Balancing the Trade-Offs Between Diversity and Precision for Web Image Search Using Concept-Based Query Expansion

Enamul Hoque, Orland Hoeber, and Minglun Gong
Department of Computer Science
Memorial University of Newfoundland
St. John's, NL, Canada A1B 3X5
Email: {enamulp, hoeber, gong}@mun.ca

***Abstract*— Even though Web image search queries are often ambiguous, traditional search engines retrieve and present results solely based on relevance ranking, where only the most common and popular interpretations of the query are considered. Rather than assuming that all users are interested in the most common meaning of the query, a more sensible approach may be to produce a diversified set of images that cover the various aspects of the query, under the expectation that at least one of these interpretations will match the searcher's needs. However, such a promotion of diversity in the search results has the side-effect of decreasing the precision of the most common sense. In this paper, we evaluate these competing factors in the context of a method for explicitly diversifying image search results via concept-based query expansion using Wikipedia. Experiments with controlling the degree of diversification illustrate this trade-off between diversity and precision for both ambiguous and more specific queries. As a result of these experiments, an automatic method for tuning the diversification parameter is proposed based on the degree of ambiguity of the original query.**

I. Introduction

The primary method for performing image retrieval on the Web is based on document search techniques [1]. Images are indexed based on the text that is related to their use on the Web (keywords, tags, and/or associated descriptions). User-supplied queries are matched to this text to produce a set of images, which are ranked based on the relevance of their associated textual information to the query. This approach can work well when the contents of the images are concisely and accurately described within their source Web pages, and when the searchers' needs are clearly specified.

Recent studies on user behaviour with respect to image search have found that queries are often very short and ambiguous [2]. This ambiguity comes from the difficulty that searchers experience in finding the words to describe an idea or image they have in their mind. From the perspective of image retrieval, the difficulty with ambiguous queries is that they can be open to many different interpretations. It is possible that different searchers may enter the same query, but their intentions and needs may vary significantly from one another. In situations such as this, the matching algorithms used by image search engines promote the interpretation that is most common and popular. However, a more sensible approach may be to produce a diversified set of images that cover the various aspects of the query, under the expectation that at least one of the interpretations matches the searcher's intent. Providing searchers with an overview of the images and allowing them to zoom-in and focus on a particular interpretation [3] may improve the effectiveness and efficiency of the image search process.

Moving beyond traditional relevance ranking, diversification approaches aim to improve the coverage of the search results set with respect to the different senses of the original query. A common method for diversification is query expansion, whereby additional terms are added to the query to generate a collection of new queries that, when taken together, are more broad than the original [4]. However, in doing so there is a danger in broadening the query too much, resulting in a potentially significant decrease in precision. That is, the more broad and diverse the search results are, the less chance that a particular search result will be relevant to the searcher's information need. As such, this trade-off between diversity and precision must be studied in order to understand the situations where more or less diversification is beneficial.

Maintaining a balance between diversity and precision requires an automatic modelling of the searcher's query to determine an appropriate degree of diversification to promote. In many cases, image search queries are inherently ambiguous. For example, "Washington" might be interpreted as "Washington (state)", "Washington D.C.", or "George Washington". Even within a more specific query, searchers might have an interest in seeing images that are related to but not explicitly identified in the query. For example, if a searcher submits a query such as "Hong Kong", they may wish to see some representative images of different landmarks in the Hong Kong area. In other cases, a query might be very specific, and the scope for

This paper is an extended version of "Evaluating the Trade-Offs Between Diversity and Precision for Web Image Search Using Concept-Based Query Expansion," by E. Hoque, O. Hoeber, and M. Gong, which appeared in the Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence Workshops (International Workshop on Intelligent Web Interaction), Lyon, France, August 2011.

doi:10.4304/jetwi.4.1.26-34

broadening the query is limited. For example, queries for a particular landmark within a specific setting like "Eiffel Tower Bastille Day" may leave little room for diversifying the search results.

In previous research, we presented a method for diversifying image search results using concept-based query expansion, which is based on information derived from Wikipedia [3]. In this work, we have modified our approach such that the degree of diversification can be controlled, allowing us to evaluate the trade-offs between diversification and precision among the image search results. Using this knowledge, we have developed a simple method for automatically determining the degree of diversification based on the level of ambiguity of the query, resulting in a balance between diversification and precision.

The remainder of this paper is organized as follows. Section II provides an overview of the existing methods for search results diversification. Section III outlines our method for diversifying image search results, with a specific focus on a parameter that we use to control the degree of diversification. Section IV describes an evaluation that explores how the diversification parameter can be used to find a balance between diversity and precision within the search results, and proposes a formula for automatically determining this parameter. Section V provides a discussion on the outcomes of this study and the implications for diversifying image search results. The paper concludes with a summary of the research contributions in Section VI.

II. Related Work

The problem of enhancing diversity in search results has been recognized as an important topic in the research community. Diversification can be stated as an optimization problem, where the objective is to maximize the probability of finding a relevant document/image among the top selected results, while minimizing its redundancy with respect to different aspects of the query. The general problem is an instance of maximum coverage problem, which is NP-hard [5].

