

Adaptive diversification for tag-based social image retrieval

Amel Ksibi · Anis Ben Ammar · Chokri Ben Amar

Received: 25 June 2013 / Revised: 27 August 2013 / Accepted: 5 September 2013 / Published online: 2 October 2013
© Springer-Verlag London 2013

Abstract With the great popularity of social photos sharing websites, a tremendous volume of digital images is hosted together with their associated tags. Thus, extensive research efforts have been dedicated to tag-based social image search which enables users to formulate their queries using tags. However, tag queries are often ambiguous and typically short. Diversifying search results is a common solution in the absence of further knowledge about the user's intention. Such approach aims to retrieve relevant images covering as much of the diverse meanings the query may have. However, not all queries are uniformly ambiguous and hence different diversification strategies might be suggested. In such a context, two new processes are jointly investigated at query pre-processing and post-processing levels. On the one hand, we propose a multi-view concept-based query expansion process, using a predefined list of semantic concepts, which aims to weight concepts from different views or contexts, aggregate the obtained weights and select the most representative ones using a dynamic threshold. On the other hand, we propose a new ranking process called “adaptive diverse relevance ranking” which automatically predicts an effective trade-off between relevance scores and diversity scores according to the query ambiguity level. Thorough experiments using 12 ambiguous queries over the NUS-WIDE dataset show the effectiveness of our approach versus classical uniform diversification approaches.

Keywords Social image retrieval · Ambiguity · Optimization · Adaptive diversity · Diversification trade-off

1 Introduction

With the spread of Web 2.0, photo-sharing services are hosting a tremendous volume of digital images associated with their users' generated tags [21]. Just as an illustration, 4.5 million new photos are added per day on Flickr¹, 136,000 million new photos and 510,000 posted comments are added every day to Facebook² and 40 million photos are posted daily and 1,000 comments are added per second to the service Instagram³. Given such figures, tag-based social image retrieval is expected to be an intuitive way to perform search, which presents two specific challenges:

- *Tags mismatch*: It occurs when tag query fails to appear with the associated tags to images that are relevant to the query due to the use of synonyms; or due to the incomplete semantic representations (e.g. not containing the tag) [11].
- *Tag ambiguity*: It occurs when a query can be interpreted in several ways other than user's expectation [25].

In the literature, these two challenges have been well studied separately. As far as the 'tags-mismatch' problem is concerned, concept-based approach has been developed to overcome this problem by searching social images based on concept matching rather than tag matching [4, 17].

A. Ksibi (✉) · A. Ben Ammar · C. Ben Amar
Research Groups on Intelligent Machines, University of Sfax,
BP 1173, Sfax 3038, Tunisia
e-mail: amel.ksibi@ieee.org

A. Ben Ammar
e-mail: anis.benammar@ieee.org

C. Ben Amar
e-mail: chokri.benamar@ieee.org

¹ <http://www.kullin.net/2012/01/instagram-now-growing-faster-than-flickr/>.

² <http://thesocialskinny.com/100-social-media-statistics-for-2012/>.

³ <http://instagram.com/press/>.

Indeed, queries and images are transformed from their low-level tag features into semantic concepts as a standardized representation.

As for the 'tag ambiguity' problem, an effective approach is to provide diverse results that cover multiple topics underlying a query. Diversity-based approaches can be categorised as either explicit or implicit [20]. Explicit approaches seek to promote images with maximum coverage of query aspects as characteristic of the query itself, while implicit-based approaches rely on characteristics of the retrieved images to identify diverse images, under the assumption that similar images will cover similar aspects.

In this paper, we attempt to cope with these two problems by taking advantage of the semantic concept-based modelling and diversification approaches. In fact, we investigate the concept-based query expansion using weighted semantic concepts that should overlap the maximum of query aspects. To this end, we reformulate the tag query using a new process called "Multi-view concept-based query expansion (MVCQE)" that aims to weight semantic concepts from different views or contexts, aggregate the obtained weights and select the most representative ones using a dynamic threshold. Despite the fact that query expansion process guarantees the detection of the query aspects, it is unable to ensure the coverage of all query aspects in the first top- k results and to avoid the problem of results redundancy.

To overcome these drawbacks, we need to control the diversification in the ranking process by adding a query post-processing process. To this end, we define a new re-ranking process called "adaptive diverse relevance ranking (ADRR)" that promotes diversity over relevant results using an adaptive diversification trade-off. The main intuition is that not all queries are uniformly ambiguous. Thus, all queries cannot be equally performed using the same diversification degree. Therefore, we would require a retrieval process that does not harm precision for slightly ambiguous queries while respecting the topical diversity of highly ambiguous queries.

This paper is organized as follows: In Sect. 2, we provide an overview of the existing orientations to deal with the ambiguity challenge. In Sect. 3, we describe the general architecture of the proposed "Adaptive tag-based social image retrieval system" where we detail the "MVCQE" process and the "adaptive diverse relevance ranking" process. In Sect. 4, we give experimental setup and results before drawing our main conclusions.

2 Related work

The query ambiguity problem (QAP) has long been recognized as a hard issue in document retrieval. A common method to handle this query ambiguity is to diversify results by displaying a good variety of results covering different

meanings of the query [16]. Diversifying search results can be stated as an NP-Hard optimization problem [19] where the objective is to retrieve a ranking of documents $R(q)$ with the maximum relevance to a given query q and a minimum redundancy ensuring its coverage of all possible aspects underlying q .

The following discussion presents the existing works related to diversifying search results over image retrieval. Then, we detail the related approaches to diversification trade-off optimization.

2.1 Overview of diversification approaches over images retrieval systems

Through the literature, there has been extensive research on how to diversify returned results to satisfy the different users. The challenge is to find the best strategy to order top- k results in such a way to maximize the number of satisfied users relying on only one ranking list. Given an ambiguous query, two strategies can be considered to diversify results [7]: a pre-processing strategy and a post-processing strategy.

2.1.1 Pre-processing strategy

The query expansion process has been considered as an intuitive and promising way to diversify results by adding new meaningful terms from knowledge resources [22]. Different knowledge bases have been exploited for this query expansion. For instance, in [1], authors proposed to expand the query through an open-source knowledge such as WordNet and ConceptNet based on synonyms and concepts. Myoupo et al. [18] proposed to reformulate queries using Wikipedia Knowledge by adding terms that are close to the query. Similarly, Hoque et al. [8] explored Wikipedia resources to ensure query expansion. Given an ambiguous query, they tried to capture its various aspects, and for each aspect, a dynamic number of terms pertaining to the query were discovered from Wikipedia. Weinberger et al. [25] introduced a new tool to disambiguate a tag query using a probabilistic framework. In this work, ambiguity is detected when the same tag generates two tags that occur in two divergent contexts.

