIVE does quite well on relevance on some queries, which is expected due to the importance given to relevance in the objective.

In another set of experiments, we study the positional variation of both relevance and novelty for all the three objective functions, i.e. we study the relevance and diversity values by restricting the output ranking to top m ranks, where $m = \{1, 2, \cdots, k\}$. We set $\lambda = 1.0$ and $\theta = 0.5$ for this experiment. Figure 3(a) plots the diversity at each position for the three objective functions and the search engine results. We note that the number of topics covered increases with position, and the results $D(q)$ produced by all three algorithms cover a larger number of categories even at the high ranked positions (low m) compared to the search engine. Although the difference in novelty between the search engine results and the diverse results is noticeable, the three diversification objectives perform quite comparably to each other. A similar plot for relevance is shown in Figure 3(b), which plots the positional variation in the value of relevance. We observe that all the orderings are equally relevant in the higher positions (lower m) while differences appear in lower positions. We note on the overall, the MONOOBJECTIVE formulation does the best in terms of relevance as it has the smallest distance to the ground truth. The MAXSUMDISPERSION algorithm comparatively produces the least relevant result sets among the three diversification objectives as it is more likely to add a more diverse result with respect to S than the other algorithms.

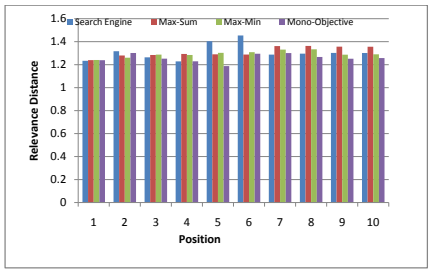
5.2 Product Disambiguation

In this set of experiments, we evaluate the performance of the three objective functions using different notions of distance and relevance. Specifically, we use the distance function described in Section 4.2 and the relevance score based on the popularity of a product. We note that the categorical distance can be tuned using the parameter e to

[8]The spike at 0 is due to queries where the search engine results had no overlap with the Wikipedia topics.



(a) Novelty



(b) Relevance

Figure 3: [Best viewed in color] The positional variation in the novelty (a) and relevance (b) of the result set.

effectively distinguish between various features of a product. For example, we would two different CD players produced by the same company to be closer to each other than to a CD player from a different company in a product taxonomy. Our data set (obtained from a commercial product search engine) consists of a set of 100 product queries and the top 50 results ranked according to popularity.

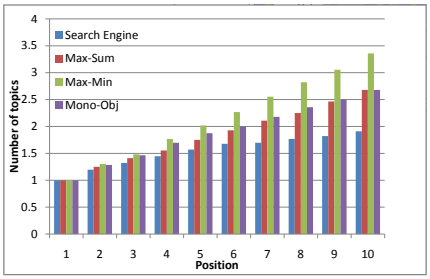
There is one drawback when we rank products based on their popularity. In the case where the popularity is skewed toward one or a few brands, the results can be dominated by products belonging to that brand with (sometimes) slight variations in the product descriptions. To observe the effectiveness of our formulation and the distance function, we diversified the results for 100 queries using the MAXSUMDISPERSION algorithm and compared the ordering with the results produced by a commercial product search engine. The parameters in this experiment were as follows: $n = 30$, $k = 10$, and $\lambda = 1.0$. Table 1 illustrates the difference between the orderings for the query `cd player`. Even though the search engine results for `cd player` offers some dominant brands, it does not include the other popular brands that the diversified results capture. Note that we do not alter the relative positions of the popular brands in the diversified results.

Similar to novelty evaluation for semantic disambiguation, we compute the number of categories represented in the list L of top-k results for a given query q which we denote by $\text{Novelty}_q(L)$. To compute the representation of a category in the result set, we require that the category is not the descendant of any other category in the result set. The fraction of covered topics gives us a measure of novelty of the set S:

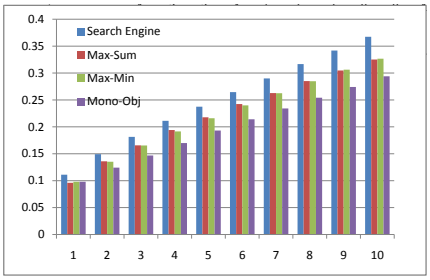
$$\text{Novelty}_q(L) = \frac{2}{|L|(|L|-1)} \sum_{u,v \in L} \mathbf{I}\left(lca(u,v) \notin \{u,v\}\right)$$

