J Ambient Intell Human Comput
DOI 10.1007/s12652-011-0094-7

ORIGINAL RESEARCH

Combining conceptual query expansion and visual search results exploration for web image retrieval

Enamul Hoque · Orland Hoeber · Grant Strong · Minglun Gong

Received: 2 May 2011 / Accepted: 25 August 2011

Abstract Most approaches to image retrieval on the web have their basis in document search techniques. Images are indexed based on the text that is related to the images. Queries are matched to this text to produce a set of search results, which are organized in paged grids that are reminiscent of lists of documents. Due to ambiguity both with user-supplied queries and with the text used to describe the images within the search index, most image searches contain many irrelevant images distributed throughout the search results, and are often focused on the most common interpretation of the query. We propose a method for addressing these problems in which conceptual query expansion is used to generate a diverse range of images, and a multi-resolution extension of a self-organizing map is used to group visually similar images. The resulting interface acts as an intelligent search assistant, automatically diversifying the search results and then allowing the searcher to interactively highlight and filter images based on the concepts, and zoom into an area within the image space to show additional images that are visually similar. Evaluations show that the precision of the image search results increase as a result of concept-based focusing and filtering, as well as visual zooming operations, even for uncommon interpretations of ambiguous queries.

Keywords Conceptual query expansion · Image search results organization · Web image retrieval · Interactive exploration

E. Hoque · O. Hoeber (✉) · G. Strong · M. Gong
Department of Computer Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada
e-mail: hoeber@cs.mun.ca

1 Introduction

Web image retrieval has traditionally followed an approach that is an extension of web document search techniques (Kherfi et al. 2004). The textual information surrounding and associated with a particular image is used as the core information that describes the image. With this textual information associated with an image collection, web search engines are able to leverage their existing expertise and infrastructure associated with searching for documents. This approach can work well when images are concisely and accurately described within web pages, and when searchers provide good descriptions of their image needs. Unfortunately, the accuracy of the descriptions given to images on the web cannot be guaranteed; nor can we rely upon searchers to provide clear textual descriptions of the images they seek. When these conditions are not met, the search results may include many non-relevant images.

Image search tasks can be fundamentally categorized into two types: rediscovery and discovery. Within rediscovery tasks, the searcher knows precisely what image they are looking for. By contrast, for discovery tasks, the mental model of the image that is being sought may be incomplete or vague; searchers may find that many images match their needs to various degrees. Often, web image search tasks fall into this second category. It has also been noted that many web image search queries are associated with conceptual domains that include proper nouns (e.g., people's names, locations, landmarks) (André et al. 2009; Jansen et al. 2003). Searchers of this type will often have a specific focus, but a less specific aim with respect to resolving the image need (i.e., many different images may be viewed as relevant). Complicating matters further, web image search queries are often very short (André et al. 2009), and open to many different interpretations. It is possible that different

Published online: 27 November 2011

searchers may enter the same query, but their intentions and needs may vary significantly from each other. As such, it may be beneficial to have a web image search interface that is able to understand the various conceptual aspects of a query, present a broad range of images covering those aspects, and allow the searchers to actively manipulate and explore within the image search results.

Such an approach moves beyond the traditional information retrieval focus of matching queries to documents, instead providing a web information retrieval support systems (WIRSS) that aims to support the human element of web search (Yao 2002; Hoeber 2008). The goal is to allow searchers to take an active role within the search process, which is particularly valuable for search tasks that are ambiguous or exploratory in nature (Hoeber 2008). By taking advantage of human decision-making processes, additional information can be provided at a high level to improve the searcher's ability to find images that are relevant.

In this paper we present a conceptual query expansion method combined with a neural network based image organization technique. The system retrieves a highly diversified collection of images, and provides an interactive interface for assisting searchers to conceptually and visually explore the image search space. It automatically extracts a list of concepts from Wikipedia that are related to the query, which are used as the source of the query expansion to retrieve a broad range of images. The images are visually clustered using a similarity-based approach that employs a multi-resolution extension to self-organizing maps (SOM). Concepts selected for query expansion are also presented to the searcher in a hierarchical manner based on a conceptual ontology.

The key benefit of this approach is that it allows searchers to begin with short and ambiguous queries, which are automatically expanded to provide a diverse range of images. The searcher can then browse the hierarchy of concepts that produced the expanded queries, focusing on those that provide a more accurate description of their needs and filtering to remove those that are not relevant. In addition, the searcher can visually explore the search results space, zooming into an area that includes images that look like what they are seeking. Used together, these operations empower the searcher to take an active role in the image retrieval process, supporting their ability to refine their image needs as they explore the image search space.

The primary contributions of this research are the novel technique of conceptual query expansion for image retrieval using Wikipedia, and the prototype implementation of a WIRSS system that provides interactive support for the searcher to explore and refine the image search space. An evaluation of the approach illustrates the ability of the interactive features to increase the precision of the search results.

The remainder of the paper is organized as follows. In the next section, an overview of related work is provided. Section 3 outlines the techniques used to extract the most relevant concepts from the given query and generate expanded queries. Section 4 explains the approach used for visually organizing the search results. Section 5 provides an overview of the interactive features that enhance the image search experience. An analysis of how the precision of the search results improves as a result of interactive exploration activities is presented in Sect. 6. The paper concludes with a summary of the research contributions and an overview of future work.

2 Related work

2.1 Query expansion

Query expansion is the process of adding a number of meaningful terms to an initial query in order to produce a better set of search results (Efthimiadis 1996; Bhogal et al. 2007). This process of adding terms can either be manual, interactive, or automatic. Manual query expansion relies on a searcher's expertise and knowledge to identify additional terms to add to their query. Interactive query expansion identifies and presents to the searcher an ordered list of potential query expansion terms, allowing them to explicitly choose which to add to their query. Automatic query expansion calculates and assigns weights to a set of candidate terms; those with the highest weights are added to the initial query. Since different weighting functions produce different results, retrieval performance depends on how the weights have been calculated.