2.1.2 Post-processing strategy

Various post-processing techniques have been proposed to deal with the diversity challenge. These techniques can be categorized as search results clustering (SRC) and search results diversification (SRD). While SRC aims to gather similar results in the same cluster, SRD considers the pair-wise similarity inter-documents to iteratively select the document that is not only relevant to the query but also dissimilar to the previously selected documents.

Search result clustering: SRC techniques have shown to be effective to promote diversity ranking thanks to their discriminative power [15, 16, 23]. For instance, in [16], authors created a visually diverse ranking of the image search results through clustering the images based on their visual characteristics. However, such approach is inefficient for practical use since on-line visual clustering is a highly time-consuming process. As an alternative, textual clustering algorithms have been applied such as in IGroup [23], an image search engine, which presents the results in semantic clusters. In Google Image Swirl [10], image search results are clustered into groups and sub-groups, based on their visual and semantic similarity, to reach results diversity.

Search result diversification: Most of existing SRD techniques retrieve a set of documents based on only their relevance scores, and then re-rank them to be diverse, based on a greedy approximation [3]. For instance, Wang [26] proposed a diverse relevance ranking (DRR) scheme for social image search by estimating the relevance scores and the semantic similarities of social images based on their tags. The ranking list was generated by a greedy ordering algorithm which optimizes average diverse precision. However, there are two limits for this approach. First, the relevance scores closely rely on semantic similarity obtained from tags which might harm the search performance. Second, the diversity score evaluated between images neglects the visual diversity.

2.2 Overview of optimizing diversification trade-off

Promoting diversifying search results leads to the danger of broadening the query too much and resulting in a topic drift which significantly decreases precision. Therefore, the trade-off between diversity and precision must be studied to estimate how much diversification is beneficial. Typically, this trade-off is uniformly optimized by maximizing the average diversification performance on a set of training queries. However, not all queries are similarly ambiguous. Thus, different queries might benefit from different trade-offs since any uniform choice for all queries would be suboptimal. The key challenge becomes how to automatically determine an appropriate degree of diversification to promote for an unseen query.

Although the diversification issue has been well investigated in social image retrieval, tuning this trade-off has only been studied recently by hoque et al. [8]. They proposed to automatically estimate the trade-off based on the level of ambiguity of the query itself. This trade-off denotes the number of most related concepts within the query expansion process based on the number of meanings of the query as determined by Wikipedia. That is to say, a highly ambiguous query requires a higher degree of diversification than a spec-

ified query. The main weakness of this approach consists in using lexical resources such as Wikipedia to extract concepts and their weights. Such knowledge resource may only extract the lexical relatedness between the query tag and extracted concepts and cannot reflect the visual relatedness between them.

Motivated by the above observations, we will estimate two diversification parameters, one to control diversity at the query expansion level and another to control diversity at the results re-ranking level. More precisely, we will define

- *An adaptive threshold* per-query by predicting the standard deviation of concept weights for a given query. Indeed, for highly ambiguous queries, this standard deviation should be low as the majority of concepts will be representative of the query and will have similar weights, while for somewhat ambiguous queries, concepts weights will be dispersed, and thus, the standard deviation should be high. With a low level of standard deviation, the threshold should be high to avoid topic drift. In contrast, with a high level of standard deviation, the threshold should be low, since relevant concepts and irrelevant ones are distinguishable.
- *An adaptive search results diversification trade-off* called ‘AmbIDC’ indicator that predicts the level of ambiguity with respect to the number of captured query tag clusters by requesting Flickr.getTagClusters web service. Here, the semantic clusters identify the implicit definitions of an ambiguous tag query based on the social intelligence, resulting in an easy detection of its different contexts. With a smaller value of AmbIDC, the search results will remain more focused and specified. If we increase AmbIDC, the search results will be more diversified.

3 Proposed adaptive tag-based social image retrieval process

Given a set of N predefined concepts, we model a query q and each image in the collection by a vector containing concepts weights. Indeed, we present the query q by a semantic vector $\{c_1^q, c_2^q, \dots, c_N^q\}$ where c_k^q defines the relevance score of concept k to the query q . For an image x_i , we denote, by $\{c_1^i, c_2^i, \dots, c_N^i\}$ the semantic vector containing concepts weights using our annotation approach described in [12] and [14], and we denote, by $\{v_1^i, v_2^i, \dots, v_M^i\}$ the visual vector V_i containing M visual feature values. We denote, also, by $D_q = \{x_1^q, x_2^q, \dots, x_{|D_q|}^q\}$ the set of images that are associated with the set of query concepts C_q . This collection, which is a part of the large set $D = \{x_1, x_2, \dots, x_{|D|}\}$ is obtained by the inverted file generation [5].

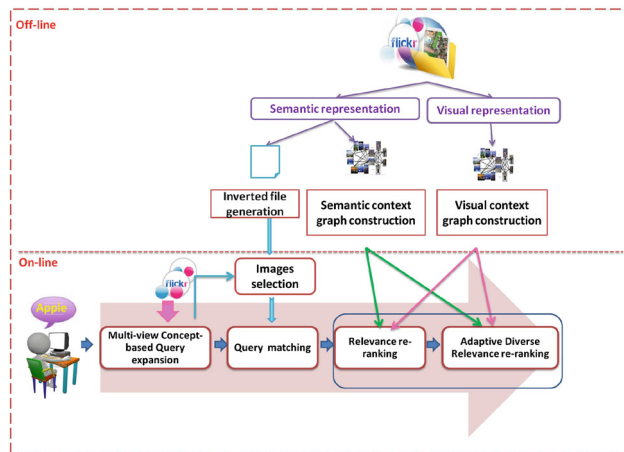


Fig. 1 Proposed adaptive tag-based retrieval process scheme

The proposed system is illustrated in Fig. 1. It consists of two modules: off-line module and on-line module. In the off-line module, we extract the visual and the semantic representation of all images in the collection. Moreover, we build two context graphs: visual context graph and semantic context graph, which will be detailed later on. In addition, we construct the inverted file to minimize the search space. In the on-line module, we analyse the query by applying the multi-view query expansion process. Then, we select a set of images corresponding to selected concepts by MVCQE process. After that, we perform query-images matching to estimate the relevance score of each image. These images will be ordered by relevance. Finally, we perform an Adaptive diverse relevance re-ranking process to get relevant and diverse results.