In the literature of document search techniques, most of the previous approaches to search results diversification can be categorized as either implicit or explicit [6]. Implicit search results diversification approaches perform direct comparison between the retrieved documents, under the assumption that similar documents will cover similar aspects. In an early work, Carbonell et al. [7] proposed a greedy approximation based ranking method called Maximal Marginal Relevance (MMR), which attempted to maximize the relevance of a search result while minimizing its similarity to higher ranked documents. In another implementation of an MMR-based method, Zhai et al. [8] proposed a subtopic search method using statistical language models, which aimed to return results that cover more subtopics.

An alternative approach to diversification is to explicitly utilize different aspects associated with a query by directly modelling these aspects. For instance, Agrawal et al. [5] investigated the diversification problem based on the assumption that a taxonomy is available for both queries and documents. In their work, documents retrieved for a query were considered similar if they were confidently classified into one or more common categories covered by the query. They applied a greedy algorithm that started with an empty list of results and selected the next result with the highest marginal utility until k results were selected. Here, the marginal utility measured the probability that the result satisfied a category the current result set did not yet cover. In a related approach, Radlinski and Dumais [9] proposed to filter the results retrieved for a given query so as to limit the number of those satisfying the same aspect of the query, represented as different query reformulations obtained from a large query log from a commercial search engine. Recently, Santos et al. [10] introduced a probabilistic framework for search results diversification, which explicitly represented query aspects as "sub-queries". They defined diversity based on the estimated relevance of documents to multiple sub-queries and the relative importance of each sub-query.

The diversity problem is more challenging in Web image search for a number of reasons. First of all, image search may involve not only semantic diversity but also visual diversity of the search results [11]–[14]. Although there may be some benefit to promoting both types of diversity, the focus of our research is on the semantic side. Secondly, in the document centric retrieval methods, the documents themselves provide some useful ways to allow direct comparisons and similarity calculations when promoting diversification in the search results. But in image retrieval, the limited amount of textual metadata associated with the retrieved images are not reliable enough nor sufficient to allow for the computation of similarities between their associated images.

A somewhat common approach to dealing with the diversification problem for image search results is to use semantic and/or visual clustering [11]–[13], [15], [16]. Although these approaches differ in the information used and the methods employed for clustering, the end result is a grouping of images that represent the different aspects of the image collection. There are, however, a number of challenges associated with clustering approaches, including determining a suitable similarity measure, the efficiency of clustering algorithms, determining an appropriate number of clusters to use, and deciding how to order or organize the clusters.

Re-ranking methods have also been proposed with an aim to enhance topic coverage in search results. Song et al. [17] addressed the diversity problem using a re-ranking method based on topic richness analysis. The goal here is to enhance topic coverage in the image search results, while maintaining acceptable retrieval performance. A topic richness score was computed by analyzing the degree of mutual topic coverage between an image pair based on the assumption that images are annotated by several words.

Some have also begun to study the trade-off between precision and diversification within the context of image search. For example, van Zwol et al. proposed a method for optimizing this trade-off by estimating a query model from the distribution of tags that favour the dominant sense of an image search query [18].

A promising direction for enhancing the diversity of image search results is the use of query expansion methods based on concepts that are related to the query. Unlike the general domain of document-centric information retrieval, query expansion has been studied in only a few research works in Web image retrieval. Those that have explored such techniques have shown them to be promising [3], [19]. However, one of the challenges associated with query expansion is to find an appropriate source of knowledge required for the expansion process. It has been noted that many image search queries are associated with conceptual domains that include proper nouns (i.e., people's names and locations) [2], [20]. As such, finding a suitable knowledge base that has sufficient coverage of a realistic conceptual domain is very important first step in this approach to query expansion. Wikipedia is a good candidate for such a knowledge base since it includes a large number of articles describing people, places, landmarks, animals, and plants. The challenge in using Wikipedia is to design efficient and effective algorithms that can process the semi-structured knowledge to derive meaningful terms for use in the query expansion process.

III. IMAGE SEARCH RESULTS DIVERSIFICATION

For the diversification framework used in this research, image search results are explicitly diversified based on concept-based query expansion. For the short and ambiguous queries that are common in image search, query expansion attempts to capture the various aspects of the query. We model the original query by discovering different possible senses, and for each sense a number of concepts pertaining to the query are discovered from within Wikipedia. These concepts are ranked according to the semantic relatedness to the original query, and only the top-N most related concepts are used within the query expansion process to retrieve a range of images that provides a broad view of what is available.

Within this process, the value of N is an explicit indicator of the degree of diversification, which we call the diversification parameter. With a smaller value of N, fewer concepts will be used, and the search results will remain more focused. If we increase N, then more concepts will be used for query expansion, and in turn the search results will be more diversified covering more aspects associated with the query. The fundamental trade-off between diversification and precision is based on the fact that as we increase N, there is a higher chance that a concept will be selected for the query refinement process that is not relevant to the searcher's information needs, resulting in the associated irrelevant images being included in the search results. As such, our goal is to fulfill the diversification objective by setting N sufficiently high to capture the broad range of concepts associated with the query, but not so high as to have a significant adverse affect on the average precision across all of the senses of the query.

In this precision-diversity trade-off context, we can state our diversification objective as follows: *Given a query Q, perform query expansion based on N concepts to retrieve a results set R, so that it will maximize the value of N, and at the same time maximize the precision P with respect to each of the possible senses of the query over the results set R.* Note that an inherent feature of such diversification is that it will increase the precision of the search results with respect to the less common senses of the query, at the expense of the most common sense.