Query expansion techniques may also be classified according to the type of information used in order to generate the expansion, including relevance feedback, corpus dependent knowledge models, and corpus independent knowledge models (Bhogal et al. 2007). Relevance feedback is an established technique for modifying an initial query using words from user-identified relevant documents, or top-ranked documents (pseudo-relevance). Techniques based on corpus dependent knowledge models generate query expansion by matching the original query to a knowledge base that was created based on some or all of the collection being queried. The problem with traditional relevance feedback techniques and corpus dependent query expansion is that they are content driven. The corpus content is analyzed to extract candidate terms for query expansion. This can only work if there are sufficient relevant documents to work with and also that these documents contain a reasonable set of terms that represent the subject area for the query. Corpus independent knowledge models eliminates this problem by taking advantage of an external knowledge base. Such a knowledge base can be in the form of a thesaurus or an ontology. Generic thesauri [e.g., WordNet (Miller 1995)] may be too broad and shallow to provide comprehensive coverage of specific topics, and domain-specific thesauri are expensive to produce and may not be available in many domains.

Unlike the general domain of document-centric information retrieval, query expansion has been studied in only a few instances within the context of web image retrieval. Those that have explored such techniques have shown them to be promising (Myoupo et al. 2009). Since query expansion can be an effective tool in order to promote diversity and provide assistance in query refinement, its use for image search may provide significant benefits for searchers conducting discovery tasks. However there are a number of challenges associated with conceptual query expansion. The first problem is finding a suitable knowledge base that has sufficient coverage of the conceptual domains that are common in image search. While WordNet (Miller 1995) has been used to improve image retrieval (Joshi et al. 2006), it does not contain information pertaining to the proper nouns that are common in image search queries. Using Wikipedia for reformulating queries has shown promise (Myoupo et al. 2009), and is used as the knowledge base in our work.

The second challenge is to design efficient and effective algorithms that can process such semi-structured knowledge to derive and rank the meaningful terms to be used in the query expansion process. A useful approach to this problem is to measure the semantic relatedness between the original query and each of the concepts derived from that query. A number of different methods have been devised to use Wikipedia for this purpose, including WikiRelate! (Strube and Ponzetto 2006), explicit semantic analysis (ESA) (Gabrilovich and Markovitch 2007), and Wikipedia link-based measure (WLM) (Milne and Witten 2008a). Due to the computationally efficiency and accuracy of WLM, we use this approach in our work.

The third issue to resolve is to decide whether to make the query expansion process interactive or automatic. Although it would be possible to provide searchers with a list of the extracted concepts and allow them to explicitly choose how to expand the query, doing so forces them to continue to think about how to describe their image needs, when what they really want to do is look at the images. We have instead chosen to automatically expand the query using the most similar concepts to the original query. In doing so, the process query expansion and image search result diversification process becomes seamless to the searcher, allowing them to delve into the exploration of the images without delay.

The last challenge is to choose a useful method for organizing the image search results. By expanding the query, the number of images within the search results can grow very large. Further, by broadening the search results for a somewhat ambiguous query, many images that are not relevant to a particular interpretation of the query may be included in the search results. As such, the rank of the search results from a particular expanded query may not be as important as the visual features of the images. A useful approach is to take advantage of the visual features of the images when organizing the search results, and then allow the searcher to browse and explore within the search results space.

2.2 Organizing and visualizing image search results

Among the popular modern search engines, the presentation techniques of image search results have changed very little over the last decade: a paged grid layout is used to organize images base on their rank. While such grid interfaces are easy to use, they provide limited ability to manipulate and explore the search results. Further, simply ordering by rank does not promote diversity within the search results set. Rather, similar images from the most common interpretation of the query are grouped near the top of the grid.

A number of works have considered clustering images based on visual features alone (van Leuken et al. 2009), or the combination of tags and visual content (Moëllic et al. 2008). The premise in these approaches is that by clustering images, the diversity within the collection can be shown. Within image search, the problem has been approached from the perspective of jointly optimizing the search precision and diversity using dynamic programming (Deselaers et al. 2009). Unfortunately, such approaches may not be sufficient to capture semantic diversity, which requires a deeper level of knowledge about the query. Cai et al. (2004) organized search results into different semantic clusters with the goal of facilitating browsing activities. They proposed a hierarchical clustering method using visual, textual, and link analysis that is mainly targeted at clustering the search results of ambiguous queries. In a related paper, Wang et al. (2007) proposed an approach for semantic clustering of image search results, based on a textual analysis of the results. Through a user study, they reported a significant improvement in terms efficiency, coverage, and satisfaction. The difficulty with such explicit clustering approaches is that they partition the images, requiring the searcher to explore each cluster separately.

To improve the organization of image search results, a recent trend employs traditional document search techniques combined with Content-Based Image Retrieval (CBIR), with the goal of diversification. In work that was a pre-cursor to Google Swirl, a similarity graph was generated among the results of a textual query using visual features (Jing and Baluja 2008). The PageRank algorithm was applied to the similarity graph to select the authority nodes as the representative images. However, this approach does not analyze semantically related concepts of the textual query; rather it only considers visual features for diversifying the results.

Similarity-based image browsing (SBIB) is an approach that takes advantage of the fundamental aspects of CBIR, but eliminates the need for the searcher to identify a set of relevant images a priori. Images are organized based solely on their features, allowing searchers to explore the collection even if they do not have a clearly defined information need (Heesch 2008). The challenge of SBIB is to arrange images based on visual similarities in such a way as to support the browsing and exploration experience. While a number of different approaches have been proposed in the literature (Heesch 2008; Chen et al. 2000; Torres et al. 2003; Snavely et al. 2006), we use a different method that employs a multi-resolution extension to SOMs. This approach provides both an organizing structure for the images and a measure of importance that can be used to determine which images to display when space is limited (Strong and Gong 2008, 2009). The interactive features within this approach have been shown to be very useful and easy to use (Strong et al. 2010).