Product Search Engine	Diversified Results
Sony SCD CE595-SACD changer	Sony SCD CE595-SACD changer
Sony CDP CE375-CD changer	Sony CDP CE375-CD changer
Sony CDP CX355-CD changer	Teac SR-L50-CD player/radio
Teac SR-L50- CD player/radio	Bose Wave Music System Multi-CD Changer
Bose Wave Music System Multi-CD Changer	Sony S2 Sports ATRAC3/MP3 CD Walkman D-NS505
Sony RCD-W500C-CD changer/CD recorder	COBY CX CD109-CD player
Sony CD Walkman D-EJ011-CD player	JVC XL PG3-CD player
Sony S2 Sports ATRAC3/MP3 CD Walkman D-NS505	Pioneer PD M426-CD changer
Sony Atrac3/MP3 CD Walkman D-NF430	Sony SCD XA9000ES-SACD player
COBY CX CD109-CD player	Yamaha CDR HD1500-CD recorder/HDD recorder

Table 1: The difference in the top 10 results for the query `cd player` from a commercial product search engine and the diversified results produced by running MaxSumDispersion with $n = 30$ and $k = 10$



(a) Diversity



(b) Relevance

Figure 4: [Best viewed in color] The positional variation in the diversity (a) and relevance (b) of the product result set averaged over 100 queries with $n = 30$, $k = 10$, $\lambda = 1.0$ and $\theta = 0.5$.

where $\mathbf{I}(\cdot)$ is the indicator function and $lca(\cdot,\cdot)$ returns the least common ancestor of the two given nodes. Figure 4(a) illustrates the positional variation in the novelty of both the product search results ranked by popularity and the diverse set of results produced by MaxMinDispersion. For the relevance evaluation, we did not have any ground truth to compare the different orderings. Instead, we posit that a result is relevant to the query depending on how related the result and the query categories are in the taxonomy. We consider two categories to be completely related if one subsumes the other. Given a query q, we compute the relevance of a list L as

$$\frac{1}{\text{Relevance}_q(L)} = \frac{1}{|L|}\sum_{u\in L}\frac{1+d(lca(q,u),q)}{pos(u)},$$

where $pos(\cdot)$ is rank of the result u and $d(\cdot,\cdot)$ is computed using Equation 7. In abuse of notation, we use q and u to represent the nodes in the taxonomy as well. Figure 4(b) shows the positional variation in relevance for the product search engine and the diverse results produced by MaxMinDispersion. Surprisingly, the relevance of MonoObjective decreases relative to the other objectives.

6. CONCLUSIONS

This work presents an approach to characterizing diversification systems using a set of natural axioms and an empirical analysis that qualitatively compares the choice of axioms, relevance and distance functions using the well-known measures of novelty and relevance. The choice of axioms presents a clean way of characterizing objectives independent of the algorithms used for the objective, and the specific forms of the distance and relevance functions. Specifically, we illustrate the use of the axiomatic framework by studying three objectives satisfying different subsets of axioms. The empirical analysis on the other hand, while being dependent on these parameters, has the advantage of being able to quantify the trade-offs between novelty and relevance in the diversification objective. In this regard, we explore two applications of web search and product search, each with different notions of relevance and distance. In each application, we compare the performance of the three objectives by measuring the trade-off in novelty and relevance.

There are several open questions that present themselves in light of these results. In terms of the axiomatic framework, it would be interesting to determine if the impossibility proof still holds if the distance function is a metric. Relaxations of the axioms (for instance, weakening the stability axiom) are also an avenue for future research. Another direction of future research could be to explore the facility dispersion link, and identify optimal objective functions for settings of interest (such as web search, product search etc).

Acknowledgments

We would like to thank Kunal Talwar for helpful discussions, and Panayiotis Tsaparas for providing data and helping us with the classification of product queries.

7. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *Proc. 2nd ACM Intl Conf on Web Search and Data Mining*, 2009.

[2] A. Altman and M. Tennenholtz. On the axiomatic foundations of ranking systems. In *Proc. 19th International Joint Conference on Artificial Intelligence*, pages 917–922, 2005.

[3] Kenneth Arrow. *Social Choice and Individual Values*. Wiley, New York, 1951.

[4] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *STOC*, pages 161–168, 1998.

[5] Andrei Z. Broder, Moses Charikar, Alan M. Frieze, and Michael Mitzenmacher. Min-wise independent permutations. *Journal of Computer and System Sciences*, 60(3):630–659, 2000.

[6] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336, 1998.

[7] Barun Chandra and Magnús M. Halldórsson. Approximation algorithms for dispersion problems. *J. Algorithms*, 38(2):438–465, 2001.

[8] H. Chen and D.R. Karger. Less is more: probabilistic models for retrieving fewer relevant documents. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 429–436, 2006.

[9] 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. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 659–666, 2008.

[10] Sreenivas Gollapudi and Rina Panigrahy. Exploiting asymmetry in hierarchical topic extraction. In *CIKM*, pages 475–482, 2006.