To achieve this objective in our diversification method, the possible senses of the query, the number of concepts to be selected for each sense, and the number of images to be retrieved for each concept are determined automatically based on an analysis of the original query and the candidate concepts extracted from Wikipedia. The purpose here is to distribute the concepts and the number of images retrieved for each concept unevenly, so that we can alleviate the problem of harming precision due to the images retrieved for irrelevant concepts that are only loosely related to the senses of the original query.

The process of diversifying the image search results using concept-based query expansion follows three steps: extracting concepts from Wikipedia; ranking the extracted concepts to select the top-N related concepts; and retrieving images based on the expanded queries derived from these concepts. The details for each of these steps are explained in the remainder of this section.

A. Concept Extraction Using Wikipedia

In order to use Wikipedia as the source knowledge base in this work, a dump of the Wikipedia collection was obtained and preprocessed to support the type of knowledge extraction required for our purposes. Matching a user-supplied query Q to this knowledge base is simply a matter of selecting the best matching article (referred as the home article) using Wikipedia's search feature. In the case where the query is ambiguous and Wikipedia suggests multiple senses of interpretations, the ones with higher commonness values are used as the home articles. Here the commonness value of an article is calculated based on how often it is linked by other articles.

In analyzing Wikipedia, we observed that the in-link articles (ones having links to a home article) and out-link articles (ones to which a home article links) often provide meaningful information that is closely related to the concept of the home article, and hence the user-specified query. Therefore, these linked articles are located and their titles are extracted as candidates for related concepts.

We also found that the captions surrounding the images present within a given article can often provide a valuable perspective on the visual features associated with the article's concept. Due to the importance of this information for the purposes of image retrieval, it is important to ensure that all relevant concepts associated with the image captions are extracted. We use Wikifier [21] to augment the captions with links to relevant Wikipedia articles that may have been missed by the author of the article, and use these links to extract their associated concepts.

The end result of this process is the selection of a set of home articles $\{h_s | 1 \leq s \leq q\}$ (for q senses of given query Q), along with a list of all the candidate articles C_{h_s} for each home article h_s that originate from the in-link articles, out-link articles, and the image captions. These concepts provide the basis for the automatic query expansion process.

B. Ranking the Extracted Concepts

Due to the rich and interconnected nature of Wikipedia, the number of concepts obtained in the process described above may become very large. Thus, a filtering step is necessary to ensure the quality of the concepts that are extracted. Here, our objective is to select the top-N concepts from among all the candidate articles. Considering the difference in the importance of each of the senses, we distribute these top-N concepts among the candidate concepts C_{h_s} of each home article h_s. As such, the number of concepts N_{h_s} that are to be selected for a particular home article h_s is determined as follows:

$$N_{h_s} = \frac{|C_{h_s}| \times N}{\sum_{j=1}^{q} |C_{h_j}|}$$

Note that the sum of all N_{h_s} values equals N.

To select these N_{h_s} concepts for each home article h_s, it is necessary to rank the candidate concepts C_{h_s} based on their relevance to their associated home article. Our approach to this problem is to measure the semantic relatedness between the home article and each of the candidate concepts. A number of different methods have been devised to use Wikipedia for this purpose, including WikiRelate! [22], Explicit Semantic Analysis (ESA) [23], and Wikipedia Link-based Measure (WLM) [24]. Given the computationally efficiency and accuracy of WLM, we use this approach in our work.

For each of the candidate articles $c_i \in C_{h_s}$ extracted from the home article, WLM is applied between the home article h_s and the candidate articles. WLM takes advantage of the hyperlink structure of the associated articles to find out how much they share in common. In order to give preference to the concepts that have been extracted from the image captions within the home article, we use a re-weighting function to determine the relatedness score:

$$r(c_i, h_s) = \min(WLM(c_i, h_s)(1 + \alpha_s), 1)$$

Since WLM provides a value in the [0,1] range, we ensure that the relatedness score remains in this range with the min function. The re-weighting factor α_s is provided according to the following function:

$$\alpha_s = \begin{cases} k\frac{C_{h_s}}{N_{h_s}} & \text{if concept } c_i \text{ originates from a caption} \\ 0 & \text{otherwise} \end{cases}$$

Here, C_{h_s} and N_{h_s} are as defined above, and k is a system parameter that controls the importance of the concepts derived from the captions. In our prototype implementation $k = 0.01$, which results in a 10 - 20% increase in the score for the concepts derived from the captions, with proportionally more importance being given when there are more concepts extracted from the home article.

The outcome of this process is that the top-N concepts are selected from among the candidate articles, such that the ones that came from the image captions are given preference over those from in-link and out-link articles. These concepts are used as the source for the query expansion process. The value of N serves here as an explicit diversification parameter. How it affects the diversification and precision of search results are discuss in Section IV.

C. Concept-Based Query Expansion and Image Retrieval

In order to ensure that the expanded queries remain focused on the topic of the query itself, the top-N related concepts $\{c_r | 0 \leq r \leq N\}$ are prepended with their associated home article h_r, resulting in queries of the form $< h_r, c_r >$. We define c_0 to be null and h_0 to the the original query Q, producing the original query plus N expanded queries.

Given that individual expanded queries have differing degrees of relevance to the original query, we dynamically determine how many images to retrieve for each expanded query based on their relatedness score to the home articles. This way we can ensure that more images are retrieved for concepts that are most similar to the sense associated with their home article, even when the original query has multiple meanings. This is done to minimize the number of images retrieved for concepts that are only loosely associated with the sense of the query.