3 Conceptual query expansion for image search

One of the main features of this work is the method by which the image search query is automatically expanded. For the short and ambiguous queries that are common in image search, query expansion attempts to capture the various aspects of the query. The objective is to diversify the range of images retrieved, providing a broad view of what is available. The problems of organizing the image search results and allowing the searcher to interactively explore within the image search results are addressed in Sects. 4 and 5.

The process of performing conceptual query expansion of image search queries follows three steps: extracting concepts from Wikipedia, ranking the extracted concepts, and generating the expanded queries. While others have used Wikipedia for query expansion in the context of general web search (Milne et al. 2007), the approach we use is novel in that it takes advantage of specific aspects of image search. The details are explained in the remainder of this section.

3.1 Extracting concepts from Wikipedia

For this work, we use Wikipedia as the knowledge base for the query expansion process. It is an excellent source of information for the purposes of image search since it includes a large number of articles describing people, places, landmarks, animals, and plants. It is densely structured, with hundreds of millions of links between articles. Most of the articles also contain various representative images and associated textual captions.

A dump of the Wikipedia collection was obtained in June 2010, and was pre-processed 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 all of the matching articles (referred to as the home articles) from Wikipedia using its search feature. In the case where the query is ambiguous and Wikipedia suggests multiple interpretations (senses), 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, for each article (concept) within the collection, these linked articles were located and their titles were 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. To ensure the inclusion of all relevant concepts associated with the image captions, we use Wikifier (Milne and Witten 2008b) 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 article(s) $\{h_s | 1 \leq s \leq q\}$ (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 based on the in-links, out-links, and image captions. These concepts provide the basis for the automatic query expansion process.

3.2 Ranking the extracted concepts

Due to the richness of Wikipedia, the number of concepts obtained in the process described above may become very large. While it is good that so much information is available for the query expansion process, there is a risk in expanding the query too broadly resulting in a significant negative impact on precision for a given interpretation of the query. In order to address this potential problem, we rank the extracted concepts and use only those that are most similar to the home articles. 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 of the query (represented by each of the home articles h_s), we distribute these top-N concepts among all of the candidate concepts $\{C_{h_s} | 1 \leq s \leq q\}$. As such, the number of related 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 the home article h_s. Our approach to this problem is to measure the semantic relatedness between the home article and each of the candidate concepts. WLM (Milne and Witten 2008a) is used for this purpose, taking advantage of the hyperlink structure of the associated articles to determine how much they share in common. 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. 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 produce 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 range (0,1), we restrict the relatedness score to that range using 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 defined as above, and k is a system parameter that controls the importance of the concepts derived from the captions. In our prototype implementation $N = 30$, $k = 0.01$. This results in a 10–20% increase in the score for the concepts derived from the captions, with proportionally more importance given when there are more concepts extracted from the home articles.

The outcome of this process is that the top-N concepts are selected from among the candidate articles, such that those from the image captions are given preference over those from the in-going and out-going links of the home articles. Further, they are distributed across all of the interpretations of the query (as provided by matching the query to Wikipedia home articles). These concepts are used as the source for the query expansion.

3.3 Generating expanded queries

In order to ensure that the expanded queries remain focused on the intended topic, the original query Q is pre-pended to each of the top-N related concepts $\{c_r | 0 \leq r \leq N\}$ as $\langle Q, c_r \rangle$. We define c_0 to be null, 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 original query, even when the original query has multiple meanings.

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 $I_t = 300$ in the current prototype). Since the null expanded query (c_0) is the original query, we define $r(c_0, h_s) = 1$ in the above calculation. Each query is sent to the Google AJAX Search service, and the desired number of images are retrieved. Duplicate images are detected and resolved based on the URL of the source image provided by the search engine.

4 Visual search results organization

The difficulty with retrieving a broad and diversified range of images is how to then present these in a way that allows the searcher to focus on the specific aspect of the query in which they have an interested. A naïve approach would be to use a traditional paged grid layout of the images, ordered by their rank in the search results list and perhaps the semantic relatedness between the top-N concepts and the home articles. However, such an approach may not be all that effective in supporting image search tasks since the meaning of the organization of the images may be rather obscure. Since the expanded query produces a broad range of images representing the various interpretations of the query, it is necessary to provide an interface that can allow the searcher to easily find the aspect of their query that matches their search intentions. A visual method for organizing the broad range of images retrieved is well suited to this task.

Our approach arranges images based on their visual similarity on a 2D virtual canvas. In previous work, we studied alternative versions of this type of canvas layout to determine user preferences between organizing the images based solely on their similarity (following a messy-desk metaphor) and in a more structured layout (following a neat-desk metaphor) (Strong et al. 2010). This study found that there is merit to 2D similarity-based organizations since they can improve the search time and result in a positive search experience compared with traditional list-based methods for organizing image search results. This work uses a grid-based layout that follows the neat-desk metaphor, since it provides a familiar interface while visual encoding the degree of similarity between images.

In order to organize images based on visual similarity using a self-organizing map (SOM), we must first extract the visual features from the images. While a number of approaches have been studied within the domain of image processing (Datta et al. 2008), we use a colour-gradient correlation vector since it is efficient to calculate and offers improved organizational performance relative to comparable feature extraction methods such as the colour histogram (Strong and Gong 2009). We then train a 2D SOM in a process similar to (Strong and Gong 2008) to organize the images.

The SOM used here consists of a 2D lattice of interconnected nodes or units, each of which has a weight vector that is of the same dimension as those extracted from the images. Initially the weight vectors are assigned random values. The map is trained by repeatedly choosing random image vectors and finding the best match unit (BMU) that is closest to that vector in terms of Euclidean distance. It then updates all of the units in the BMU's neighbourhood to be similar to the image vector. The amount that the image vector influences the units within the neighbourhood varies with distance from the BMU. Over time the neighbourhood shrinks to one unit forcing the training to converge. At this point we can determine relative positions of the images in 2D by using the positions of their BMUs in the map.

