

Article

Supporting Keyword Search for Image Retrieval with Integration of Probabilistic Annotation

Tie Hua Zhou ¹, Ling Wang ^{2,*} and Keun Ho Ryu ^{1,*}

¹ Database/Bioinformatics Laboratory, School of Electrical & Computer Engineering, Chungbuk National University, Cheongju 362-763, Korea; E-Mail: thzhou@dblab.chungbuk.ac.kr

² Department of Information Engineering, Northeast Dianli University, Jilin 132000, China

* Authors to whom correspondence should be addressed;

E-Mails: smile2867ling@gmail.com (L.W.); khryu@dblab.chungbuk.ac.kr (K.H.R.);

Tel.: +86-156-88952200 (L.W.); +82-43-261-2254 (K.H.R.); Fax: +86-432-62493676 (L.W.); +82-43-275-2254 (K.H.R).

Academic Editor: Jason C. Hung

Received: 10 February 2015 / Accepted: 11 May 2015 / Published: 22 May 2015

Abstract: The ever-increasing quantities of digital photo resources are annotated with enriching vocabularies to form semantic annotations. Photo-sharing social networks have boosted the need for efficient and intuitive querying to respond to user requirements in large-scale image collections. In order to help users formulate efficient and effective image retrieval, we present a novel integration of a probabilistic model based on keyword query architecture that models the probability distribution of image annotations: allowing users to obtain satisfactory results from image retrieval via the integration of multiple annotations. We focus on the annotation integration step in order to specify the meaning of each image annotation, thus leading to the most representative annotations of the intent of a keyword search. For this demonstration, we show how a probabilistic model has been integrated to semantic annotations to allow users to intuitively define explicit and precise keyword queries in order to retrieve satisfactory image results distributed in heterogeneous large data sources. Our experiments on SBU (collected by Stony Brook University) database show that (i) our integrated annotation contains higher quality representatives and semantic matches; and (ii) the results indicating annotation integration can indeed improve image search result quality.

Keywords: multi-label image; image annotation; annotation integration; semantic matching; keyword search; image retrieval

1. Introduction

In recent years, large repositories of digital photos, such as social networks including Facebook, Google, Instagram and Flickr *etc.*, have become valuable resources for image retrieval technologies, especially for enriching documents with semantic annotations—annotations that label photo snippets as referring to either certain entities or to elements of particular semantic categories. Thus, more and more images with their descriptions are generated, acquired, distributed, analyzed, classified, stored, and made accessible via the interactive user interfaces. Unfortunately, a recent study reported in [1,2] shows that the user-provided annotations associated with images are rather imprecise, with only about a 50% precision rate. On the other hand, the average number of annotations for each image is relatively low [3], which is far from the number that can fully describe the contents of an image. For example, to see how far we have come, just try searching through your own images on Google or Flickr. Without accurate and sufficient annotations, images on the Web cannot be well indexed via search engines and consequently, cannot be easily accessed by the users.

Social networks have boosted the need for efficient and intuitive queries to access large scale image repositories. Therefore, efficient and effective approaches to retrieve images from large repositories have become an extremely difficult problem for research and development. With the rapid development of computer visioning and information technology, the image retrieval community has already developed advanced techniques for effective searches. Text-based and content-based approaches are the most common approaches used for image retrieval. Text-based methods match keywords to retrieve images or image sets. Further, content-based methods retrieve images and image sets by visual characteristics such as color, texture, and shape. Rich textual features are utilized as annotations for Web images due to the high variance of image content that combines visual features and high-level features.

Keyword-based image indexing is defined as the representation of images through the textual descriptions of their contents: what they are of or about, subject headings, presenter, date of acquisition, Uniform Resource Locator (URL), and the accompanying text as the annotations are integral parts of text-based image indexing and retrieval. Consider the real-life image in Flickr displayed in Figure 1a; all these textual features provide an access path to images and have proved to be very helpful in many commercial Web image search-engines. As introduced in [4,5], image annotation, also known as image tagging, is a process by which labels or tags are manually, automatically, or semi-automatically associated with the images. Keywords search over annotations, but direct application of these solutions to relational image sources, where the image is typically labeled in multiple vocabularies, is neither efficient nor effective. Such Web services not only allow individuals to upload and share images, but they also allow individuals to collaboratively describe the resources with their own tags (or annotations) via tagging services. While originally focused on a few standard classes of annotations, the ecosystem of annotators is now becoming increasingly diverse. Currently, modern annotators often have very different vocabularies, with both common-knowledge and specialist concepts; they also have many semantic interconnections. As illustrated in Figure 1b, the same vocabulary may describe different content (*i.e.*,