The number of images to retrieve for each expanded query is given by the following formula:

$$I_r = \frac{r(c_r, h_s) \times I_t}{\sum_{k=0}^{N} r(c_k, h_s)}$$

Here, r is the same function used to generate the relatedness score in the concept ranking process, and I_t is the total number of images to be retrieved by all of the queries. We set $I_t = 60$ for the purposes of performing the evaluation within this paper, but it can be set to any reasonable number of images. Since the null expanded query (c_0) is the original query, we define $r(c_0, h_s) = 1$ in the above calculation. All of the queries are sent to the Google AJAX Search, and the desired number of images are retrieved. Duplicate images are deleted based on the URL of the source image (as provided by the underlying search engine).

IV. Evaluation

The goal of this approach is to automatically diversify the images retrieved for a given query, using Wikipedia as the source for a query expansion process. However, it is unclear to what degree such diversification should be promoted during the search process. In this evaluation of the approach, our goal is to study the inherent trade-off between precision and diversity in detail. In particular, for a set of queries, we explore how the precision changes as diversity in the image search results is promoted. Using this information, we propose a simple approach to automatically determining the degree of diversification based on features of the user-supplied query.

A. *Experimental Setup*

For these experiments, we chose 12 query topics, split between those we deemed to be highly ambiguous (having four senses), moderately ambiguous (three senses), slightly ambiguous (two senses), and non-ambiguous (one sense). This distribution of different degrees of ambiguity allowed us to examine the effect of the experimental condition (i.e., the varying of the degree of diversification) in the context of ambiguity. For each of these different degrees of ambiguity we selected two queries, except for the moderate ambiguity, for which we selected six queries. The moderate degree of ambiguity was examined more carefully since it represents the most common case of ambiguity.

To evaluate the effect of diversification on precision, we retrieved the top 60 search results from Google Image Search using our concept-based query expansion method with ten different values of N, ranging from 0 to 40 ($N = 0, 2, 4, 6, 8, 10, 15, 20, 30, 40$). Here, $N = 0$ implies that no query expansion has occurred (i.e., the search results are not diversified, and are simply the results provided by the underlying image search engine). At the other extreme, $N = 40$ causes the system to return a highly diversified set of image search results from 40 different associated concepts chosen in the query expansion procedure. Data was collected more frequently in the low end of this range in order to more closely observe the effect of a low degree of diversification. Preliminary experiments illustrated that the effect of the degree of diversification at the higher range became rather stable [25].

For each of the different senses of the query, assessors were asked to judge the relevance of each image. This assessment of relevance provided the ground truth information in the calculation of the precision scores (the ratio of relevant images to the total number of images retrieved). Since there were ten trails (i.e., ten different values of N) and 60 images retrieve with each trial, this resulted in the evaluation of a total of 600 images for each test query.

B. *Results*

In these experiments we measured the precision for each of the test queries as the diversification parameter N was varied from 0 to 40. Our hypothesis was that as N increased, the distribution of the senses would become more balanced across all of the meanings of the query. This would result in a reduction in the precision for the most common senses of the query, and an increase in the precision for the less common senses. This feature can be readily identified in the graphs in Figures 1, 2, and 3. To further understand this effect, we plotted the average precision (the red lines with the triangle markers) and the total precision (the dark red lines with the x marker) across all of the senses.

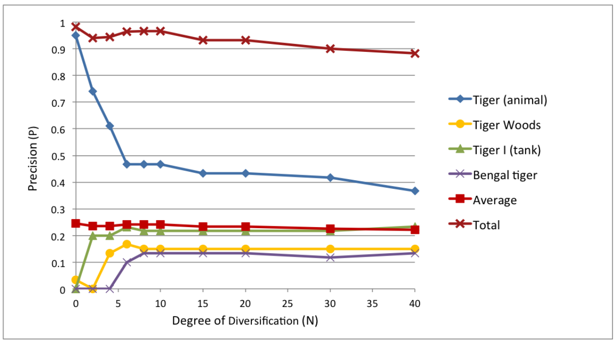
Figures 1 and 2 show the results from the highly and moderately ambiguous queries. The general effect that can be seen here is that the precision for the most common sense (i.e., the blue line in each graph) was automatically reduced as a result of the diversification. In most cases this occurred in a more or less smooth fashion even with very low values of N. At the same time, the precision for all other senses increased. In some cases, it was necessary for the value of N to be set higher than six for images from some of the less common senses of the query to be represented in the search results.

Although there was, in some cases, a minor reduction in the total precision for very low values of N, this was often accompanied by a subsequent increase in the total precision as more diversification occurred. This effect is a result of the method by which the number of images retrieved for each expanded query is dynamically determined. With very few expanded queries, more images are retrieved for each (which may result in the inclusion of some less relevant images deeper in the search results list). As N is increased and more expanded queries are generated, the images that are retrieved have higher rankings with respect to their source query.