The topology-favouring property of the SOM ensures that images with similar feature vectors are mapped to units within the SOM that are closer to each other, and vice versa. One thing to note is that if the SOM is too small with respect to the number of images to be organized, multiple images may be mapped to the same unit. To avoid this, we ensure that there are approximately three times as many SOM units as there are images. This gives enough room for the images to find unique positions within the SOM.

Due to the large difference between the size of the SOM and the number of images, using the positions from the SOM directly in the organization of the images results in a layout that follows the messy-desk metaphor. However, as noted above, we prefer to provide a more structured organization that follows the neat-desk metaphor. In order to achieve this, we apply a space partitioning algorithm on the image positions provided by the SOM in order to align them to a grid. The algorithm generates a kd-tree (Bentley 1975) of the images by first finding the median value among the horizontal coordinates of all images, and uses this to split the collection of images into left and right halves. It then computes the median value among the vertical coordinates of images in each half, so that each half is further split into top and bottom quarters. The above two steps are repeated until each node contains at most one image. At the end, all images are contained in the leaves of a balanced binary tree. Based on the location of each leaf, a unique and aligned position is assigned to its associated image.

At this stage of the organization, each image has a position on a grid that makes sense based on its similarity to its neighbours. However, very often there will be more images retrieved than can be shown at a reasonable size within the available display space. To address this problem, we provide a priority order for the images and sort them based on this. The order is obtained by generating multiple grids at different resolutions. This multi-resolution structure is constructed in a bottom-up approach from the grid of images generated by the SOM, whereby the dimensions of the grid are repeatedly divided in half. In each step, groups of 2×2 adjacent images become mapped to the same location in the higher-level grid. In order to assign the priority order which determines which image to show at the given grid resolution, the average feature vector of the images for each group is generated, and the image whose feature vector is closest to the average is chosen to be most representative of this space.

As a result, only images with high display priority are shown when there is insufficient space to display all images. The amount of space available is relative to the screen resolution, as well as two interactive parameters: zoom level and image size. As the searcher zooms into a particular area of interest, the viewport becomes smaller, more space is made available, and images with a lower display priority can be shown. This zoom operation, and the associated panning, allow the searcher to identify an area in the search results that is visually interesting, and then zoom to view other similar images that are hidden below. Once the searcher has zoomed in to a level that shows all of the images that are available, further zooming increases the sizes of the images. As a result, there is a smooth zoom effect that moves from creating more space to show more images, to showing more image detail once all of the images are visible.

Since conceptual information is also associated with each image, we have enhanced this interface further to allow the searcher to dynamically adjust the display priority of each image based on concept-based focusing. When a concept is selected, all the images that were retrieved as a result of this concept are temporarily assigned a high display priority, bringing them to the foreground of the image organization and hiding some of the images that had previously been shown to make space for these new high-priority images. These interactive features are described in more detail in the section that follows.

5 Exploration within the image search space

The conceptual query expansion process described above may result in the introduction of many non-relevant images within the search results, especially for ambiguous queries where there are multiple interpretations of the searcher's intent. As such, it is important to provide a mechanism by which the searcher can easily narrow down the search results to those that match their interests. Our work allows the searcher to do this in two complimentary ways: they may use a concept hierarchy to focus and filter the search results, or they may perform visual filtering through pan and zoom operations.

5.1 Conceptual focusing and filtering

In addition to arranging the images from the search results visually, our approach also uses the concepts from which the expanded queries were derrived to support focusing and filtering operations. Each of these concepts is mapped to an ontology using DBPedia (Bizer et al. 2009). This ontology is displayed to the searcher as a hierarchy, with the most semantically similar concepts to the original query placed at the top.

The searcher can use this hierarchy of concepts for both focusing and filtering operations. By clicking on any of the concepts, all the images that were retrieved as a result of that concept are pulled to the front of the display (temporarily increasing their display priority within the image organization process, as described in the previous section); the remaining images are dimmed giving the focused images more visual prominence. An example of this is shown in Fig. 1.

In addition, the searcher can use checkboxes associated with each node in the concept hierarchy to filter the search results, removing the associated images from the display. This feature allows the searcher to quickly scan the names of the concepts, removing those that are obviously not relevant to their meaning for the query.

Together, these two features can allow the searcher to quickly inspect a particular concept (clicking on it will focus the display on the images that were retrieved as a result of this concept), and decide whether to keep it as part of the search results set or remove it. This interactivity allows searchers to easily explore the search results, based on the Wikipedia concepts that generated the expanded queries.