3.1 MVCQE

The proposed MVQCE process is illustrated in Fig. 2. Our aim is to transform the tag query q to a vector $\{c_1^q, c_2^q, \dots, c_N^q\}$ defining the relevance score of each concept to the query q [13]. Concepts having the highest scores should cover as many of the diverse meanings the query may have. To this end, a step of tag disambiguation is, first, performed, consisting in capturing the different contexts related to this query. Each context defines a view of the query, characterizing a set of related aspects to this view. These contexts are harvested using the “flickr.tags.getClusters” service of Flickr. This service provides the related semantic clusters for a given tag. As the illustration in Fig. 3, the query ‘Apple’ has four semantic clusters: mac, fruit, ipod and newyork. These clusters are obtained by performing a clustering algorithm of all the related tags to the tag in question.

After tag disambiguation, each query is expanded by its related contexts. For example, the query “Apple” becomes “Apple, mac, fruit, ipod, newyork”. Next, concept-based

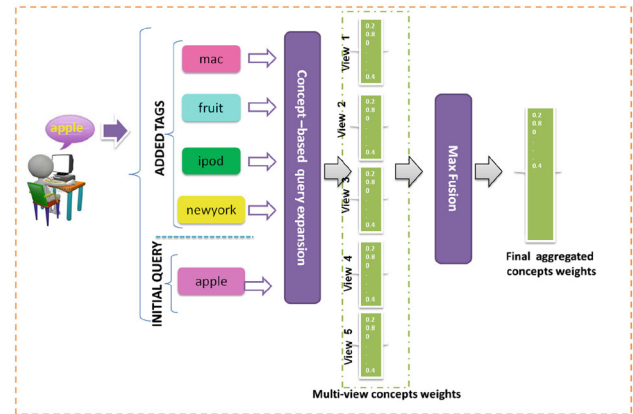


Fig. 2 Proposed multi-view concept-based query expansion process scheme



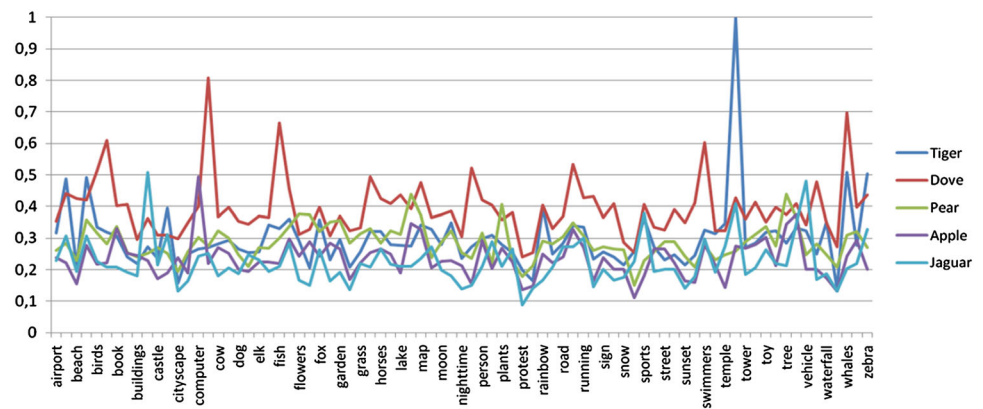
Fig. 3 Flickr-based tag disambiguation

query expansion is performed for each view by mapping each term over the semantic space. By doing so, each view is represented by a linear combination of weighted concepts. After that, we aggregate the obtained combinations by selecting the maximum value of each concept over the different representations to build the multi-view expanded query vector.

The final step is concept selection where the objective is to choose an optimal subset of weighted concepts from the available set that are able to capture the majority of query aspects and avoid topic drift risk. A challenging question is where to stop selecting concepts from a ranked list of concept weights? In other words, how can one select such a rank threshold given a ranked list [2]?

In order to select concepts, such ranking can be thresholded at an arbitrary rank or score. This threshold improves the diversification performance of the retrieval process as it can also be considered as a diversification parameter: too tight a threshold would extract only the common senses of tag-query, while too loose a threshold would produce too query broadening resulting in a topic drift problem. Typically, this threshold is uniformly optimised so as to maximize the precision performance on a set of training queries. So, a fixed

Fig. 4 Concept weight distribution for different queries



threshold is estimated for all queries. Figure 4 shows the distribution of concept weights for different queries. From this figure, it is clear that different queries have a different degree of dispersion among the weights. So, any uniform choice of the threshold for all queries would be suboptimal. A main factor determining the optimal threshold consists in weights distribution. Since this factor is query dependent, the threshold should be selected dynamically per query. To achieve this objective, we develop a new method defining an optimal trade-off score τ , per query, using concept scores and their distribution as input. Indeed, we opt for threshold optimization per-query by focusing on the dispersion degree among the scores. The threshold will be estimated using the following formula:

$$\tau = \begin{cases} \frac{1}{N} * \sum_{i=1}^N c_i^q & \text{if } \sigma \geq \epsilon \\ \frac{1}{N} * \sum_{i=1}^N c_i^q + \sigma & \text{else} \end{cases}$$

where $\epsilon \in [0, 1]$ is a heuristic parameter, c_i^q is the weight of concept i , σ is the standard deviation of concepts weights and N is the number of concepts. For a given query vector, if the dispersion value among the concept scores is high, relevant concepts can be distinguishable from irrelevant ones. So, we estimate the threshold as the average of all scores. In contrast, relevant concepts and irrelevant ones are indistinct. In such a case, we add the value of standard deviation to the average to estimate the threshold.

3.2 Query-image matching

Several measures are employed to evaluate the similarity between the user request and a given image. In [6], experimental studies have argued that cosine similarity performs the best for our context. Given two semantic vectors corresponding to the query q and a given image x_i , the semantic similarity between them is defined using the cosine similarity measure as follows:

$$S_s(q, x_i) = \frac{\sum_{t=1}^N c_t^q * c_t^i}{\sqrt{\sum_{t=1}^N c_t^{q^2}} * \sqrt{\sum_{t=1}^N c_t^{i^2}}} \tag{1}$$

This semantic similarity is computed between the vector of weighted concepts for the query and each vector of weighted concepts for all images that belong to the sub-collection D_q relative to this query. Images are, finally, ranked according to associated relevance values.

