# Visual Query Suggestion *

Zheng-Jun Zha †‡, Linjun Yang ‡, Tao Mei ‡, Meng Wang ‡, Zengfu Wang †

† University of Science and Technology of China, Hefei 230027, P. R. China
‡ Microsoft Research Asia, Beijing 100190, P. R. China
junzzustc@gmail.com, {linjuny, tmei, mengwang}@microsoft.com, zfwang@ustc.edu.cn

## ABSTRACT

Query suggestion is an effective approach to improve the usability of image search. Most existing search engines are able to automatically suggest a list of textual query terms based on users' current query input, which can be called Textual Query Suggestion. This paper proposes a new query suggestion scheme named Visual Query Suggestion (VQS) which is dedicated to image search. It provides a more effective query interface to formulate an intent-specific query by joint text and image suggestions. We show that VQS is able to more precisely and more quickly help users specify and deliver their search intents. When a user submits a text query, VQS first provides a list of suggestions, each containing a keyword and a collection of representative images in a dropdown menu. If the user selects one of the suggestions, the corresponding keyword will be added to complement the initial text query as the new text query, while the image collection will be formulated as the visual query. VQS then performs image search based on the new text query using text search techniques, as well as content-based visual retrieval to refine the search results by using the corresponding images as query examples. We compare VQS with three popular image search engines, and show that VQS outperforms these engines in terms of both the quality of query suggestion and search performance.

## Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information search and retrieval—*Query formulation*

## General Terms

Algorithms, Experimentation, Human Factors.

## Keywords

Query suggestion, image search.

* This work was performed when Zheng-Jun Zha was visiting Microsoft Research Asia as a research intern.

## 1. INTRODUCTION

With the rapid advances in both hardware and software technologies, large collections of images have been made available on the Web. To help users find images on the Web, image search has been intensively studied [9, 17, 22]. Many popular search engines (e.g., Google [3], Microsoft Bing [4], and Yahoo! [5]) have developed technologies that allow users to search web images.

Most of existing popular search engines allow users to represent their search intents by issuing the query as a list of keywords. However, keyword queries are usually ambiguous especially when they are short (one or two words). This ambiguity often leads to unsatisfying search results. For example, the query "apple" covers several different topics: fruit, computer, smart-phone, and so on. With such ambiguous query, search engines often return results mixed with "apple"—the fruit, "apple"—the computer, and "apple"—the smart-phone(see Figure 1)[1]. Such results are unsatisfying since users typically prefer search results that could be aligned with their interests, rather than those mixed with diverse categories. Therefore, showing images from one or more categories in which users are truly interested is much more effective and efficient than just returning images from all interpretations. By using a suggested list of expanded queries, users can easily figure out what they are exactly searching and find the target images.

Recently, many query suggestion techniques have been proposed to address the query ambiguity problem. Some existing image search engines such as Google, Yahoo!, and Ask [1] also attempt to address this problem by providing query suggestions. However, these systems usually simply adopt the technique of textual query suggestion. In other words, they suggest a list of keywords based on users' current and history queries. As we know, compared with text, image carries more information that can be perceived more quickly, just like an old saying, "*one image is worth of thousands of words.*" Moreover, there are times and situations that we can imagine what we desire, but are unable to express this intent in precise words [9] [11]. For example, when we saw a Lamborghini car in the street, we may want to search some images about it without knowing its name. How can we formulate the query to find the desired images more effectively? Probably we will input query "car" which is overly general. To help us formulate a specific query, conventional query

---

[1] We use Engine I, II, III to represent three popular image search engines to preserve anonymity.

**Figure 1: The search results of ambiguous query "apple" from three popular image search engines: Engine I, II, and III. It is observed that the search results are mixed with different prototypes of "apple."**

suggestion approaches may suggest the keyword "Lamborghini" (see Figure 2 (a)). However, we have no idea whether it is the one we are interested in. In this case, if one visual example is associated with this suggested keyword "Lamborghini," then we will exactly know it is the one we want and can reformulate a better query that expresses our search intent more clearly (see Figure 2 (b)).

Motivated by the above observations, we argue that images can better help users specify their search intents, and therefore providing visual (i.e., image) query suggestion is a more natural way for image search than only showing textual suggestions. If we can suggest a list of joint visual-textual queries based on users' current queries, not only the ambiguity can be reduced in query formulation but also a better matching between the original text query and images can be achieved. In this way, users would have better search experience. In this paper, we propose a novel query suggestion scheme named Visual Query Suggestion (VQS), which provides users a better query interface to formulate an intent-specific query by simultaneously providing text and image suggestions. It is able to help users express the search intents more precisely. Specifically, it assists users in formulating intent-specific queries by suggesting related keywords to the initial text queries. For each suggested keyword, the representative images associated with this keyword are leveraged to provide visual suggestions in order to further encapsulate users' search intents.

Figure 3 shows the entire procedure of query suggestion for image search in the proposed VQS system. When user are inputting the query, VQS system provides a list of suggestions each containing both representative image and keyword in a dropdown menu, which is quickly responsible for user's operation. User can choose one keyword-image suggestion from the list. Then, VQS system expands the initial query with the corresponding keyword. This results in a composite query, with which VQS system performs image search using text-based search techniques firstly. Then, VQS system regards the corresponding image suggestion as *query example* and refines the initial search results by leveraging visual information, which is potentially useful to improve text-based image search [9, 13, 17, 22]. The re-ranked results are then presented to users, which meet user's intent much better.

To the best of our knowledge, this work represents the first attempt towards formulating query suggestion with both text and image. The main contributions of this paper can be summarized as follows:



(a) Textual query suggestion
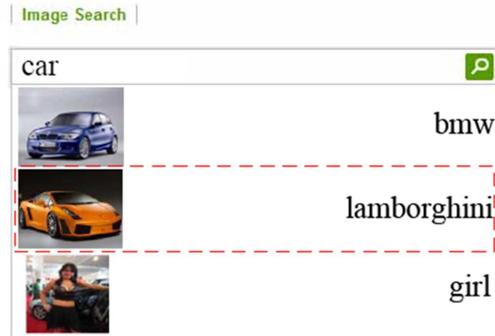


(b) Visual query suggestion

**Figure 2: Comparison between conventional Textual Query Suggestion (TQS) and Visual Query Suggestion (VQS). TQS suggests a list of keywords, while VQS suggests not only the keywords but also the visual examples for the keywords.**

- We propose a new query suggestion scheme named Visual Query Suggestion (VQS) for image search. VQS assists users to formulate an intent-specific query by simultaneously providing both text and image suggestions.

- We develop an easy-to-use query interface, which is able to help users specify and deliver their search intents in a more precise and efficient way.

- As a byproduct, VQS can refine the text-based image search results by exploiting visual information, such that the search results can meet users' information need much better.