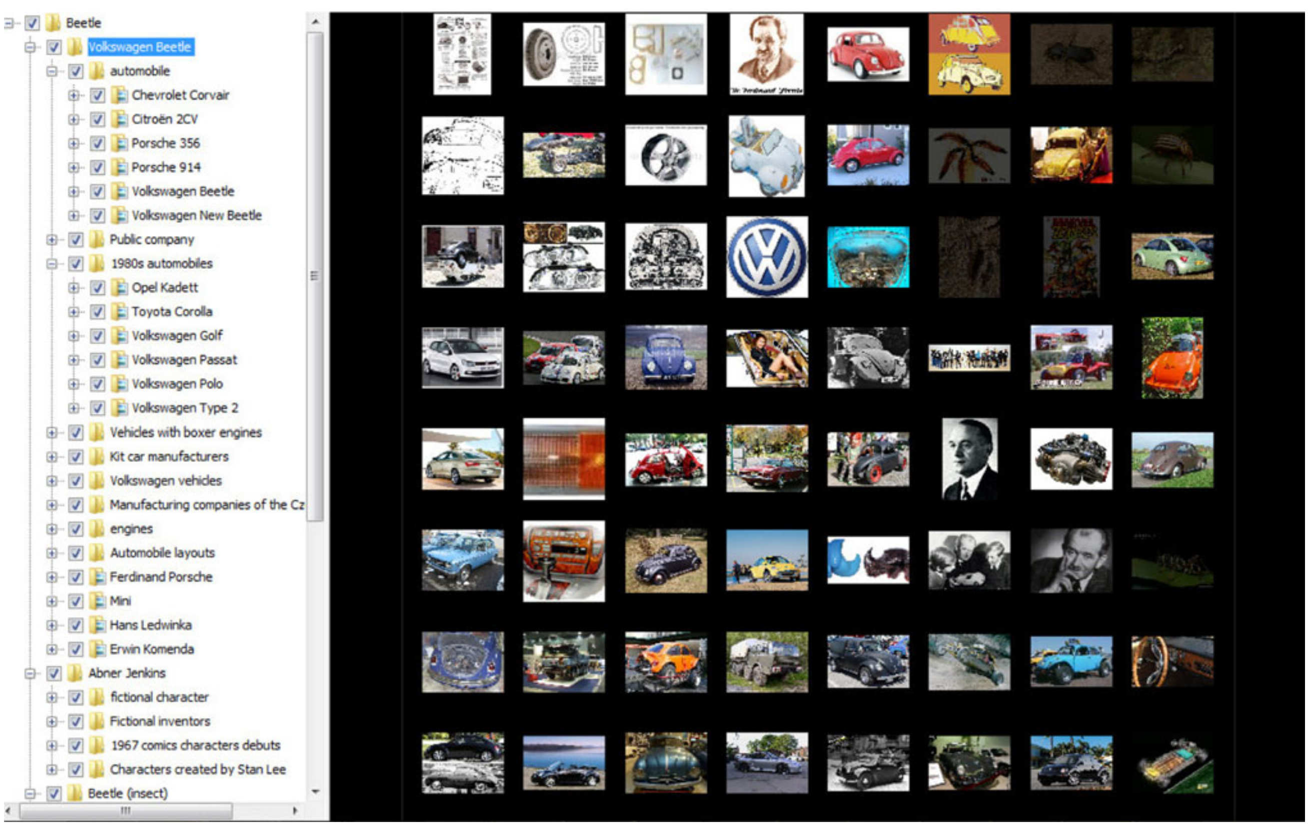


Fig. 1 Concept-based focusing allows searchers to inspect the images that were retrieved as a result the concepts derived from Wikipedia for the query. In this particular example using the query *Beetle*, once the searcher selects *Volkswagen Beetle*, images from this concept are given a high display priority and are pulled to the *front*; all others are *dimmed*

5.2 Visual focusing and filtering

Since the images comprising the search results are organized visually based on a multi-resolution SOM, a logical method for exploring within this image space is through pan and zoom operations. As the searcher zooms into an area of interest and more space is created between the images that are visible, additional images that were previously hidden are shown. At the same time, images that are distant from the focal point of the zoom operation are pushed out of the viewport.

As a result, we can think of this operation as a both a focusing operation (where hidden images that are in the vicinity of the focal point are shown) and a filtering operation (where images that are not of interest to the searcher are removed). Searchers are able to take advantage of their ability to easily scan the visual features of the images to identify those that look like what they are seeking. By zooming into these regions, the searchers can then explore other images that look similar but were previously hidden due to space constraints of the display. Figure 2 illustrates a number of steps in this zoom operation. Note that the empty grid locations within the bottom-level zoom are due to there being fewer images in the search results set than what would be required to entirely fill the grid at this resolution.

6 Evaluation of approach

6.1 Methodology

In order to evaluate the potential benefit of our approach, we studied how the image retrieval performance changed as the image search results space is explored. Within this analysis, it was assumed that searchers were able to make correct judgements regarding which concepts were relevant to their search needs, and were able to identify regions within the visual organization of the images that appeared to include relevant images. This assumption can be justified by the fact that image search is more visual in nature than document search. That is, a searcher generally has a mental picture of the images being sought prior to searching, and when an area of images having visual similarity to this mental picture is shown, it can be easily identified even by a moderately skilled searcher. Within this context, the goal of this evaluation was to measure the ability of the interactive features to support focusing and filtering of the image search results, but not the performance of individual searchers.

We analyzed the system performance using a set of six ambiguous queries. Each of these queries were chosen such that they had a similar level of ambiguity. In particular, each query produced a strong match to three different interpretations within Wikipedia. Using the conceptual query expansion process described previously, 300 images were retrieved for each query. A pair of assessors were chosen from the student body at our university to judge the relevance of each image in the expanded search results set with respect to each of the different senses of the query. Relevance scores were determined by examining each image and the corresponding text snippets; the relevance scores by the assessors were consistent across all images examined. This assessment of relevance was used as ground truth information in the calculation of the precision scores (the ratio of relevant images to the total number of images retrieved).

Since both visible and hidden images may be included within the viewport, we define two different variants of the precision metric for this analysis. *Pv* is the precision

(a) Top-level zoom **(b)** Mid-level zoom **(c)** Bottom-level zoom

Fig. 2 Zooming into a region of the search results filters out images that are distant from the focal point, and includes images that were previously hidden due to space constraints. For search results retrieved using the query *Beetle*, as the searcher zooms into the region that contains automobiles, more images of *Volkswagen Beetles* are shown and images that are relevant to other interpretations are pushed out of the viewport. The *red box* in **a** and **b** shows the zoom areas in **b** and **c**, respectively considering only the visible images within the viewport, and *Pa* is the precision calculated over both the visible and hidden images within the viewport. Images that are outside of the viewport (due to zoom operations) are considered as not having been retrieved within these calculations.