3.3 Results re-ranking

In this section, we give a detailed description of the proposed re-ranking process called Adaptive diverse relevance reranking.

3.3.1 Relevance score estimation

A large proportion of the initial retrieved results may not be related to the query causing noisy results. To surmount such a problem, we apply a relevance re-ranking model using random walk with restart process [5] by exploring contextual correlation inter-images, assuming that visually and semantically similar images to top ranked images should be upward. For this purpose, we define a similarity measure between images x_i and x_j , called hybrid similarity, by fusing both their semantic and visual similarities, as follows:

$$S_h(x_i, x_j) = \max(S_s(x_i, x_j), S_v(x_i, x_j)) \tag{2}$$

S_s referring to semantic similarity is also estimated using cosine measure. The visual similarity S_v is calculated based on the Gaussian kernel function with a radius parameter σ as follows:

$$S_v(x_i, x_j) = \exp\left(-\frac{\|V_i - V_j\|^2}{\sigma^2}\right), \tag{3}$$

where V_i and V_j are, respectively, the weighted bag of visual features of images x_i and x_j . The value of the parameter σ is empirically computed as the average of all the pair-wise Euclidean distances between images. We denote by W the

context matrix whose element W_{ij} indicates the hybrid similarity between images x_i and x_j . Elements on the diagonal of W are null.

Implementing the random walk with restart process (RWS) needs two important inputs: the importance of nodes which are the relevance scores of images and the probability transition matrix \tilde{W} which is the normalized matrix of W using the normalized Laplacian graph, as follows:

$$\tilde{W} = D^{-1/2} * W * D^{-1/2}, \quad (4)$$

where

$$D_{ii} = \sum_{j=1}^{|D|} \tilde{W}_{ij} \quad (5)$$

\tilde{W}_{ij} indicates the probability of the transition from node i to node j . We denote by $r_k(x_i)$ the relevance score of image x_i with respect to the set of query concepts C_q at iteration k . The process of RWS is formulated as follows:

$$r_k(j) = c * \sum_i^N r_{k-1}(i) * \tilde{W}_{ij} + (1 - c) * r_0(j), \quad (6)$$

where $c \in [0, 1]$ is a weight parameter to be determined. The initial value of the relevance score vector r_0 is v containing the initial relevance scores of the images. In the process of re-ranking, random walk proceeds until convergence.

3.3.2 Diversity score estimation

Intuitively, a convincing diversity score estimation should contain both a semantic diversity and visual one. First, we define the semantic diversity score of the image x_i as follows:

$$D_s(x_i) = \frac{1}{\text{rank}(x_i) + 1} * \sum_{x_j > x_i} (1 - S_s(x_i, x_j)) \quad (7)$$

In fact, we consider that semantically diverse results should include at least one representative image from each query subtopic. If an image is, on average, semantically dissimilar to the other images ranked higher, it will have an important diversity score.

Second, to define the visual diversity score of an image, we search its minimal difference with the images appearing before it for a given query. To be diverse, an image must be visually different from all the images ranked before it.

$$D_v(x_i) = 1 - \max_{\text{rank}(x_j) < \text{rank}(x_i)} (S_v(x_i, x_j)) \quad (8)$$

These two proposed measures are the ones we adopt in our system, but it is worth mentioning that we can also adopt other diversity score measures.

Finally, to balance between the visual and semantic diversity values, we use a trade-off parameter λ as follows:

$$D_f(x_i) = \lambda * D_s(x_i) + (1 - \lambda) * D_v(x_i) \quad (9)$$

3.3.3 Balancing diversity and relevance scores

Studying the trade-off between diversity and precision is strongly required to estimate how much diversification is beneficial. In such a context, our contribution is to fuse relevance-based ranking and diversity-based ranking for each query unevenly according to the level of query ambiguity. To this end, we suggest the AmbIDC indicator that quantifies this level. Thus, the final ranking $r_{dr}(x_i)$ of a retrieved image x_i is generated by the following formula:

$$r_{dr}(x_i) = (1 - \text{AmbIDC}) * r(x_i) + \text{AmbIDC} * d_f(x_i) \quad (10)$$

Our aim is, then, to fulfil the diversification objective by setting AmbIDC high enough to capture the broad range of query aspects, but not so high as to have a significant adverse affect on precision. To achieve this objective, the quantification of ambiguity level is estimated proportionally to the number of obtained clusters related to the corresponding query. These clusters are captured, as a social tag disambiguation process, using Flickr.tag.GetClusters service. The greater the number of clusters is, the more the tag is ambiguous, the more is the need to diversify results. Therefore, The AmbIDC indicator is estimated using the following formula:

$$\text{AmbIDC}(t) = 1 - \left(\frac{1}{|\text{clusters}(t)|} \right) \quad (11)$$

where t is an ambiguous query tag, $|\text{clusters}(t)|$ defines the number of clusters deduced by the Flickr service.

4 Experiments and results

4.1 Experimental setup

To validate our proposed retrieval process, we conduct experiments on the challenging real-word NUS-WIDE dataset. It is one of the largest social media datasets which contains 269,648 Flickr images accompanied by their associated tags and their visual features (500 Bag of visual words). Each image is also indexed by 81 concepts. All the images are employed as the database images for retrieval, and all the 81 concepts are used for query processing. In addition, we select a set of 12 common ambiguous tag-queries, including Apple, Jaguar, Dove, Pear, Jordan, Eagle. . . For performance evaluation, we use average diverse precision (ADP) [21], Subtopic Recall at rank n (S-Rec@ n) and Diversity at rank n (Div@ n).

- Subtopic Recall at rank n (S-Rec@n): Consider a query Q with h subtopics $\{S_1 \dots S_h\}$ and a ranking $\{d_1 \dots d_m\}$ of m documents. Let subtopics D_i be the set of subtopics to which d_i is relevant. the Subtopic recall at rank n, denoted S-Rec@n, is defined as the number of subtopics retrieved up to a given rank n divided by the total number of subtopics:

$$S - \text{Rec}@n = \frac{|\text{subtopics}(d_i)|}{h} \tag{12}$$

- Diversity at rank n (Div@n): To describe the diversity of ranking, we use the Diversity at rank n (Div@n) defined as

$$\text{Div}@n = \frac{|\text{subtopics}(d_i)|}{N_{\text{rel}_n}}, \tag{13}$$

where N_{rel_n} is the number of relevant images in top n results.