The rest of this paper is organized as follows. Section 2 reviews related research on query suggestion. Section 3 provides an overview of the VQS system. The details of VQS and the search strategy with the selected suggestions are
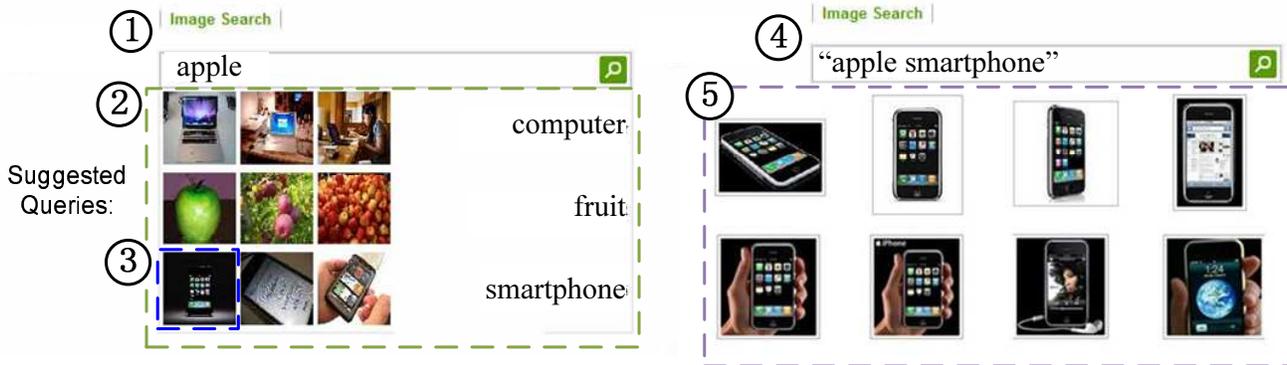
**Figure 3: The workflow of VQS system. A user can interactively: (1) submit an keyword query; (2) browse the keyword-image suggestions provided by VQS system; and (3) select one suggestion to specify the search intent. Then, VQS system: (4) expands the original query with the corresponding keyword; (5) performs image search based on the new query in (4) using text search techniques, preforms content-based visual retrieval to refine the search results by using the corresponding images as *query example*, and returns the final search results.**
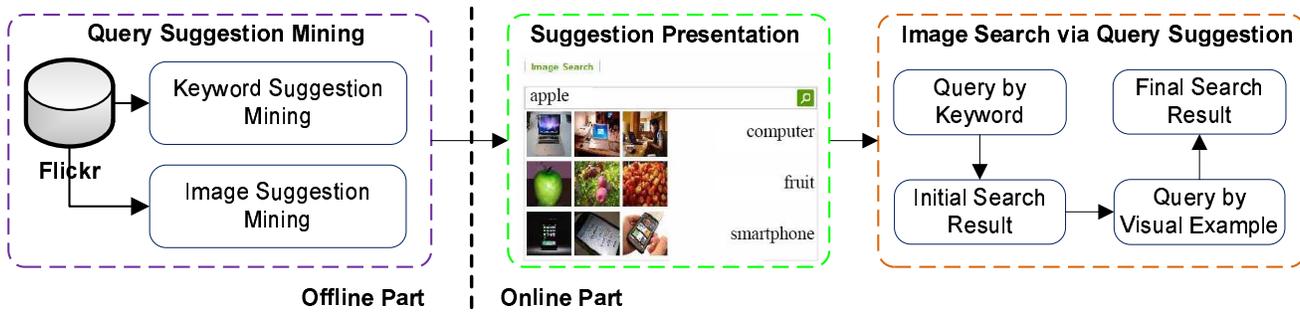


**Figure 4: System framework of VQS. VQS contains three components: (1) query suggestion mining: discovering both image and keyword suggestions for user's current query; (2) suggestion presentation: showing the keyword-image suggestions in a dropdown menu; and (3) image search with query suggestion: performing image search using text search techniques and refining the search results by using the selected image suggestion as *query example*.**

elaborated in Section 4. Section 5 gives the experiments and evaluations, followed by the concluding marks in Section 6.

## 2. RELATED WORK

In recent years, many query suggestion approaches have been proposed to address query ambiguity problem in the information retrieval community. A commonly adopted solution is to find keyword suggestions from the documents retrieved by initial query [8, 16, 25, 26]. For example, Xu *et al.* [25] and Lam *et al.* [16] extracted keywords from the top-ranked documents that are regarded as the relevant results of initial query. Carpineto *et al.* proposed to select the keywords that maximize the divergence between the language model defined by the top-ranked documents and that defined by the entire document collection [8]. Recently, Yu *et al.* selected the keywords from vision-based segments of Web pages to deal with the multiple topics residing problem [26]. Another kind of solution to textual query suggestion is to mine similar queries from search logs [6, 20, 24]. The mined queries are then used as the suggestion for each other. The basic assumption is that two queries are similar to each other if they share a large number of clicked URLs. For exam-

ple, Beeferman *et al.* adopted a hierarchical agglomerative method to mine similar queries in an iterative way [6]. Wen *et al.* used a density-based method to find similar queries by exploiting query content and click-through information [24]. Baeza-Yates *et al.* adopted k-means algorithm to discover similar queries [20]. After the clustering process, the queries within the same cluster were used as suggestions.

Although these methods that designed for text search can be directly applied for image search, they only expand the queries by keywords and thus ignore the visual representativeness of images, which can help users deliver their search intents more precisely.

## 3. SYSTEM OVERVIEW

Figure 4 illustrates the system framework of VQS, which contains three components, i.e., query suggestion mining, suggestion presentation, and image search with query suggestion. In query suggestion mining module, we discover both image and keyword suggestions in order to help user express the search intent more clearly. Specifically, the keyword-image suggestions are generated by exploiting the knowledge from the popular photo-sharing service Flickr [2]. Flickr

contains more than two billion photos that are tagged by billions of tags (keywords). Discovering keyword-image suggestions from such plentiful images associated with abundant keywords is rational and advantageous in the following two aspects: (1) the suggestions can be generated without performing initial search for the original query, which leads to the proposed method being more efficient; and (2) the provided suggestions will not suffer from the unsatisfying quality of the initial search results, which leads to more effective suggestions. A two-step approach is developed to discover the keyword-image suggestions. The first step involves a statistical method which can suggest keywords (i.e., tags) to reduce the ambiguity of the initial query (see Section 4.1). In the second step, for each keyword suggestion, we first collect the images associated with both the initial query and the suggested keyword, and cluster these images, with each cluster representing one aspect of the combined query. We then select the most representative images from these clusters to form the image suggestions (see Section 4.2).

# 4. APPROACH

In this section, we elaborate the implementation of VQS. We will show how the keyword and image suggestions are discovered and how image search is performed with the joint keyword-image suggestions.