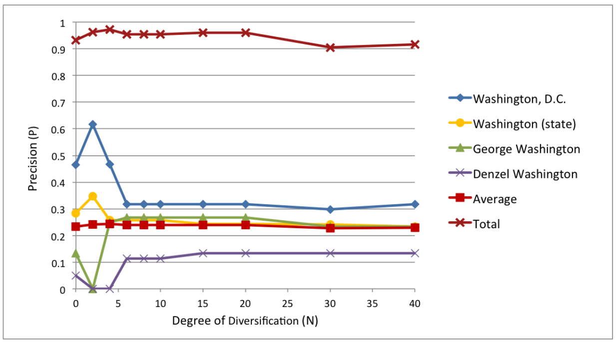
For all of these queries, the precision of the images over each of the senses of the query did not change significantly once the value of N was set beyond six to ten, depending on the specific query. Furthermore, if the value of N was set too large (i.e., 30 or 40), the average and total precision started to decrease, indicating that some non-relevant concepts and their associated images were being included in the search results set due to over-diversification.

Our expectation when designing these experiments was that for queries that have a high degree of ambiguity, it would be necessary to set the diversification parameter rather high in order to capture enough information on all of the different senses. However, it is clear that even with a diversification parameter set at $N = 10$, the desired effect appears.

In some cases, the most common sense remains the most common regardless of the level of diversification (e.g., Figure 1a and b, and 2d, and e). In the other cases, senses that were less common in the original search results become dominant. The reason for this change is that there may be disagreement between what the underlying search engine assumes is the desired information need (i.e., the most common sense of the query), and the amount of information that can be extracted from Wikipedia on the other senses and their associated concepts.

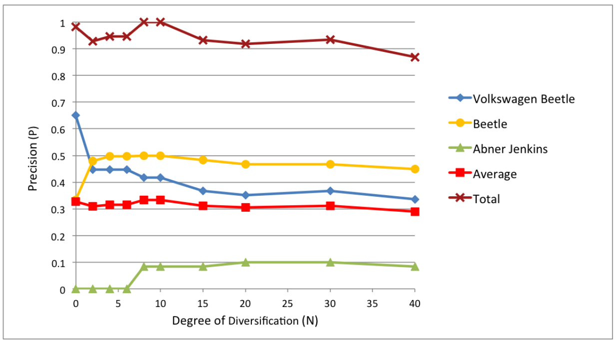


(a) Tiger

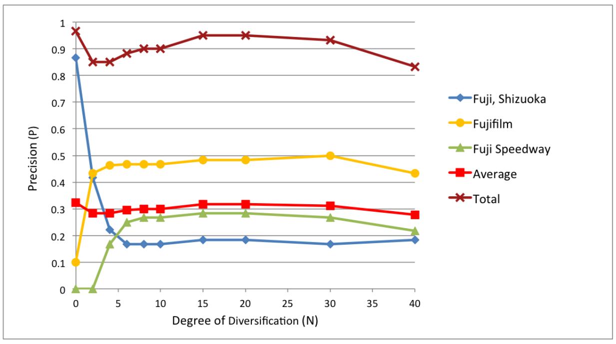


(b) Washington

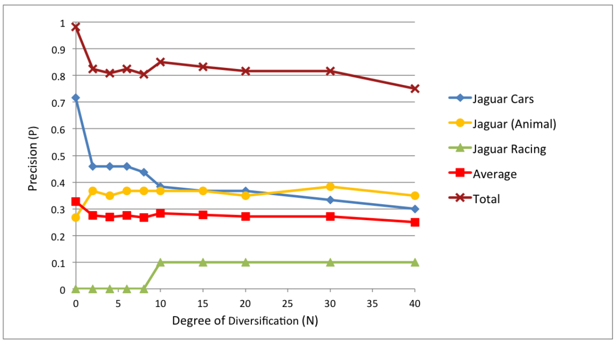
Figure 1. The effect of varying the degree of diversification (N) on precision (P) for highly ambiguous queries with *four* different senses.