Before performing evaluation, we should define our approach parameter values. On the one hand, the parameter c is a damping factor that ensures the convergence of the random walk process. c fluctuates between 0 and 1. we test different values of c in this interval and we choose empirically $c=0.8$. On the other hand, the parameter which defines the convergence criteria in the random walk process is fixed empirically as 0.00001.

4.2 Study of MVCQE effectiveness

We study the effectiveness of MVCQE process by responding to the following questions:

- To which extent does social knowledge improve the detection of context performance underlying an ambiguous query?
- What is the impact of using Multi-view on diversifying results?
- What is the impact of using adaptive threshold in MVCQE?
- What is the impact of knowledge resource selection in adaptive threshold in MVCQE?

4.2.1 What is the impact of using multi-view on diversifying results?

In this experiment, we compare our proposed MVCQE approach with the baseline concept-based query expansion(CQE) [24] and tag-based query expansion (TQE) [9].

Concepts	CQE	Concepts	MVC-QE	Concepts	TQE
tiger	1	tiger	1	cat	1
whales	0,5091	animal	0,6847	animal	1
zebra	0,5039	computer	0,5689	animals	1
bear	0,4915	whales	0,509	nature	1
animal	0,4861	zebra	0,5039	wildlife	1
rainbow	0,3968	bear	0,4915	bigcat	1
cat	0,3951	leaf	0,4757	white	1
sports	0,3624	rainbow	0,3967	stripes	1
flags	0,3611	cat	0,3951	wild	1
fox	0,3573	book	0,3788	flickrbigcats	1
waterfall	0,351	flowers	0,3707	canon	1
mountain	0,3482	birds	0,3674	nikon	1
temple	0,3471	sports	0,3624	tigre	1
map	0,3429	garden	0,3614	cats	1
fire	0,3417	flags	0,3611	feline	1

Fig. 5 Different interpretations for query “Tiger”

In TQE, tag query is expanded by its top- k related tags from Flickr that frequently co-occur with the original query.

In Fig. 5 below, we illustrate the obtained tag query reformulations for the query ‘Tiger’ using the aforementioned approaches. Top-15 related concepts and tags are extracted. For TQE, the coverage of query aspects is low as all the selected tags are related to one context for the query ‘Tiger’. CQE outperforms TQE by capturing more aspects and contexts than TQE. This improvement is due to the ability of the predefined list of semantic concepts to cover the comprehensive human world knowledge. However, we note that top-5 selected concepts belong to the same context resulting in a lack of capturing all related contexts of a tag query (such as Tiger Operating System). MVCQE performs the best by detecting more diverse aspects in the top-5 selected concepts than in TQE and CQE when including “computer” in its top selection. The success of our approach MVCQE is due to the diversification of top- k selected concepts yielding diverse query aspects and diverse search results. In fact, the multi-view concept weighting with respect to different contexts is responsible for this diversification.

4.2.2 What is the impact of using adaptive threshold in MVCQE?

In this experiment,we investigate two types of thresholding: static and adaptive. In the static thresholding, the same fixed pre-selected rank threshold τ is applied to all queries. We test different scores of τ at 0.2, 0.3 and 0.4. In the dynamic thresholding, we estimate the optimal threshold score for each query with respect to the dispersion degree among the concept weights as described above. We obtain the following results in Figs. 6 and 7.

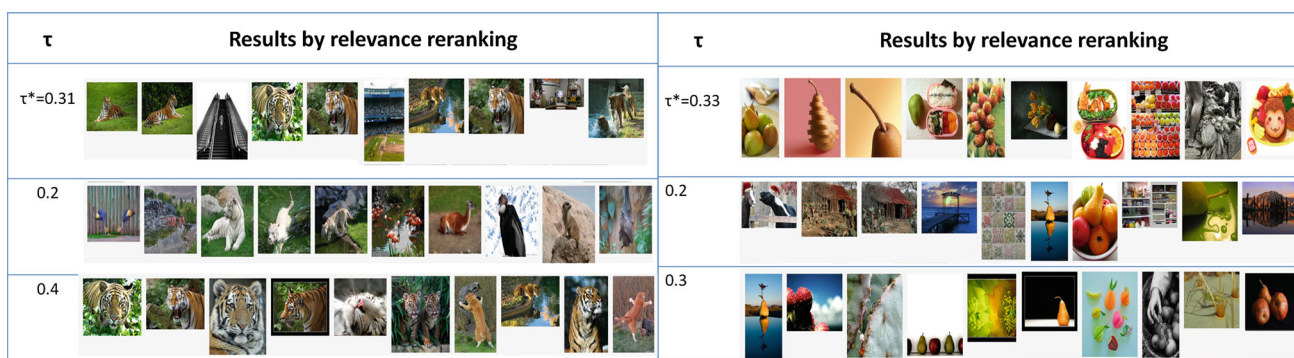


Fig. 6 Obtained results for queries ‘Tiger’ and ‘Pear’ using different threshold values

	σ	τ	mean
dove	0.104	0.39	0.39
pear	0.053	0.33	0.28
apple	0.059	0.29	0.231
tiger	0.137	0.31	0.31

Fig. 7 Optimal thresholds for different queries with respect to dispersion degree among concept weights

It can be seen from these figures that defining a unified threshold score for all queries would be suboptimal, which proves the need for a dynamic thresholding. Actually, the static threshold hurts the retrieval precision and leads to topic drift risk in case of low value of the threshold or under-estimation of the query in case of a high value of the threshold.

4.2.3 What is the impact of knowledge resource selection on adaptive threshold in MVCQE?

In this experiment, we study the impact of knowledge resource choice in the threshold estimation. For this purpose, we compare Flickr and Wikipedia resources. Indeed, we estimate the correlation between concepts and query using Flickr Context Similarity measure (FCS) over Flickr resources and DiscoDistance over Wikipedia resources. Figures 8 and 9 illustrate the results:

From Figs. 8 and 9, we notice that weights of concepts for queries ‘tiger’ and ‘jaguar’ are more dispersed than those for queries ‘apple’ and ‘pear’. Thus, we deduce that the dispersion of concept weights can be the main factor to determine the optimal threshold. In addition, we note that the interval of weight repartition differs when changing the knowledge resource. Thus, the mean of all concept weights for each query can be also a factor to determine the desired threshold. These deductions prove, heuristically, the efficiency of our proposed threshold formula.