## 4.1 Keyword Suggestion

Given an ambiguous query $Q$ (i.e., a keyword or a list of keywords), our goal is to find a set of keywords $\mathcal{S}_Q$ from the whole set of keywords $\mathcal{S}$. These keywords should be able to resolve the ambiguity of $Q$. Therefore, they should satisfy the following two properties [23]:

- **Relatedness**: Each of the selected keywords $q \in \mathcal{S}_Q$ is inherently related to the initial query $Q$;

- **Informativeness**: The selected keywords $\mathcal{S}_Q$ are informative enough such that they can reflect different aspects of the initial query $Q$.

A good example of an ambiguous query is "apple," since it has various meanings. For the query "apple," the keywords "fruit," "computer," "smartphone" are all good suggestions because they are inherently related to "apple" and reflect the different aspects of "apple."

Here we present a probabilistic formulation that simultaneously addresses the above two properties in a single framework. To address the first property, we measure the relatedness between $q_i \in \mathcal{S}_Q$ and $Q$ with their co-occurrence [21]. We calculate the co-occurrence between $q_i$ and $Q$ as the following probability which is normalized by the frequency of $Q$.

$$p(q_i|Q) = \frac{I(q_i \cap Q)}{I(Q)}, \qquad (1)$$

where $I(Q)$ denotes the number of images associated with $Q$, while $I(q_i \cap Q)$ is the number of images associated with both the keyword $q_i$ and the query $Q$. Then, we can define the relatedness between $q_i$ and $Q$ as

$$R(q_i, Q) = f\big(p(q_i|Q)\big), \qquad (2)$$

where $f(\cdot)$ is certain monotonically increasing function. We defined $f(\cdot)$ as the standard sigmoid function in the experiments (see Section 5). Accordingly, the relatedness between

---

**Algorithm 1** Generating keyword suggestions

**Input**: $\mathcal{S}$, $Q$
**Output**: $\mathcal{S}_Q^*$
**Initialization**: set $\mathcal{S}_Q^* = \emptyset$
1: **for each** iteration $t$ **do**
2:     $\mathcal{S}_Q^* = \emptyset$, $L(\mathcal{S}_Q^{(t)}) = 0$;
3:     randomly select the first keyword $q$ from $\mathcal{S} \setminus \mathcal{S}_Q^{(t)}$;
4:     $\mathcal{S}_Q^{(t)} = \mathcal{S}_Q^{(t)} \cup \{q\}$;
5:     select the next keyword $q_i$ from $\mathcal{S} \setminus \mathcal{S}_Q^{(t)}$ by solving
    $\arg\max\limits_{q_i} L(q_i) =$
    $\arg\max\limits_{q_i} \left\{ \lambda R(q_i, Q) + \frac{(1-\lambda)}{|\mathcal{S}_Q^{(t)}|} \sum_{q_j \in \mathcal{S}_Q^{(t)}} D(q_i, q_j, Q) \right\}$
    $L(\mathcal{S}_Q^{(t)}) = L(\mathcal{S}_Q^{(t)}) + L(q_i)$;
6:     **if** $\Delta L(\mathcal{S}_Q^{(t)}) > e$ where $e$ is a threshold **do**
        $\mathcal{S}_Q^{(t)} = \mathcal{S}_Q^{(t)} \cup \{q_i\}$, go to step 5;
    **else**
        end this iteration;
    **end if**;
7: **end for**
    return $\mathcal{S}_Q^* = \arg\max\limits_{t} L(\mathcal{S}_Q^{(t)})$

---

a keyword set $\mathcal{S}_Q$ and $Q$ is given by

$$R(\mathcal{S}_Q, Q) = \sum_{q_i \in S_Q} R(q_i, Q). \qquad (3)$$

To address the second property, we find a set of keywords $\mathcal{S}_Q$ that can diversely reflect various aspects of the initial query $Q$. Each selected keyword $q_i \in \mathcal{S}_Q$ should be informative enough such that it is able to reflect one facet of $Q$. Meanwhile, this facet should be different from those characterized by other keywords $q_j \in \mathcal{S}_Q \setminus \{q_i\}$. We assume that $q_i$ and $q_j$ reflect two different aspects of $Q$ if appending $q_i$ or $q_j$ to $Q$ can result in very different distribution over the remaining keywords $q \in \mathcal{S} \setminus \{q_i, q_j\}$. That is to say, $q_i$ and $q_j$ can resolve the ambiguity of $Q$ if the distribution $p(q|Q \cup \{q_i\})$ and $p(q|Q \cup \{q_j\})$ are quite different [23]. For example, given the query "apple," the keywords co-occurring with {"apple," "fruit"} and those with {"apple," "computer"} are quite different. Therefore, appending "fruit" or "computer" to "apple" leads to two different distributions, i.e., $p(q|\text{"apple"}, \text{"fruit"})$ and $p(q|\text{"apple"}, \text{"computer"})$. Here, we use the symmetric Kullback-Leibler (KL) divergence [15] to measure the distribution difference between $p(q|Q \cup \{q_i\})$ and $p(q|Q \cup \{q_j\})$ as

$$\widetilde{KL}(q_i||q_j) = KL(q_i||q_j) + KL(q_j||q_i), \qquad (4)$$

where

$$KL(q_i||q_j) = \sum_q p(q|Q \cup \{q_i\}) \log \frac{p(q|Q \cup \{q_i\})}{p(q|Q \cup \{q_j\})}. \qquad (5)$$

Accordingly, we define the informativeness of $\{q_i, q_j\}$ with respect to $Q$ as

$$D(q_i, q_j, Q) = g\big(\widetilde{KL}(q_i||q_j)\big), \qquad (6)$$

where $g(\cdot)$ is a monotonically increasing function. The informativeness of a keyword set $\mathcal{S}_Q$ can be measured as $\sum_{q_j, q_k \in \mathcal{S}_Q} D(q_j, q_k, Q)$.

Accordingly, the keywords $\mathcal{S}_Q$ that simultaneously satisfy both the relatedness and informativeness properties can be

found by solving the following equation:

$$\mathcal{S}_Q^* \quad = \quad \arg\max_{\mathcal{S}_Q} \Big\{ \frac{\lambda}{N} \sum_{q_i \in \mathcal{S}_Q} R(q_i, Q) \tag{7}$$

$$+ \quad \frac{(1-\lambda)}{C_N^2} \sum_{q_j, q_k \in \mathcal{S}_Q} D(q_j, q_k, Q) \Big\}$$

where $N = |\mathcal{S}_Q|$ is the number of the selected keywords. $\lambda$ ($0 \leqslant \lambda \leqslant 1$) is a weighting parameter that is used to modulate the two properties.

However, it is computationally intractable to solve equation (7) directly since it is a non-linear integer programming (NIP) problem [7]. Alternatively, we resort to a greedy strategy which is simple but effective in solving NIP problem. The process is illustrated in Algorithm 1.