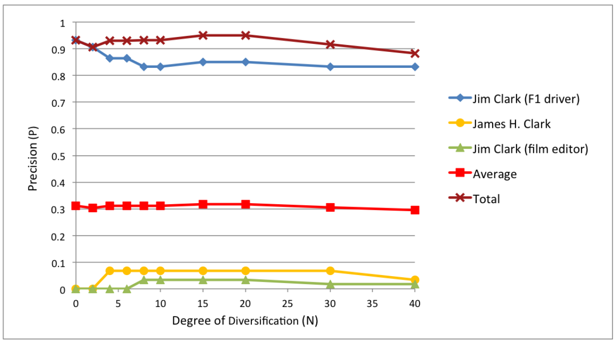
(a) Beetle



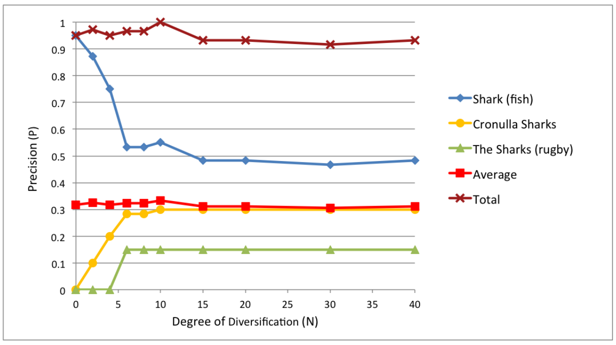
(b) Fuji



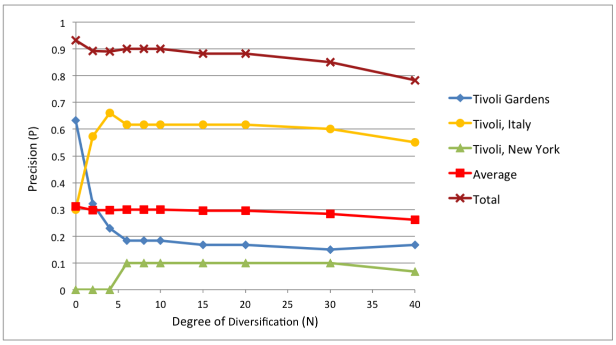
(c) Jaguar



(d) Jim Clark



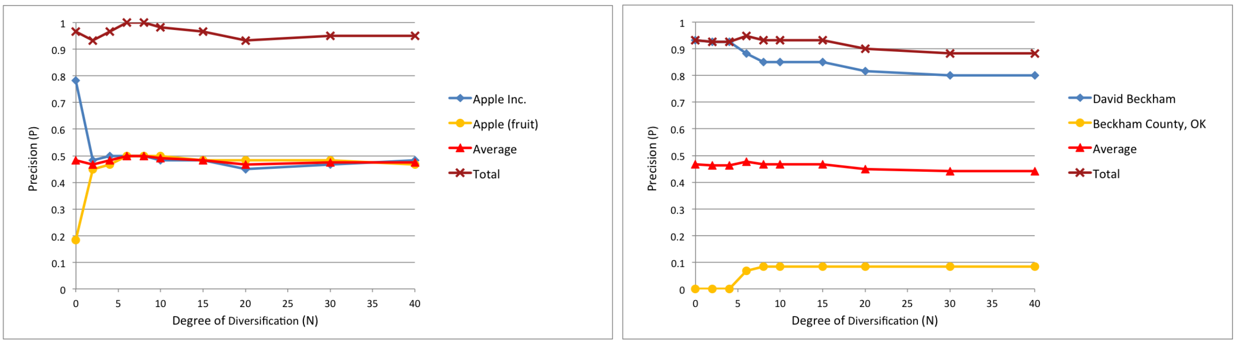
(e) Sharks



(f) Tivoli

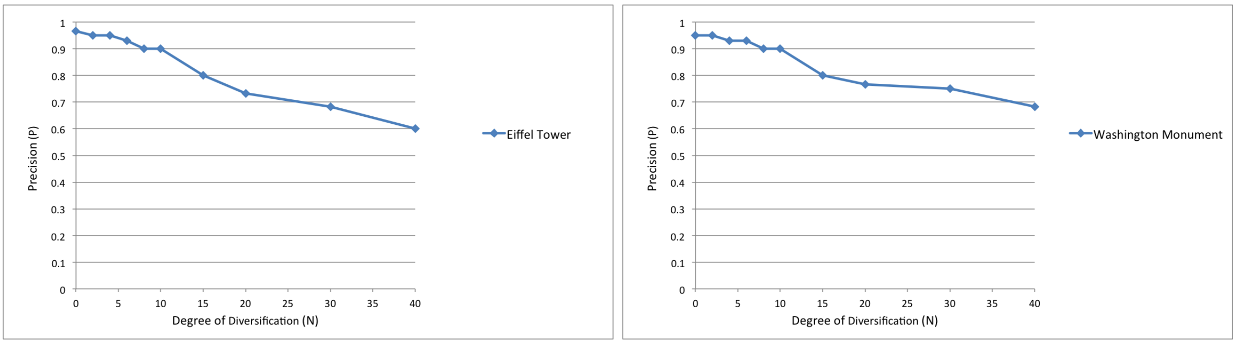
Figure 2. The effect of varying the degree of diversification (N) on precision (P) for moderately ambiguous queries with *three* different senses.

Figure 3 shows the results from the slightly ambiguous queries. As with the highly ambiguous queries, even with a low degree of diversification (from two to six), the outcome of the diversification is a balancing of the precision between the two senses of the query. In addition, the effect of a dropping then increasing total precision over low values of N is present, as is the dropping average and total precision when N is set too large. However, the later effect occurs with lower values of N than with the highly or moderately ambiguous queries (15 - 20 for the



(a) Apple (b) Beckham

Figure 3. The effect of varying the degree of diversification (N) on precision (P) for slightly ambiguous queries with *two* different senses.



(a) Eiffel Tower (b) Washington Monument

Figure 4. The effect of varying the degree of diversification (N) on precision (P) for non-ambiguous queries with only *one* sense.

slightly ambiguous queries).

For the queries where there was only one sense (Figure 4), it is clear that diversifying the search results can very quickly have a negative effect on the precision. However, for small values of N (e.g, 2-4) this effect is negligible. This indicates that even for very specific queries, a small degree of diversification might be tolerable and perhaps even beneficial as highly related images are drawn into the search results set.

As a result of this analysis, we conclude that diversifying the image search results can be very useful for addressing the situation where an ambiguous query has multiple senses. Rather than relying on the search engine to choose the most common sense, we can diversify the image search results and let the user focus on those images that match their needs. The more senses that can be inferred from a query, the more diversification is necessary to sufficiently balance all of these senses in the search results. However, when there are few different senses, the degree of diversification should be limited to avoid including irrelevant concepts and their associated images in the search results.

C. Automatically Determining the Degree of Diversification

One of our goals in this research is to automatically determine the degree of diversification needed to enhance the searcher's ability to find relevant documents. As illustrated in the experiments, the degree of ambiguity of the query has an impact on the value of diversification. That is, a highly ambiguous query can benefit from more diversification, whereas a very specific query may be harmed by diversifying the search results too much.