4.3 Study of ADRR effectiveness

4.3.1 Trade-off between visual diversity and semantic diversity evaluation

In this experiment, we aim to study the sensitivity of our ranking process with respect to the parameter λ . This parameter is a trade-off between visual diversity and semantic one. We vary λ for two ambiguous queries and measure the coverage capability of all query aspects in the top-n results using S-Rec@n. Four values of λ are tested as illustrated in Fig. 10:

- $\lambda = 0$: corresponds to semantic diversity (SemanticDiv)
- $\lambda = 1$: corresponds to visual diversity (VisualDiv)
- $\lambda = 0.3$: corresponds to giving more importance to the semantic diversity than the visual diversity (0.3V + 0.7S)
- $\lambda = 0.7$: corresponds to giving more importance to the visual diversity than the semantic diversity (0.3S + 0.7V).

The obtained results show that the optimal value of λ is 0.7. This reveals that the semantic diversity and the visual diversity are unequally harmonized, which confirms the influence of each modality on diversity ranking.

4.3.2 The trade-off between diversity and relevance

In this experiment, we aim to study, mainly, whether applying an adaptive trade-off for queries having different levels of ambiguity improves performance. To investigate this, we compare our approach with different existing approaches:

- relevance-based ranking [26] (RR)
- Diverse Relevance ranking [26] (DRR)
- Adaptive Diverse Relevance Ranking(ADRR)
- Flickr (relevance).

The comparison of ADP measurements between these different approaches is shown in Fig. 12. From this figure, the

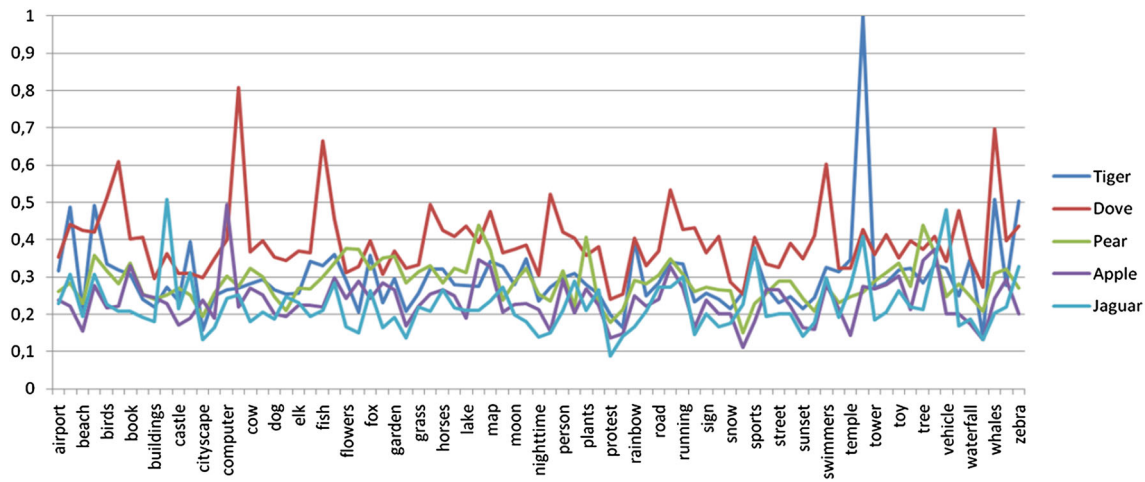


Fig. 8 Weights of concept estimation over Flickr resources

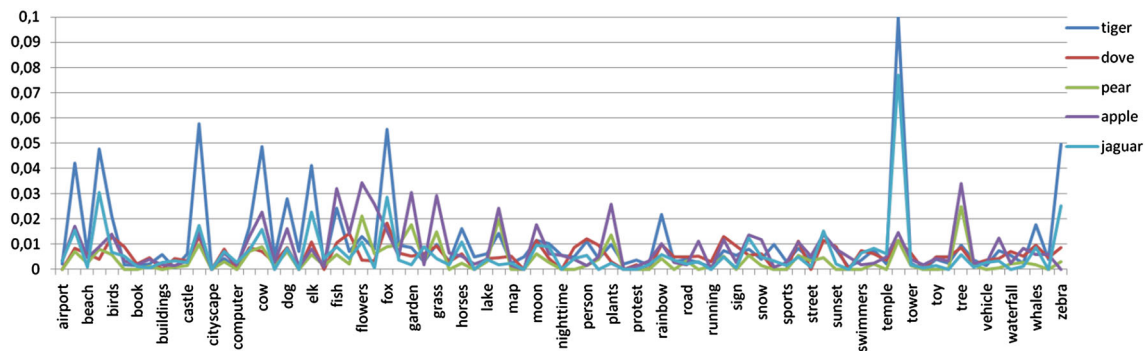


Fig. 9 Weights of concept estimation over Wikipedia resources

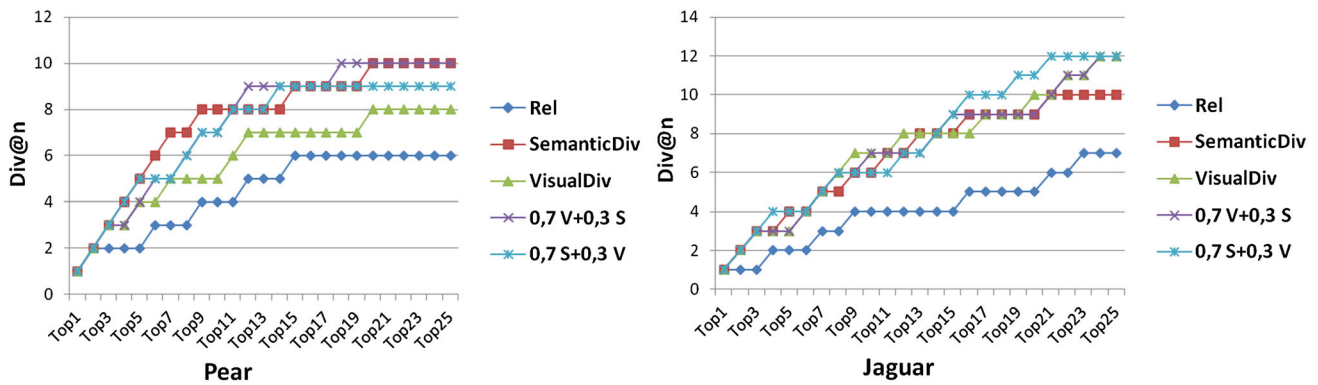


Fig. 10 Evolution curve of S-Rec@n with different values of λ for query ‘Pear’ and ‘jaguar’