In real-world problems, most keywords are unrelated to $Q$. Therefore, we perform pre-filtering to filter out the keywords with small values of $R(q, Q)$. As a result, only the keywords with large values of $R(q, Q)$ should be considered. This will further accelerate the suggestion generation process. With the discovered keywords $\mathcal{S}_Q$, we move our effort to generate the visual suggestions in the next section.

## 4.2 Image Suggestion

As aforementioned, VQS system provides not only keyword suggestions but also image suggestions. Here we select representative images for each suggested keyword to form the image suggestions. Consider a suggested keyword $q$ for the original query $Q$, we first collect the images associated with both $q$ and $Q$ from our Flickr image set. Then these representative images are selected from the image collection. As the visual content of the images usually varies largely, the selected images should be diverse enough so that they can comprehensively represent the corresponding keyword. Here, we resort to Affinity Propagation (AP) method which is proposed to identify small number of images that accurately represent a data set of images [10].

Based on the collected image set $\mathcal{I} = \{I_i\}_{i=1}^N$ for $(Q, q)$, and the similarity measure $s(I_i, I_j)$ between two images, our goal is to cluster $\mathcal{I}$ into $M$ ($M < N$) clusters, each represented by the most representative image called "exemplar". In affinity propagation, all the images are considered as potential exemplars [10] [14]. Each of them is regarded as a node in a network. The real-valued message is recursively transmitted via the edges of the network until a good set of exemplars and their corresponding clusters emerge. Let $\mathcal{I}_e = \{I_{ei}\}_{i=1}^M$ denote the final exemplars and $e(I)$ represent the exemplar of image $I$. In brief, the AP algorithm propagates two kinds of information between images: 1) the "responsibility" $r(i, j)$ transmitted from image $i$ to image $j$, which measures how well-suited $I_j$ is to serve as the exemplar for $I_i$ by simultaneously considering other potential exemplars for $I_i$, and 2) the "availability" $a(i, j)$ sent from candidate exemplar $I_j$ to $I_i$, which reflects how appropriately $I_i$ chooses $I_j$ as exemplar by simultaneously considering other potential images that may choose $I_j$ as their exemplar (see Figure 5). This information is iteratively updated by

$$r(i, j) \leftarrow s(I_i, I_j) - \max_{j \neq j'} \big\{ a(i, j') + s(I_i, I_{j'}) \big\},$$
$$a(i, j) \leftarrow \min\{0, r(j, j)\} + \sum_{i' \notin \{i, j\}} \max \big\{ 0, r(i', j) \big\} \tag{8}$$
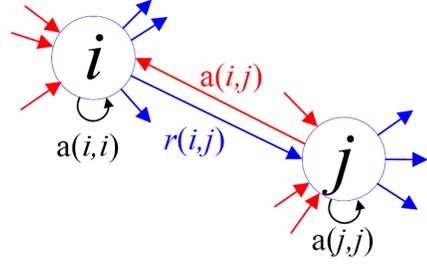


**Figure 5: Example of the affinity propagation over images.** $i$ **and** $j$ **are image index;** $r(i, j)$ **and** $a(i, j)$ **are the "responsibility" and "availability" between image** $i$ **and** $j$**, respectively;** $a(i, i)$ **is the "self-availability" of image** $i$**.**

The "self-availability" $a(j, j)$ is updated by:

$$a(j, j) := \sum_{i' \neq j} \max \big\{ 0, r(i', j) \big\}. \tag{9}$$

The above information is iteratively propagated until convergence. Then, the exemplar $e(I_i)$ of image $I_i$ is chosen as $e(I_i) = I_j$ by solving

$$\arg\max_j \big\{ r(i, j) + a(i, j) \big\}. \tag{10}$$

As pointed out in [14], the original AP algorithm that uses full connected network may lead to a high computational cost of $O(N^2 T)$ where $T$ is the number of iterations. A solution to improving the speed is to perform AP on a sparse similarity matrix instead of the full one. This can be accomplished by constructing a sparse graph structure $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$. $\mathcal{V}$ is the image set and $\mathcal{E}$ represents the edges between images. We construct the graph using the $k$-nearest neighbor strategy [12]. For each data point, we find $k$-nearest neighbors, each of which is connected to a datum point via an edge. Based on the sparse graph, the AP algorithm can be implemented much more efficiently since the information propagation only needs to be performed on the existing edges. However, when we perform AP on such sparse graph, each data point can and only can be the exemplar of $(k+1)$ data points. That is to say, there are at least $N/k$ exemplars, which are much more than expected. To address this problem, we adopt an edge refinement method proposed in [14], which is summarized in Algorithm 2. In each iteration, multiple exemplars may be merged into one cluster. The AP performed on the re-constructed graph may generate fewer exemplars. Once the number of exemplars is reduced to a desirable value, the iteration can be ended. The final exemplars are representative and regarded as the image suggestions.

## 4.3 Image Search via Joint Keyword-Image Suggestion

After the user chooses a keyword-image suggestion, the keyword is appended to the initial query. This results in a composite query, with which VQS system performs image search using text-based search techniques [19]. However, due to the mismatch between the image content and the associated text, the performance of text-based image search is usually unsatisfying. On the other hand, the user-selected image suggestion inherently reflects the user's search intent

**Algorithm 2** Generating image suggestions for each keyword suggestion

---

**Input**: $\mathcal{I}$, $\mathcal{G}$
**Output**: $\mathcal{I}_e$
**Initialization**: set $\mathcal{G}^{(0)} = \mathcal{G}$
1: **for each** iteration $t$ **do**
2:   Generate $\mathcal{I}_e^{(t)}$ with AP on $\mathcal{G}^{(t-1)}$, $\mathcal{I}_e = \mathcal{I}_e^{(t)}$;
3:   Construct $\mathcal{G}^{(t)}$ based on $\mathcal{I}_e^{(t)}$ and $\mathcal{G}^{(t-1)}$
    (1) for each $I_i \in \mathcal{I}_e^{(t)}$, if $I_i$ is the exemplar
       of $I_j$, then an edge between $I_i$ and $I_j$ is added;
    (2) for $I_k, I_l \in \mathcal{I}_e^{(t)}$, if there are two data points $I_m$, $I_n$
       that are the neighbor to each other and satisfy
       $e(I_m) = I_k$ and $e(I_n) = x_l$, then $I_k, I_l$ are
       connected by an edge;
    (3) for $I_k, I_l \in \mathcal{I}_e^{(t)}$, if they are connected in (2), then
       all data points that choose $I_k$ as exemplar are
       connected to $I_l$, and vice versa.
4: **end for**
   return $\mathcal{I}_e$

---

**Table 1: Sample initial queries used in the experiments.**

| Initial Query | |
|---|---|
| airshow | animal |
| apple | building |
| camping | car |
| disaster | flag |
| flight | flower |
| fruit | game |
| insect | panorama |
| Paris | plant |
| portrait | road |
| scenic | season |
| sky | sports |
| sunset | travel |
| weather | |

and the visual content of the image is potentially useful to improve text-based image search [9, 13, 17, 18, 22].