Based on this, we propose a linear scaling of the degree of diversification based on the number of senses of the query (as determined by Wikipedia). One such automatic function for determining the how aggressively to diversify the search results is $N = a \times q$ where q is the number of senses for query Q (following the terminology from Section III). Based on the experiments reported in this paper, setting $a = 3$ will produce reasonable results. That is, if a query has one sense, the diversification parameter will be set to 3, performing a relatively low degree of diversification that does not harm the precision of the one sense of the query. For the slightly, moderately, and highly ambiguous queries, this results in the diversification parameter being set to 6, 9, and 12, respectively. By inspecting the graphs in Figures 1 - 4, we can see that this is a near-optimal value for N.

This simple approach to determining N could benefit from further research in fine-tuning this formula and its parameters for automatically determining the value of N based on some measure of query ambiguity.

V. Discussion

Without diversification, the most common sense of an ambiguous query can dominate the image search results. Images that match other senses of the query may not be very common or even represented at all within the top search results. This is a preferential outcome if the searcher's information needs match the underlying search engine's interpretation of the query. However, when this is not the case, it can be very difficult for the searcher to find images that match their needs. Their only recourse is to attempt to reformulate the query into something more specific. However, studies on Web image search behaviour indicate that this may be a rather difficult task for searchers to perform.

The key benefit of performing image search diversification is that instead of assuming a single interpretation of the query, we retrieve images from different senses discovered using Wikipedia, providing a more balanced distribution of the senses within the set of search results. As a result, this diversified set of search results may be suitable to a wider range of users with a wider range of information needs.

A common problem with query expansion in general, and search result diversification in particular, is that it can reduce the precision by including documents (images) that might not be relevant to all searchers. This remains the case for our work on image search diversification. In particular, we have shown how the precision for the most common sense will invariably be reduced. This is because images from the less common senses are being included in the search results as a result of the diversification process. However, even amid this reduction in the precision of the most common sense, and the increase in the precision of the less common senses, the average precision is not chaining noticeably when the degree of diversification is not set unreasonably large. On the contrary, in some cases the total precision actually increases, when calculated over all senses of the query. This indicates that the proposed approach is effective in keeping the expanded queries focused on the intended senses of the query, even when there are multiple such senses.

Because the process of diversification will commonly result in a mix of relevant and irrelevant images for a given interpretation of an ambiguous query, it is important that an interface to the image search results be used that makes it easy for the searcher to ignore the senses of the query that are not relevant to their needs, and focus on those that are. We have developed such an interface in previous research that organizes the search results according to both visual and conceptual similarities [26], allowing the searcher to filter the search results both visually through a pan-and-zoom operation [27], and based on a hierarchy of the concepts that produced the expanded queries [3]. Using this visual interface to explore the search results, the precision for a particular sense of the query may be improved as irrelevant images are moved out of the field of view or are filtered based concepts that are uninteresting to the searcher. An evaluation of this interface from the perspective of improving the precision of the search results through these interactive operations has shown not only the benefit of the interactive features, but also the benefit of organizing the images based on both visual and conceptual features [28].

It is also worth noting that the benefit that this approach to image search diversification can provide depends greatly on the completeness and interconnected nature of information in Wikipedia. If there is a sense of a given query that is not well represented in Wikipedia, then it will not be well represented within the diversified image search results set. As such, as Wikipedia continues to grow and be enhanced, it will become a better and better tool for providing knowledge for information retrieval research such as what has been discussed in this paper.

VI. Conclusions

In this paper, we present a novel approach for explicitly diversifying image search results using concept-based query expansion. This approach is based on our previous research, but has been modified such that the degree of diversification can be controlled through the diversification parameter N. Using this system, we have evaluated the trade-off between diversification and precision, using a set of test queries of varying ambiguity.

From these experiments, we can see that the degree to which diversification needs to be promoted in order to provide a balance across all of the senses of a query depends on the ambiguity of the query itself. That is, a highly ambiguous query can benefit from a high degree of diversification more so than a very specific query. Based on this, we propose a simple automatic calculation of the diversification parameter based on the level of ambiguity of the query that balances the increased diversification against the decrease in precision.

Instead of only satisfying the information needs for the most common interpretation of the query, our method provides a more balanced view of the different senses of the query, without a significant negative impact on the average precision across all senses. Furthermore, our method of diversification provides an effective means for ensuring that the expanded queries that produce the diversified search results remain focused on at least one sense of the query. Coupled with a visual interface for organizing the image search results based on their visual and conceptual similarity, and interaction methods that support dynamic filtering, this approach to search results diversification can be a very powerful tool for enhancing the image search experience.

Acknowledgments

This research was supported by Research Development Corporation (RDC) IgniteR&D and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants held by the last two authors.

References

[1] M. L. Kherfi, D. Ziou, and A. Bernardi, "Image retrieval from the world wide web: Issues, techniques, and systems," *ACM Computing Surveys*, vol. 36, no. 1, pp. 35–67, 2004.

[2] P. André, E. Cutrell, D. S. Tan, and G. Smith, "Designing novel image search interfaces by understanding unique characteristics and usage," in *Proceedings of the IFIP Conference on Human-Computer Interaction*, 2009, pp. 340–353.