conclusion that can be drawn is that the proposed ranking scheme using relevance and diversity criteria outperforms those using only relevance criteria. Furthermore, the experimental results indicate that the performance of our ranking process using ambiguity indicator is superior to that using a fixed ambiguity indicator. In order to better evaluate our approach, we select two queries with different levels of ambi-

guity and we look for the best diversity tuning parameter value for each one to balance diversity and relevance scores; then, we compare it with our AmbIDC indicator value associated to each query. The obtained results are illustrated in Fig. 11. For a highly ambiguous query such as ‘Jaguar’, we find that the best value of Div@n is at 0.75 which corresponds to the value of AmbIDC. In contrast, for a low ambiguous

Fig. 11 Evolution curve of Div@n with different values of diversity tuning parameter for high-level ambiguous queries ‘Jaguar’ (AmbIDC = 0.75) and ‘dove’ (AmbIDC = 0.5)

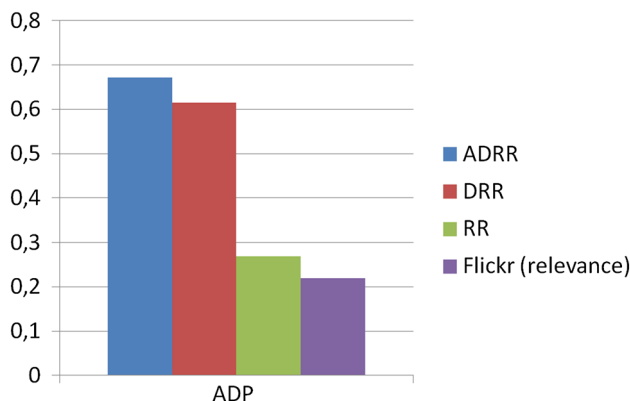
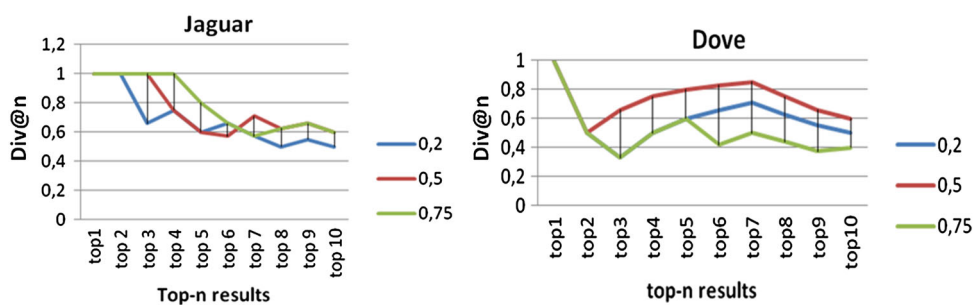


Fig. 12 ADP measurements of 12 queries for different approaches: “RR”, “DRR”, “ADRR”, “Flickr (relevance)”

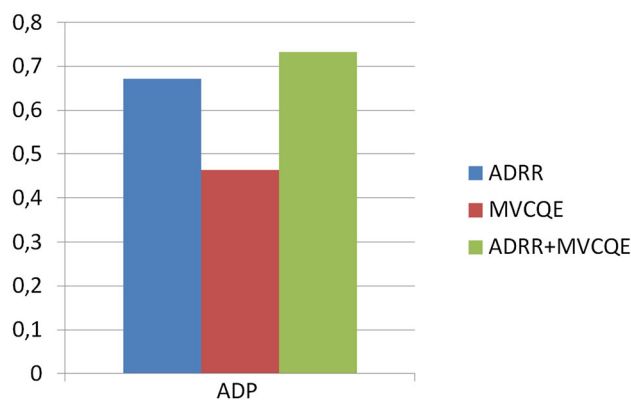


Fig. 13 ADP measurements of 12 queries for different ranking schemes: “MVCQE”, “ADRR”, “MVCQE + ADRR”

query such as ‘Dove’, we find that the best value of Div@n is at 0.5 which is the value of AmbIDC. As a conclusion, we can confirm the effectiveness of our approach by automatically balancing diversity and relevance ranking according to the level of ambiguity for a given query.

4.4 Study of the overall retrieval process

In this experiment, we study the impact of fusing “MVCQE” and “ADRR” processes in the retrieval system. The following figure shows the ADP values of different ranking schemes: “MVCQE”, “ADRR”, “MVCQE+ADRR”.

As expected, “MVCQE+ADRR” fusion outperforms “MVCQE” and “ADRR”. In fact, “MVCQE+ADRR” fusion controls the diversity at two levels: pre-processing and post-processing (Fig. 13).

4.5 Conclusion

In this paper, we have presented an approach to diversify relevant search results for ambiguous queries in tag-based social image retrieval. As not all queries are equally ambiguous and different strategies should be suggested, our approach has the criteria to adapt with the level of ambiguity to control how much we need to diversify relevant results.

Since existing social image retrieval systems suffer from a tag mismatch problem, top- k results are noisy. Since successful diversity requires relevant results, our first concern was to ensure relevance criteria. Thus, we have adapted the representation of semantic concepts to avoid tag-mismatch problem by providing a standardized representation for images and queries using vectors of concepts weights.

The key advantage of our approach is its ability to make effective use of the semantic concepts representation not only to ensure relevance but also to diversify results. In fact, the experiments have demonstrated that concept representation can guarantee the complete coverage of all query aspects. Indeed, the set of concept used can provide a wide coverage of human knowledge. We demonstrated also that in order to ensure diversity for ambiguous queries, we need to jointly apply the query expansion and the search result diversification processes. In addition, we have investigated the need to control the diversity per query. So, we have proposed an adaptive trade-off that balances the relevance and diversity scores with respect to the level of query ambiguity.