Therefore, we propose to refine the text-based search results by exploiting visual information. A re-ranking method is developed to re-rank the returned images according to the visual similarities between them and the selected image suggestion. It is worth noting that our system is extensible as any other re-ranking algorithm can be easily integrated. Suppose there are $K$ visual modalities (e.g., color, shape, and texture), we first calculate the visual similarities $\mathcal{S}_k = \{s_{ki}\}_{i=1}^N$ between the returned images and the user-selected image suggestion $I_q$ on the $k$-th modality, where $N$ is the number of returned images. Then, all the $K$ visual information are aggregated to refine the initial search results through the following equation.

$$r_i = \alpha_0 r_{0i} + \sum_{k=1}^K \alpha_k s_{ki}, \qquad (11)$$

$$s.t. \qquad \alpha_0 + \sum_{k=1}^K \alpha_k = 1, \quad i = 1, \cdots, N$$

where $r_{0i}$ denote the initial relevance score between the query and the image $I_i$, which is generated by text-based search method [19]. $r_i$ is the refined relevance score, $\alpha_0$ and $\alpha_k$ are the weighting parameters used to modulate the textual and visual information. Since the similarities over different modalities may vary significantly, the visual similarities $\{s_{ki}\}_{i=1}^N$ over each modality are normalized such that $s_{ki}$ is with zero mean and unified variance. The initial relevance scores $r_{0i}$ are also normalized in the same way.

After obtaining the final relevance scores $\mathcal{R} = \{r_i\}_{i=1}^N$, the VQS system presents the images sorted by the relevance scores with a descending order.

## 5. EXPERIMENTS

We conducted extensive experiments and evaluations, including both subjective and objective evaluations, as well as the comparison between the proposed VQS and three popular image search engines. We first evaluated the performance

of query suggestion provided by VQS system, and then investigated the performance of image search via VQS.

### 5.1 Data and Methodologies

To generate keyword-image suggestions, We used Flickr images as our database. Flickr is the most popular image sharing sit that allows users to upload, share, and tag their photos [2]. We have collected 3 million Flickr images, which were associated with about 15 million keywords in total. To evaluate the performance of image search with visual query suggestions, one popular image search engine (i.e., Engine III) was adopted as the baseline search engine. We used Engine III to retrieve images with each initial query and the corresponding renewed query, which is the combination of the initial query and the keyword suggestion. The top 1,000 returned images of each query are crawled to construct the experimental data set . To obtain the ground truth of the relevance orders of the returned images, we resorted to a manual labeling procedure. Specifically, each image was labeled as three relevance levels with respect to the query: level 2—"highly relevant," level 1—"relevant," and level 0—"irrelevant." We invited 20 subjects to manually label the relevance levels of these returned images. Each image was labeled by at least three subjects. The ground truth is obtained through the majority voting of subjects' labeling.

To represent the image content, we extract three types of visual features, including (1) 225-dimensional block-wise color moment based on 5-by-5 division of the image, (2) 128-dimensional wavelet texture, and (3) 75-dimensional edge distribution histogram [19]. The visual similarity between two images is calculated as $\exp(-||x_i - x_j||^2)$ , where $x_i$ is the feature vector of image $I_i$. The monotonically increasing function $f(\cdot)$ in equation (2) and $g(\cdot)$ in equation (6) are defined as the sigmoid function, i.e., $f(x) = g(x) = \frac{1}{1+e^{-x}}$. The tradeoff parameter in equation (7) is empirically set to 0.7. All the tradeoff parameters in equation (11) are set to be the same, i.e., $\alpha_0 = \alpha_k = \frac{1}{K+1}$.

### 5.2 Evaluation of Query Suggestion

We conducted two user studies to evaluate the VQS. The first study aimed to compare VQS with two existing query suggestion services provided by Engine I and Engine II. While the second study evaluated the usefulness of VQS.
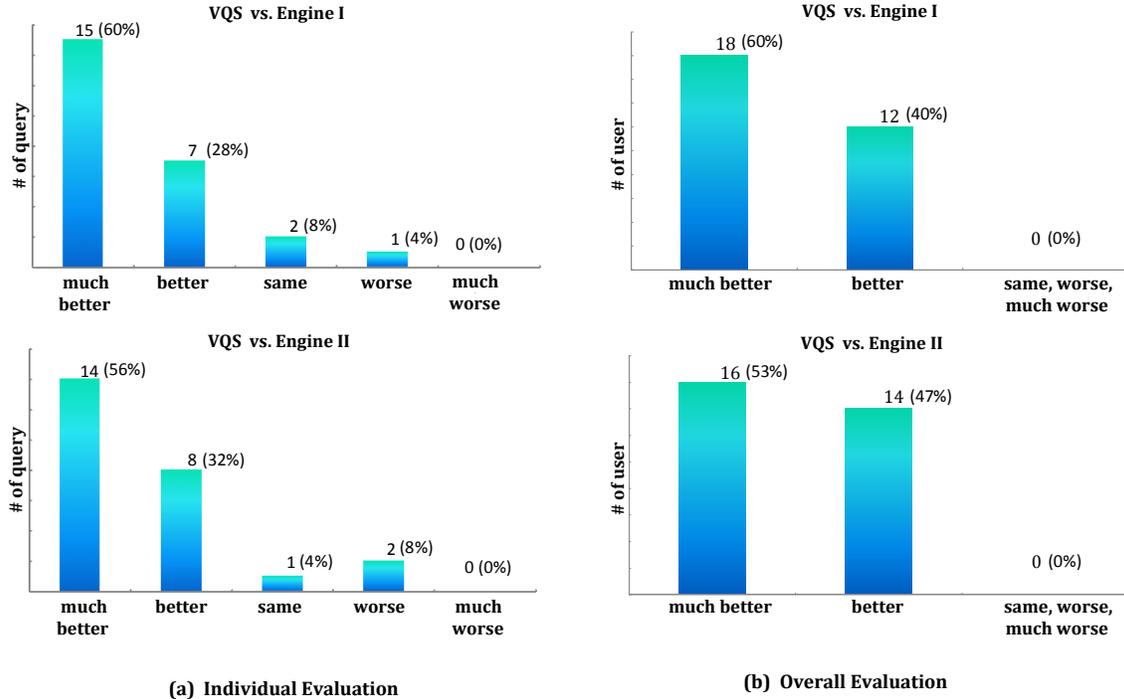
**VQS vs. Engine I**