[3] E. Hoque, G. Strong, O. Hoeber, and M. Gong, "Conceptual query expansion and visual search results exploration for Web image retrieval," in *In Proceedings of the Atlantic Web Intelligence Conference*, 2011, pp. 73–82.

[4] C. J. van Rijsbergen, *Information Retrieval*. Butterworths, 1979.

[5] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong, "Diversifying search results," in *Proceedings of the ACM International Conference on Web Search and Data Mining*, 2009, pp. 5–14.

[6] R. Santos, J. Peng, C. Macdonald, and I. Ounis, "Explicit search result diversification through sub-queries," in *Proceedings of the European Conference on Information Retrieval*, 2010, pp. 87–99.

[7] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1998, pp. 335–336.

[8] C. X. Zhai, W. W. Cohen, and J. Lafferty, "Beyond independent relevance: methods and evaluation metrics for subtopic retrieval," in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003, pp. 10–17.

[9] F. Radlinski and S. Dumais, "Improving personalized Web search using result diversification," in *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2006, pp. 691–692.

[10] R. L. Santos, C. Macdonald, and I. Ounis, "Exploiting query reformulations for Web search result diversification," in *Proceedings of the International Conference on the World Wide Web*, 2010, pp. 881–890.

[11] D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen, "Hierarchical clustering of WWW image search results using visual, textual and link information," in *Proceedings of the ACM International Conference on Multimedia*, 2004, pp. 952–959.

[12] B. Gao, T.-Y. Liu, T. Qin, X. Zheng, Q.-S. Cheng, and W.-Y. Ma, "Web image clustering by consistent utilization of visual features and surrounding texts," in *Proceedings of the ACM International Conference on Multimedia*, 2005, pp. 112–121.

[13] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang, "IGroup: presenting Web image search results in semantic clusters," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 587–596.

[14] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol, "Visual diversification of image search results," in *Proceedings of the International Conference on the World Wide Web*, 2009, pp. 341–350.

[15] P.-A. Moëllic, J.-E. Haugeard, and G. Pitel, "Image clustering based on a shared nearest neighbors approach for tagged collections," in *Proceedings of the International Conference on Content-Based Image and Video Retrieval*, 2008, pp. 269–278.

[16] H. Ding, J. Liu, and H. Lu, "Hierarchical clustering-based navigation of image search results," in *Proceedings of the ACM International Conference on Multimedia*, 2008, pp. 741–744.

[17] K. Song, Y. Tian, W. Gao, and T. Huang, "Diversifying the image retrieval results," in *Proceedings of the ACM International Conference on Multimedia*, 2006, pp. 707–710.

[18] R. van Zwol, V. Murdock, L. Garcia Pueyo, and G. Ramirez, "Diversifying image search with user generated content," in *Proceeding of the ACM International Conference on Multimedia Information Retrieval*, 2008, pp. 67–74.

[19] D. Myoupo, A. Popescu, H. L. Borgne, and P.-A. Moëllic, "Multimodal image retrieval over a large database," in *Proceedings of the International Conference on Cross-Language Evaluation Forum: Multimedia Experiments*, 2009.

[20] B. J. Jansen, A. Spink, and J. Pedersen, "An analysis of multimedia searching on AltaVista," in *Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval*, 2003, pp. 186–192.

[21] D. Milne and I. H. Witten, "Learning to link with Wikipedia," in *Proceedings of the ACM Conference on Information and Knowledge Management*, 2008, pp. 509–518.

[22] M. Strube and S. P. Ponzetto, "WikiRelate! computing semantic relatedness using Wikipedia," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2006, pp. 1419–1424.

[23] E. Gabrilovich and S. Markovitch, "Computing semantic relatedness using Wikipedia-based explicit semantic analysis," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2007, pp. 1606–1611.

[24] D. Milne and I. H. Witten, "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links," in *Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence*, 2008, pp. 25–30.

[25] E. Hoque, O. Hoeber, and M. Gong, "Evaluating the trade-offs between diversity and precision for Web image search using concept-based query expansion," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence - Workshops (International Workshop on Intelligent Web Interaction)*, 2011, pp. 130–133.

[26] G. Strong, E. Hoque, M. Gong, and O. Hoeber, "Organizing and browsing image search results based on conceptual and visual similarities," in *Proceedings of the International Symposium on Visual Computing*, 2010, pp. 481–490.

[27] G. Strong, O. Hoeber, and M. Gong, "Visual image browsing and exploration (Vibe): User evaluations of image search tasks," in *Proceedings of the International Conference on Active Media Technology*, 2010, pp. 424–435.

[28] E. Hoque, O. Hoeber, and M. Gong, "Combining conceptual query expansion and visual search results exploration for Web image retrieval," *Journal of Ambient Intelligence and Humanized Computing*, in press.

Enamul Hoque is currently an M.Sc. student in the Department of Computer Science at Memorial University of Newfoundland. His research interests include image retrieval, search interfaces, and information visualization.

Orland Hoeber received his Ph.D. from the University of Regina in 2007, and is currently an Assistant Professor in the Department of Computer Science at Memorial University of Newfoundland. His primary research interests include information visualization, Web search interfaces, Web search personalization, human-computer interaction, and geo-visual analytics.

Minglun Gong received his Ph.D. from the University of Alberta in 2003, and is currently an Associate Professor in the Department of Computer Science at Memorial University of Newfoundland. His research interests include a variety of topics in computer graphics, computer vision, image processing, pattern recognition, and optimization techniques.