To measure the potential benefit of the conceptual focusing and filtering operations, we analyzed how the precision changed as these features were used. For the focusing operation, *Pv* was measured using the default settings (with no focusing, filtering, or zooming), and when the best concept for each of the three different interpretations of the query were selected. Note that a measurement of *Pa* doesn't make sense here since the focusing operation brings all of the images linked to the selected concept into view. For the filtering operation, the middle frequency interpretation of the query was chosen, and only the top ten relevant concepts were selected using the check-boxes within the concept hierarchy. Both *Pa* and *Pv* were measured before and after this filtering operation was performed.

To study the effect of the visual focusing and filtering operations (i.e., visual zooming), the precision measurements *Pa* and *Pv* were taken at three different levels of zoom for each query. At the top-level zoom (defined as zoom = 0), the viewport was completely zoomed out so that the entire image space was shown, but with many images hidden. At the bottom-level zoom (defined as zoom = 2), the viewport was zoomed in completely so that there were no hidden images, but with many images outside of the viewport. The mid-level zoom (defined as $zoom = 1$) was set at half-way between the top-level and bottom-level zoom settings, such that there were both images outside of the viewport and images hidden. The focal points for these zoom operations were chosen separately for each of the three different interpretations of each query.

6.2 Results

6.2.1 *Conceptual focusing*

The effect of the conceptual focusing operation on the precision the visible images (*Pv*) is illustrated in Fig. 3. For each of the interpretations of the test queries, the focus operation resulted in many relevant images being brought into view. At the same time, many of the images retrieved as a result of the other non-focused concepts became hidden.

The goal of the query expansion process was to diversify the image search results so that they were not all from the most common interpretation of the query. Even so, the precision of the visible images without any visual zooming or conceptual focusing or filtering operations were not evenly distributed across all of the senses of the queries.

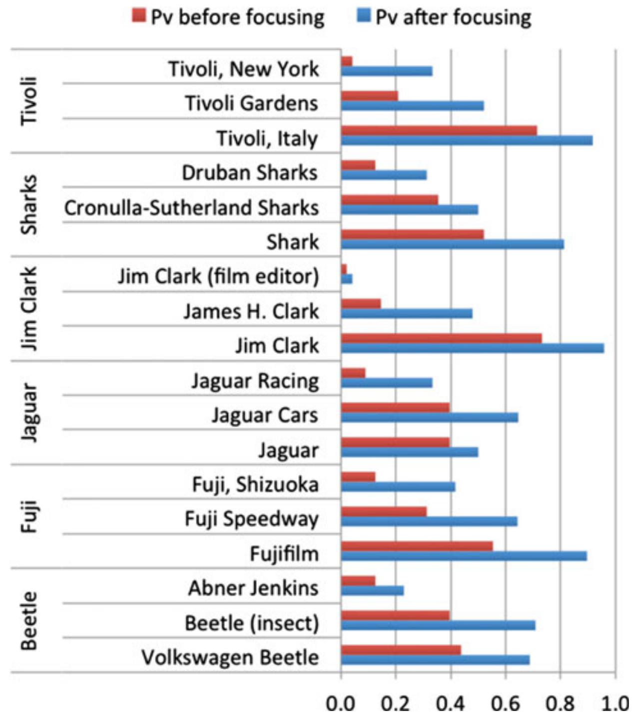


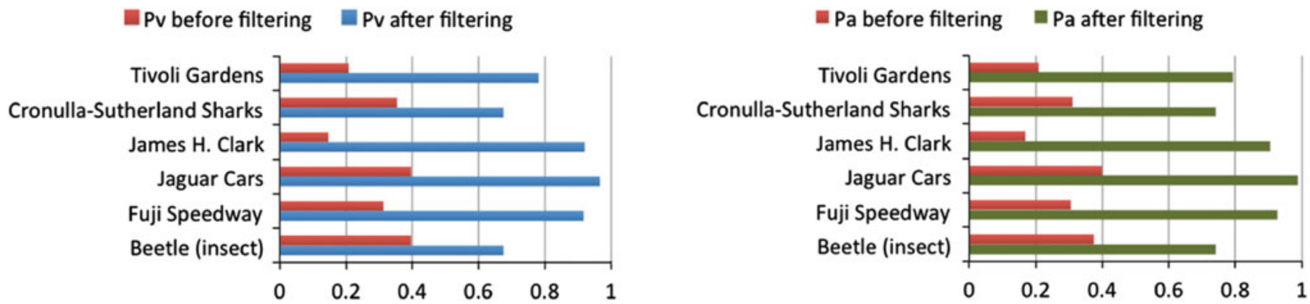
Fig. 3 Precision calculated over the visible images (*Pv*) before and after using the conceptual focusing operation. In all cases, more relevant images are pulled into view by focusing on concepts that are related to each of the interpretations of the queries

This was due to the method by which the query expansion was performed, wherein the number of concepts selected, as well as the number of images retrieved for each concept, varied depending on how closely the concept matched the home article in Wikipedia. However, as a result of the focusing operation, it was possible for many more relevant images to be shown, even for interpretations of the query that were relatively uncommon.

6.2.2 *Conceptual filtering*

The results of the conceptual filtering operation performed for the middle frequency interpretation of the queries are shown in Fig. 4, measured over both the precision of the visible images (*Pv*) and the precision of all the images within the viewport (*Pa*). Here we can see a substantial increase in the precision as a result of only showing the images that were retrieved as a result of the ten best concepts for these interpretations of the queries. That is, by de-selecting the non-relevant concepts, many of the non-relevant images were also filtered out, resulting in increases in precision for each of the test queries.