15 (60%) · 7 (28%) · 2 (8%) · 1 (4%) · 0 (0%)

much better · better · same · worse · much worse

**VQS vs. Engine I**

18 (60%) · 12 (40%) · 0 (0%)

much better · better · same, worse, much worse

**VQS vs. Engine II**

14 (56%) · 8 (32%) · 1 (4%) · 2 (8%) · 0 (0%)

much better · better · same · worse · much worse

**VQS vs. Engine II**

16 (53%) · 14 (47%) · 0 (0%)

much better · better · same, worse, much worse

**(a) Individual Evaluation**

**(b) Overall Evaluation**

**Figure 6: Comparisons between VQS and image search engine I and II from the 30 regular users.**

We invited 30 average image search users evaluate VQS system, including 28 graduate students and two researchers. All of them had the experience of using image search engines more than once per week. We also invited another 10 subjects that are unfamiliar with image search to participate the user studies. These subjects covered a wide variety of backgrounds, such as sales people, marketing people, teachers, government officers and so on. Therefore, there were 40 evaluators in total, including 30 males and 10 females. Their ages ranged from 21 to 55. To avoid any bias on the evaluation, all the participants were selected such that they did not have any knowledge about the current approach for query suggestion and search.

To facilitate the evaluation and comparison, we selected 25 representative queries (see Table 1) from the query log of Engine III. These queries belong to different types such as scene, object, and event. For each query, we selected four keyword suggestions and three image suggestions for each keyword. As a result, there were 300 ($25 \times 4 \times 3$) pairs of initial query and keyword-image suggestion for evaluation.

### 5.2.1 Comparison with Existing Search Engines

Participants were required to submit the 25 queries one-by-one to the three image search systems, i.e., Engine I, Engine II, and our VQS system. Then, they were asked to provide the following evaluations:

- **Individual evaluation** We compared VQS with Engine I and Engine II separately. Based on the observation of the suggestion service of VQS, Engine I(II) for each query, participants were asked to give a score from "2" to "-2," which indicates that VQS performs much better, better, closely, worse, and much worse than Engine I(II), respectively.

- **Overall evaluation**: Similarly, evaluators were also asked to give the overall comparisons between VQS and existing image search engines. They were asked to choose one from the five options: VQS performs "much better," "better," "same," "worse," and "much worse" than Engine I(II).

The results from 30 regular image search users are illustrated in Figure 6. Figure 6 (a) provides the average numbers of the queries with the score "2," "1," "0," "-1," or "-2" from the 30 participants. Compared to Engine I, the VQS system performs much better over 15 queries, better over 7 queries, closely over two queries, and worse on only one query. Compared to Engine II, the VQS system provides much better suggestions for 14 queries and better suggestions for eight queries. Figure 6 (b) shows the overall evaluation. All the 30 users consider VQS system outperforms existing image search engines. Specifically, 60% and 53% users reported that the query suggestions of VQS system are much better than those of Engine I and Engine II, respectively. Figure 7 shows the evaluation result from the 10 evaluators who are unfamiliar with image search. They considered that VQS system performs much better than Engine I on 17 queries, better on 5 queries, and closely on 3 queries, while they thought VQS system performs much better than Engine II over 20 queries, better over 4 queries, and closely over 1 query. In the overall evaluation, all of them thought VQS system outperforms existing image search engines.

### 5.2.2 Evaluation of Usefulness

To evaluate the usefulness of VQS system, participants were invited to answer the question "Is VQS useful for eliciting your true search intents?" They were asked to choose one from three options: "very useful" "somewhat useful," and
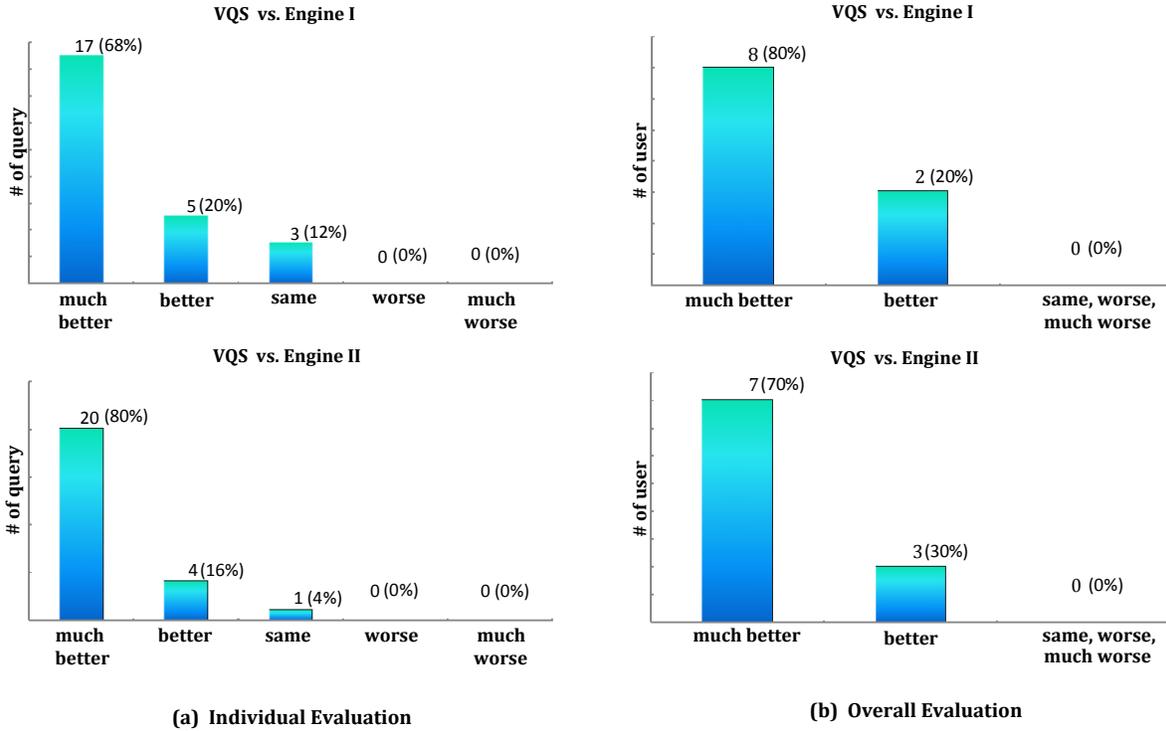
**VQS vs. Engine I**

17 (68%)    5 (20%)    3 (12%)    0 (0%)    0 (0%)
much better    better    same    worse    much worse

**VQS vs. Engine II**

20 (80%)    4 (16%)    1 (4%)    0 (0%)    0 (0%)
much better    better    same    worse    much worse

**(a) Individual Evaluation**

**VQS vs. Engine I**

8 (80%)    2 (20%)    0 (0%)
much better    better    same, worse, much worse

**VQS vs. Engine II**

7 (70%)    3 (30%)    0 (0%)
much better    better    same, worse, much worse

**(b) Overall Evaluation**

**Figure 7: Comparisons between VQS and image search engine I and II from the 10 users who are unfamiliar with image search.**