Our future work will involve studying a further refinement of the proposed system. More specifically, we are interested in enhancing our MVCQE approach by applying the Maximal Marginal Relevance algorithm in the aggregated concepts. By doing this, we attempt to further promote diversity. The rationale behind this idea is that the top selected

concepts are semantically redundant. Moreover, we are interested in testing our system over large-scale social media resources. In this case, we will need to adapt our system to a hyper-graph representation to further diversify results.

References

- Abbasi R (2011) Query expansion in folksonomies. In: Declerck T, Granitzer M, Grzegorzec M, Romanelli M, Rger S, Sintek M (eds) *Semantic multimedia. Lecture notes in computer science*, vol 6725. Springer, Berlin, pp 1–16
- Arampatzis A, Kamps J, Robertson S (2009) Where to stop reading a ranked list? Threshold optimization using truncated score distributions. In: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '09*. ACM, New York, pp 524–531. doi:[10.1145/1571941.1572031](https://doi.org/10.1145/1571941.1572031)
- Carbonell J, Goldstein J (1998) The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '98*. ACM, New York. doi:[10.1145/290941.291025](https://doi.org/10.1145/290941.291025)
- Fakhfakh R, Ksibi A, Ben Ammar A, Ben Amar C (2013) Enhancing query interpretation by combining textual and visual analyses. In: *Advanced Logistics and Transport (ICALT), 2013 International Conference on*, IEEE, pp 170–175
- Feki G, Ksibi A, Ammar AB, mar CB (2012) Regimvid at imageclef2012: Improving diversity in personal photo ranking using fuzzy logic. In: *ImageCLEF12*, pp -1-1
- Feki G, Ksibi A, Ben Ammar A, Ben Amar C (2013) Improving image search effectiveness by integrating contextual information. In: *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on*, IEEE, pp 149–154
- Hoque E, Hoerber O, Gong M (2011) Evaluating the trade-offs between diversity and precision for web image search using concept-based query expansion. In: *WI-IAT 2011*, vol 3, pp 130–133
- Hoque E, Hoerber O, Gong M (2012) Balancing the trade-offs between diversity and precision for web image search using concept-based query expansion. *J Emerg Technol Web Intell* 4(1):26–34
- Jin S, Lin H, Su S (2009) Query expansion based on folksonomy tag co-occurrence analysis. In: *GrC*, pp 300–305
- Jing Y, Rowley HA, Wang J, Tsai D, Rosenberg C, Covell M (2012) Google image swirl: a large-scale content-based image visualization system. In: *WWW '2012*, ACM, New York, pp 539–540
- Kato M, Ohshima H, Oyama S, Tanaka K (2008) Can social tagging improve web image search? In: *Web Information Systems Engineering-WISE 2008*. Springer, Berlin, pp 235–249
- Ksibi A, Ammar AB, Amar CB (2012) Effective concept detection using second order co-occurrence flickr context similarity measure socfcs. In: *CBMI*, pp 1–6
- Ksibi A, Ben Ammar A, Ben Amar C (2013) Enhanced context-based query-to-concept mapping in social image retrieval. In: *Content-based multimedia indexing (CBMI), 2013 11th International Workshop on*, IEEE, pp 85–89
- Ksibi A, Dammak M, Ben Ammar A, Mejdoub M, Ben Amar C (2012) Flickr-based semantic context to refine automatic photo annotation. In: *Image processing theory, tools and applications (IPTA), 2012 3rd International Conference on*, IEEE, pp 377–382
- Ksibi A, Feki G, Ammar AB, Amar CB (2013) Effective diversification for ambiguous queries in social image retrieval. *Computer analysis of images and patterns*. Springer, Berlin, pp 571–578
- van Leuken RH, Garcia L, Olivares X, van Zwol R (2009) Visual diversification of image search results. *WWW '09*. ACM, New York, pp 341–350
- Mejdoub M, Ben Amar C (2013) Classification improvement of local feature vectors over the knn algorithm. *Multimedia Tools Appl* 64(1): 197–218. doi:[10.1007/s11042-011-0900-4](https://doi.org/10.1007/s11042-011-0900-4)
- Myoupo D, Popescu A, Le Borgne H, Moëllic PA (2010) Multimodal image retrieval over a large database. In: *Proceedings of the 10th international conference on Cross-language evaluation forum: multimedia experiments, CLEF'09*. Springer, Berlin, pp 177–184. <http://dl.acm.org/citation.cfm?id=1885110.1885135>
- Radlinski F, Bennett PN, Carterette B, Joachims T (2009) Redundancy, diversity and interdependent document relevance. *SIGIR Forum* 43(2): 46–52. doi:[10.1145/1670564.1670572](https://doi.org/10.1145/1670564.1670572)
- Santos RL, Macdonald C, Ounis I (2010) Exploiting query reformulations for web search result diversification. In: *Proceedings of the 19th international conference on World wide web, WWW '10*. ACM, New York, pp 881–890. doi:[10.1145/1772690.1772780](https://doi.org/10.1145/1772690.1772780)
- Sun A, Bhowmick SS, Nguyen KTN, Bai G (2011) Tag-based social image retrieval: An empirical evaluation. *JASIST* 62(12):2364–2381
- Tang X, Liu K, Cui J, Wen F, Wang X (2012) Intentsearch: capturing user intention for one-click internet image search. *IEEE Trans Pattern Anal Mach Intell* 34(7):1342–1353. <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.242>
- Wang S, Jing F, He J, Du Q, Zhang L (2007) Igroup: presenting web image search results in semantic clusters. In: *CHI '07*. ACM, New York, pp. 587–596. doi:[10.1145/1240624.1240718](https://doi.org/10.1145/1240624.1240718)
- Wei XY, Ngo CW, Jiang YG (2008) Selection of concept detectors for video search by ontology-enriched semantic spaces. *Multimedia IEEE Trans* 10(6):1085–1096. doi:[10.1109/TMM.2008.2001382](https://doi.org/10.1109/TMM.2008.2001382)
- Weinberger KQ, Slaney M, Van Zwol R (2008) Resolving tag ambiguity. In: *Proceedings of the 16th ACM international conference on Multimedia, MM '08*. ACM, New York, pp 111–120. doi:[10.1145/1459359.1459375](https://doi.org/10.1145/1459359.1459375)
- Yang K, Wang M, Hua XS, Zhang HJ (2010) Social image search with diverse relevance ranking. *MMM'10*. Springer, Berlin, pp 174–184