The similarity between the *Pa* and *Pv* values here illustrates that the images hidden due to space constraints are not any more or less relevant than those that are shown. This outcome provides evidence of the value of the multi-resolution SOM method for organizing the images and



(a) Precision calculated over the visible images (*Pv*)

(b) Precision calculated over all of the images (*Pa*)

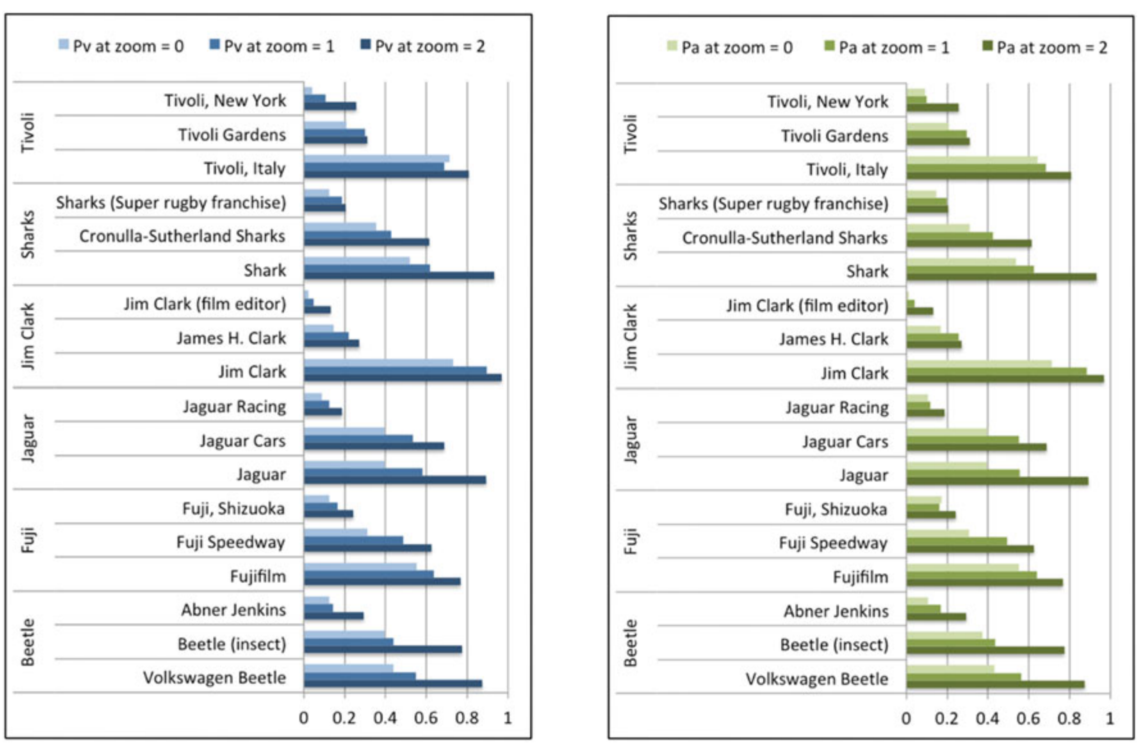
Fig. 4 Filtering out all but the ten most relevant concepts associated with the middle frequency interpretation of the queries results in a significant improvement in precision, both among the visible images and over all of the images within the viewport

defining the display priority to determine which images to show when the display space is constrained.

6.2.3 *Visual focusing and filtering*

The evaluation results for the visual focusing and filtering features, as implemented through the zoom operation, are depicted in Fig. 5. For nearly every interpretation of each of the test queries, as the image space was zoomed into a region that appeared to contain relevant images, both the precision of the visible images and the precision of all the images within the viewport increased. The zoom operation had two simultaneous effects on the image search results. Not only were images moved out of the viewport (effectively providing a filtering operation), but as more space was created, more images that were previously hidden became visible (effectively providing a focusing operation).

Note that at *zoom* $= 0$ (i.e., top-level zoom), the *Pv* values are the same as those reported for the conceptual focusing operation in Fig. 3, and the *Pv* and *Pa* values for the middle interpretation are the same as those reported for the conceptual filtering operation in Fig. 4. Furthermore,



(a) Precision calculated over the visible images (*Pv*)

(b) Precision calculated over all of the images (*Pa*)

Fig. 5 In almost all cases, zooming into regions that appear to contain images relevant to a given interpretation of a query results in an increased precision among the images, both for those that are visible and for all the images within the viewport. Here, zoom $= 0$ is the top-level zoom; zoom $= 1$ is the mid-level zoom; and zoom $= 2$ is the bottom-level zoom the similarity between the Pa and Pv at the different zoom levels is consistent with the findings from the conceptual filtering operation. Indeed, at $zoom = 2$ (i.e., the bottom-level zoom), there are no more hidden images, resulting in $Pv = Pa$.

As with the conceptual focusing operation, the degree to which the zoom operation had the ability to increase the precision depended upon the particular makeup of expanded set of image search results. For most interpretations of the queries, there were many visually similar relevant images that were grouped together, making the zoom operation very effective (e.g., Fuji Speedway). For others, the images were visually dissimilar, resulting in their distribution throughout the image space, limiting the ability of the zoom operation to increase the precision (e.g., Tivoli, Italy).

7 Conclusions and future work

In this paper, we describe an approach for performing conceptual query expansion, producing a diversified set of image search results which are then organized based on their visual features, and presented within an interactive interface. In particular, the approach uses knowledge contained within Wikipedia to extract concepts related to the source query, using these concepts to expand the query and retrieve a diverse range of images. In order to avoid retrieving too many images that are not very closely related to the original query, the approach dynamically controls the number of concepts chosen as well as the number of images that are retrieved for each concept.

Query expansion has the side effect of introducing potentially irrelevant images within the search results, especially when a query is ambiguous and can be interpreted in multiple different ways. To address this problem, a visual and interactive interface allows searchers to perform conceptual filtering and focusing operations using a hierarchical representation of the concepts. In addition, the images themselves are organized using a multi-resolution grid layout derived from a SOM, which not only groups visually similar images, but also provides a solution to the problem when there are more images to show than display space allows. Zooming into the image space results in a visual focusing and filtering operation, displaying more images as space between the images is created, and moving images that are distant from the focal point out of the viewport.