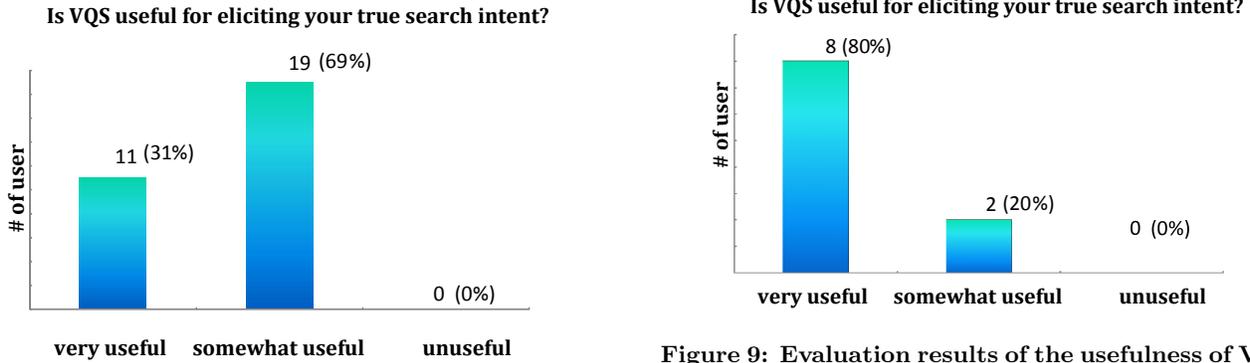**Is VQS useful for eliciting your true search intent?**

19 (69%)
11 (31%)
0 (0%)
very useful    somewhat useful    unuseful

**Figure 8: Evaluation results of the usefulness of VQS from the 30 regular image search users.**

**Is VQS useful for eliciting your true search intent?**

8 (80%)
2 (20%)    0 (0%)
very useful    somewhat useful    unuseful

**Figure 9: Evaluation results of the usefulness of VQS from the 10 users who are unfamiliar with image search.**

"unuseful." Figure 8 shows the evaluation results from the 30 regular image search users. The VQS system was regarded to be very useful by 30% users and be useful to the remaining 70% users. Figure 9 the evaluation result from the 10 users who are unfamiliar with image search. Eight out of the 10 users considered VQS system was very useful and the remaining two thought it was useful.

From the above user studies, we can find that the proposed VQS system outperforms existing popular image search engines and it is useful for better eliciting the true intents of users. We show some exemplars of the keyword-image suggestions for three initial queries in Figure 10. It can be found that the keyword-image suggestions provided by VQS system do reflect different aspects of the initial query and

resolve its ambiguity, and thus they can help users specify the search intent clearly.

## 5.3 Evaluation of Image Search via VQS

We evaluate the performance of three search strategies:

- **IQ**: searching images using the initial query;

- **IQ+KS**: searching images using the combined query consisting of the initial query and the keyword suggestion;

- **IQ+KS+IS**: re-ranking the returned images of IQ+KS based on the selected image suggestion.

The average performance over the 25 queries is reported for the evaluation.

| Initial Query | Image-Keyword Suggestion | | |
|---|---|---|---|
| apple | | | fruit |
| | | | computer |
| | | | Smartphone |
| air show | | | airplane |
| | | | balloon |
| | | | parachute |
| building | | | bridge |
| | | | apartment |
| | | | tower |

Figure 10: Sample keyword-image suggestions of VQS for three initial queries.

We adopt the Normalized Discounted Cumulative Gain at top $k$ (N$DCG@k$) as the evaluation metric. NDCG is a normalized version of DCG measure. Two assumptions of DCG measure are that highly relevant results are more useful when appearing earlier in a result list (i.e., having higher ranks), and that highly relevant results are more useful than marginally relevant ones, which are in turn more useful than irrelevant results. Since comparing search engines' performance only on one query is not comprehensive using DCG alone, the normalized DCG is adopted and NDCG@$k$ is calculated by

$$\text{NDCG@}k = \frac{1}{Z} \sum_{n=1}^{k} \frac{2^{s(p)} - 1}{\log(1+p)} \quad (12)$$

where $s(p)$ is the function that represents reward given to the retrieved image at position $p$, $Z$ is a normalization term derived from the perfect ranking of top $k$ images so that it can normalize NDCG@$k$ to be [0, 1]. In contrast to other measures, such as *precision* and *recall* that only measure the accuracies of retrieved results, NDCG@$k$ takes into account multiple levels of relevance and prefers the retrieved ranking results that are consistent with relevance order. Thus this evaluation measure can better reflect the users' requirement of ranking the most relevant images at top in a real search system.
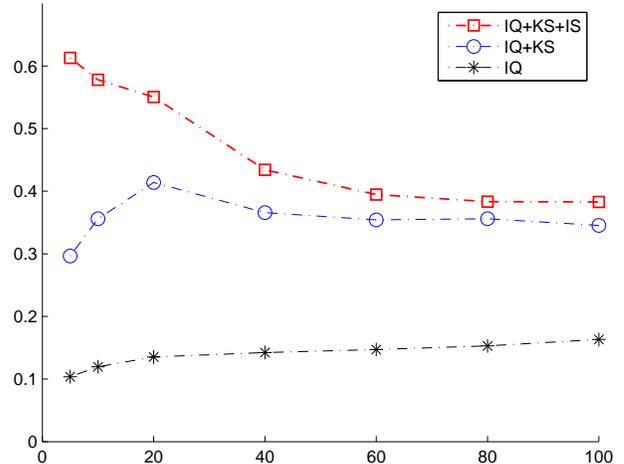


Figure 11: Comparison of NDCG@k for three search strategies: searching images using IQ, IQ+KS, and IQ+KS+IS.

Figure 11 shows the performance comparison of the three approaches (i.e., IQ, IQ+KS, and IQ+KS+IS) with respect to the rank position $k$. Figure 12 shows the top 10 images of different search strategies with two initial queries. We can see that IQ leads to unsatisfying performance due to the ambiguity of the initial query. By appending keyword suggestion to the initial query and searching images with the renewed query, IQ+KS represents the search intent more clearly and thus outperforms the IQ strategy. By further specifying the search intent using image suggestion and leveraging visual information, the IQ+KS+IS strategy gets the best performance.

## 6. CONCLUSIONS

In this paper, we have proposed a new query suggestion technique named Visual Query Suggestion (VQS), which simultaneously provides both keyword and image suggestions and thus is able to help users specify and deliver their search intents in a more precise and efficient way. If the user selects one keyword-image suggestion, the corresponding keyword is added to complement the initial text query. With the renewed query, VQS system performs image search using text search techniques. Afterwards, VQS system regards the selected image suggestion as query example and refines the initial search results by exploiting visual information. Extensive experiments have been conducted to evaluate the proposed VQS system against three popular image search engines. The experimental results show that VQS system outperforms these engines in terms of the quality of query suggestions and search performance.