[11] R. Hassin, S. Rubinstein, and A. Tamir. Approximation algorithms for maximum dispersion. *Operations Research Letters*, 21(3):133–137, 1997.

[12] J. Kleinberg. An Impossibility Theorem for Clustering. *Advances in Neural Information Processing Systems 15: Proceedings of the 2002 Conference*, 2003.

[13] B. Korte and D. Hausmann. An Analysis of the Greedy Heuristic for Independence Systems. *Algorithmic Aspects of Combinatorics*, 2:65–74, 1978.

[14] SS Ravi, D.J. Rosenkrantz, and G.K. Tayi. Facility dispersion problems: Heuristics and special cases. *Proc. 2nd Workshop on Algorithms and Data Structures (WADS)*, pages 355–366, 1991.

[15] S.S. Ravi, D.J. Rosenkrantz, and G.K. Tayi. Heuristic and special case algorithms for dispersion problems. *Operations Research*, 42(2):299–310, 1994.

[16] SS Ravi, D.J. Rosenkrantzt, and G.K. Tayi. Approximation Algorithms for Facility Dispersion. In Teofilo F. Gonzalez, editor, *Handbook of Approximation Algorithms and Metaheuristics*. Chapman & Hall/CRC, 2007.

[17] Stephen Robertson and Hugo Zaragoza. On rank-based effectiveness measures and optimization. *Inf. Retr.*, 10(3):321–339, 2007.

[18] Atish Das Sarma, Sreenivas Gollapudi, and Samuel Ieong. Bypass rates: reducing query abandonment using negative inferences. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 177–185, New York, NY, USA, 2008. ACM.

[19] E. Vee, U. Srivastava, J. Shanmugasundaram, P. Bhat, and S.A. Yahia. Efficient Computation of Diverse Query Results. *IEEE 24th International Conference on Data Engineering, 2008. ICDE 2008*, pages 228–236, 2008.

[20] ChengXiang Zhai. *Risk Minimization and Language Modeling in Information Retrieval*. PhD thesis, Carnegie Mellon University, 2002.

[21] C.X. Zhai, W.W. Cohen, and J. Lafferty. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 10–17, 2003.

[22] C.X. Zhai and J. Lafferty. A risk minimization framework for information retrieval. *Information Processing and Management*, 42(1):31–55, 2006.

[23] C.N. Ziegler, S.M. McNee, J.A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. *Proceedings of the 14th international conference on World Wide Web*, pages 22–32, 2005.

8. APPENDIX

8.1 Impossibility Result

Proof of Theorem 1. We start by fixing functions $w(\cdot)$ and $d(\cdot,\cdot)$ such that f is maximized by a unique S_k^* for all $k \geq 2$. Such a set of functions always exist, from the richness axiom. Now, fixing a k, we can use the uniqueness property and the stability axiom, to say that $\forall y \notin S_{k+1}^*$, we have:

$$f(S_{k+1}^*) > f(S_k^* \cup \{y\})$$

Note that we have a strict inequality here, as otherwise the monotonicity axiom would imply that there is no unique S_{k+1}^*, as we have that $f(S_k^* \cup \{y\}) \geq f(S_k^*)$ for all $y \notin S_k^*$. From here on, we fix one such y.

Let $x = S_{k+1}^* \setminus S_k^*$ (which follows from stability). Now, we invoke the strength of relevance (b) axiom for S_k^*, in conjunction with the independence of irrelevant attributes axiom, to imply that the function value $f(S_{k+1}^*)$ of the set S_{k+1}^* decreases by some $\delta > 0$ if $w(u)$ is fixed for all $u \in S$ and $w(x) = a_0$ for some $a_0 > 0$. Next, we do the following transformation: 1) increase $d(y,u)$ for all $u \in S_k^*$ to be equal to twice their current value, and 2) scale all the relevance and distance values by half.

Firstly, note that the first step of this transformation is consistent (in the sense defined in the consistency axiom) w.r.t both S_k^* and S_{k+1}^*. Also, the second step uses global scaling, and hence from the scale invariance and consistency axioms, we have that the output of f for size k and $k+1$ should still be S_k^* and S_{k+1}^* respectively. Further, from the above comment about the decrease in $f(S_{k+1}^*)$, we have that $f(S_{k+1}^*) - f(S_k^*)$ strictly decreases. In addition, we can use the strength of diversity (a) axiom to see that $f(S_k^* \cup \{y\})$ strictly increases. Repeating this process several times, we are guaranteed to get a state where

$$f(S_{k+1}^*) \leq f(S_k^* \cup \{y\})$$

which implies that S_{k+1}^* is not unique, and is hence a contradiction to the richness axiom. □