The evaluation of this system provides empirical evidence of the potential benefit of our approach. The conceptual focusing operation was shown to increase the precision of the images that are visible. The conceptual filtering operation was shown to be highly effective in removing non-relevant images among both those that are visible and those that are hidden. Zooming into a region of interest within the visual organization of the images also resulted in improving the precision among both the visible and all the images within the viewport. Within this analysis, all of the experiments were conducted based on the assumption that the searcher would be able to make intelligent choices regarding which concepts were relevant and which regions of the image display space appeared to contain relevant images. That is, these results assume a searcher who does not make mistakes during the search process. The ability of searchers to achieve this level of performance requires further study via a user evaluation.

Future work includes adding features to support complex multi-concept queries and adding additional features that support interactive query refinement loops and query-by-example. We are also examining the benefit of including conceptual information within the image organization process, and exploring alternative methods for visually organizing the images.

References

André P, Cutrell E, Tan DS, Smith G (2009) Designing novel image search interfaces by understanding unique characteristics and usage. In: Proceedings of the IFIP conference on human–computer interaction, pp 340–353

Bentley JL (1975) Multidimensional binary search trees used for associative searching. Commun ACM 18(9):509–517

Bhogal J, Macfarlane A, Smith P (2007) A review of ontology based query expansion. Inf Process Manag 43(4):866–886. ISSN 0306-4573

Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S (2009) DBpedia—a crystallization point for the web of data. J Web Semant Sci Serv Agents World Wide Web 7(3):154–165

Cai D, He X, Li Z, Ma W-Y, Wen J-R (2004) Hierarchical clustering of www image search results using visual, textual and link information. In: Proceedings of the annual ACM international conference on multimedia, pp 952–959. ISBN 1-58113-893-8

Chen C, Gagaudakis G, Rosin P (2000) Similarity-based image browsing. In: Proceedings of the IFIP international conference on intelligent information processing, pp 206–213

Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: Ideas, influences, and trends of the new age. ACM Comput Surv 40(2):1–60

Deselaers T, Gass T, Dreuw P, Ney H (2009) Jointy optimising relevance and diversity in image retrieval. In: Proceedings of the ACM international conference on image and video retrieval, pp 1–8

Efthimiadis EN (1996) Query expansion. Annu Rev Inf Syst Technol (ARIST) 31:121–187

Gabrilovich E, Markovitch S (2007) Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In: Proceedings of the international joint conference on artificial intelligence, pp 1606–1611

Heesch D (2008) A survey of browsing models for content based image retrieval. Multimed Tools Appl 42(2): 261–284

Hoeber O (2008) Web information retrieval support systems: the future of web search. In: Proceedings of the IEEE/WIC/ACM international conference on web intelligence, workshops (international workshop on web information retrieval support systems), pp 29–32

Jansen BJ, Spink A, Pedersen J (2003) An analysis of multimedia searching on AltaVista. In: Proceedings of the ACM SIGMM international workshop on multimedia information retrieval, pp 186–192

Jing Y, Baluja S (2008) VisualRank: Applying PageRank to large-scale image search. IEEE Trans Pattern Anal Mach Intell 30(11):1877–1890

Joshi D, Datta R, Zhuang Z, Weiss WP, Friedenberg M, Li J, Wang JZ (2006) PARAgrab: a comprehensive architecture for web image management and multimodal querying. In: Proceedings of the international conference on very large databases, pp 1163–1166

Kherfi ML, Ziou D, Bernardi A (2004) Image retrieval from the world wide web: issues, techniques, and systems. ACM Comput Surv 36(1):35–67

Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41. ISSN 0001-0782

Milne D, Witten IH (2008a) An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In: Proceedings of the AAAI workshop on Wikipedia and artificial intelligence, pp 25–30

Milne D, Witten IH (2008b) Learning to link with Wikipedia. In: Proceedings of the ACM conference on information and knowledge management, pp 509–518

Milne D, Witten IH, Nichols DM (2007) A knowledge-based search engine powered by Wikipedia. In: Proceedings of the ACM conference on information and knowledge management, pp 445–454

Moëllic P-A, Haugeard J-E, Pitel G (2008) 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, pp 269–278

Myoupo D, Popescu A, Borgne HL, Moëllic P-A (2009) Multimodal image retrieval over a large database. In: Proceedings of the international conference on cross-language evaluation forum: multimedia experiments

Snavely N, Seitz SM, Szeliski R (2006) Photo tourism: exploring photo collections in 3d. In: Proceedings of the ACM international conference on computer graphics and interactive techniques, pp 835–846

Strong G, Gong M (2008) Browsing a large collection of community photos based on similarity on GPU. In: Proceedings of the international symposium on advances in visual computing, pp 390–399

Strong G, Gong M (2009) Organizing and browsing photos using different feature vectors and their evaluations. In: Proceedings of the international conference on image and video retrieval, pp 1–8

Strong G, Hoeber O, Gong M (2010) Visual image browsing and exploration (Vibe): user evaluations of image search tasks. In: Proceedings of the international conference on active media technology, pp 424–435

Strube M, Ponzetto SP (2006) WikiRelate! computing semantic relatedness using Wikipedia. In: Proceedings of the AAAI conference on artificial intelligence, pp 1419–1424

Torres RS, Silva CG, Medeiros CB, Rocha HV (2003) Visual structures for image browsing. In: Proceedings of the international conference on information and knowledge management, pp 49–55

van Leuken RH, Garcia L, Olivares X, van Zwol R (2009) Visual diversification of image search results. In: Proceedings of the international conference on world wide web, pp 341–350

Wang S, Jing F, He J, Du Q, Zhang L (2007) Igroup: presenting web image search results in semantic clusters. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 587–596. ISBN 978-1-59593-593-9

Yao Y (2002) Information retrieval support systems. In: Proceedings of the 2002 IEEE world congress on computational intelligence, pp 1092–1097