Our future investigations may include: 1) applying the proposed visual query suggestion on different real-world data sets; 2) integrating other image search re-ranking techniques into the proposed system; and 3) studying the effects of user interaction and click-through on visual query suggestion.

## 7. ACKNOWLEDGMENTS

IQ  IQ+KS  IQ+KS+IS  IQ  IQ+KS  IQ+KS+IS

IS:  IS:

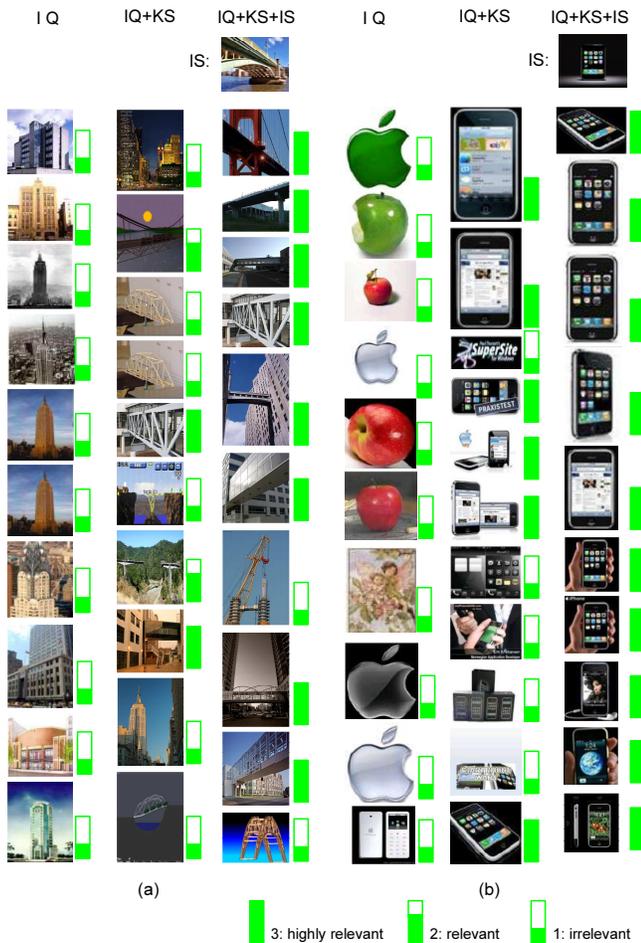(a)  (b)

3: highly relevant  2: relevant  1: irrelevant

**Figure 12: Examples of serch results of IQ, IQ+KS, and IQ+KS+IS. (a) IQ: "building;" IQ+KS: "building bridge;" (b) IQ : "apple;" IQ+KS: "apple smartphone."**

# 8. REFERENCES

[1] Ask image search: http://www.ask.com/?tool=img.
[2] Flickr: http://www.flickr.com/.
[3] Google image search: http://images.google.com/.
[4] Microsoft bing image search: http://www.bing.com/?scope=images.
[5] Yahoo! image search: http://images.search.yahoo.com/.
[6] D. Beeferman and A. Berger. Agglomerative clustering of a search engine query log. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 407–416, Boston, US, 2000.
[7] S. Boyd and L. Vandenberghe. Convex optimization. *Cambridge University Press*, 2004.
[8] C. Carpineto, R. de Mori, G. Romano, and B. Bigi. An information-theoretic approach to automatic query expansion. *ACM transactions on Information Systems*, 19(1):1–27, New York, US, 2001.
[9] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, New York, US, 2008.
[10] B. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 319(5814):726, 2007.
[11] R. J. Gerrig and P. G. Zimbardo. *Psychology and Life (16 Edition)*. Allyn & Bacon, 2001.
[12] D. Heesch and S. RÍzger. Nnk networks for content-based image retrieval. In *Proceedings of European Conference on Information Retrieval*, pages 253–266, Sunderland, UK, 2004.
[13] W. Hsu, L. Kennedy, and S.-F. Chang. Video search reranking via information bottleneck principle. In *Proceedings of ACM SIGMM International Conference on Multimedia*, pages 35–44, Santa Barbara, USA, 2006.
[14] Y. Jia, J. Wang, C. Zhang, and X.-S. Hua. Finding image exemplars using fast sparse affinity propagation. In *Proceedings of ACM SIGMM International Conference on Multimedia*, pages 639–642, Vancouver, Canada, 2008.
[15] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.
[16] A. M. Lam-Adesina and G. J. F. Jones. Applying summarization techniques for term selection in relevance feedback. In *Proceedings of ACM SIGIR Internation Conference on Information Retrieval*, pages 1–9, New Orleans, US,2001.
[17] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: state of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2(1):1–19, New York, USA, 2006.
[18] Y. Liu, T. Mei, and X.-S. Hua. CrowdReranking: Exploring multiple search engines for visual search reranking. In *Proceedings of ACM SIGIR Internation Conference on Information Retrieval*, Boston, USA, 2009.
[19] T. Mei, X.-S. Hua, W. Lai, L. Yang, Z.-J. Zha, Y. Liu, Z. Gu, G.-J. Qi, M. Wang, J. Tang, X. Yuan, Z. Lu, and J. Liu. MSRA-USTC-SJTU at TRECVID 2007: High-level feature extraction and search. In *TREC Video Retrieval Evaluation Online Proceedings*, 2007.
[20] B.-Y. Ricardo, H. Carlos, and M. Marcelo. Query recommendation using query logs in search engines. In *Proceedings of International Conference on Extending Database Technology*, pages 588–596, Heraklion, Greece, 2004.
[21] B. Sigurbjornsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *Proceedings of International conference on World Wide Web*, pages 327–336, Beijing, China, 2008.
[22] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.
[23] K. Weinberger, M. Slaney, and R. V. Zwol. Resolving tag ambiguity. In *Proceedings of ACM SIGMM International Conference on Multimedia*, pages 111-120, Vancouver, Canada, 2008.
[24] J.-R. Wen, J.-Y. Nie, and H.-J. Zhang. Clustering user queries of a search engine. In *Proceedings of International Conference on World Wide Web*, pages 162–168, Hong Kong, China, 2003.
[25] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In *Proceedings of ACM SIGIR Internation Conference on Information Retrieval*, pages 4–11, Zurich, Switzerland, 1996.
[26] S. Yu, D. Cai, J.-R. Wen, and W.-Y. Ma. Improving pseudo-relevance feedback in web information retrieval using web page segmentation. In *Proceedings of International Conference on World Wide Web*, pages 11–18, Budapest, Hungary, 2003.