objects, scenes, attributes, actions). Annotations are typically attached to tuples and represent metadata such as probability, multiplicity, comments, or provenance. However, the annotations provided are often noisy and unreliable (*i.e.*, birds in Figure 1a). Thus, effective methods to refine these unreliable annotations are needed. It also creates challenging problems, including estimating the quality of annotations and reconciling disagreeing semantic annotations for an image.

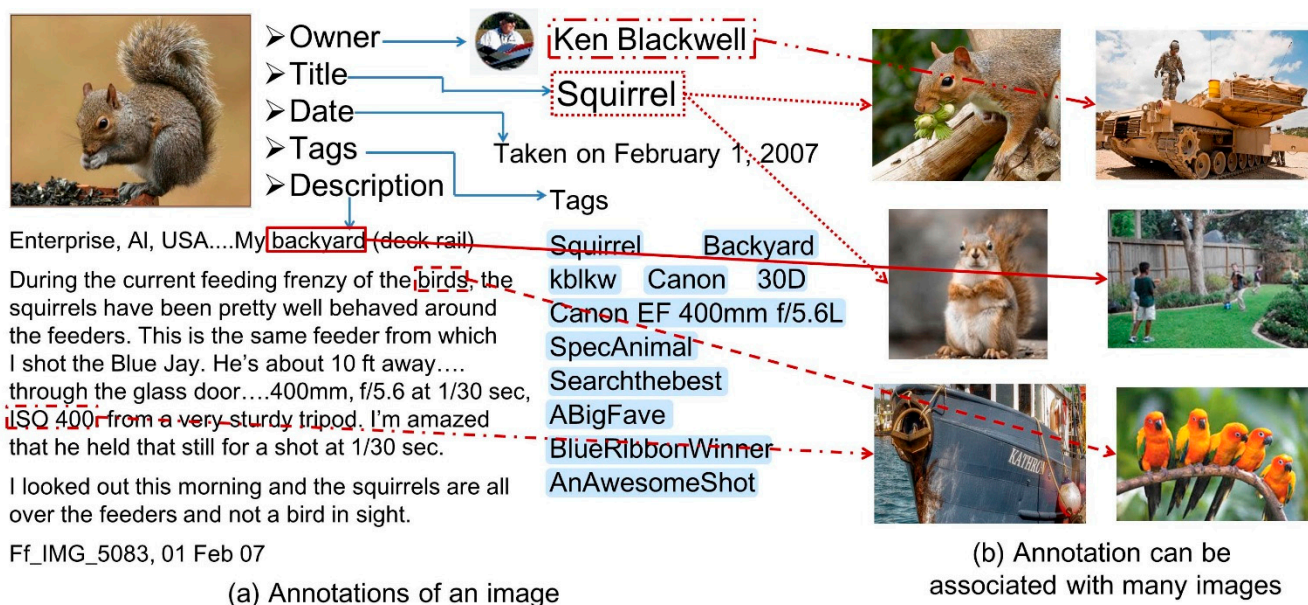


Figure 1. A real-life image in Flickr.

The main inspiration of our work comes from recent advances in rewriting semantic annotations for keyword-based queries [6–8]. In this work, we propose integration of probabilistic model based on keyword query architecture that integrates annotations of images. Our approach is motivated by the need for managing and integrating semantic annotations: allowing users to see the satisfactory results of image retrieval via the integration of multiple annotations. We focus on annotation integration step in order to specify the meaning of each image annotation, thus leading to the most representative annotations of the intent of a keyword search. We present a Query-Frequency Pair (QF-P) algorithm that detects annotations based on their interrelation and diversity for advanced querying. The main goal of our work is to develop an advisable model for image retrieval that can be used in all keyword-based search engines to optimize result quality. Demonstration are: (i) the combination of two perspectives, one taking into account the “user” perspective, and other considering the “database” perspective; (ii) the flexibility of the approach to adapt to different working conditions; and (iii) the ability to query fully accessible databases and databases which provide a reduced search. Towards a refined solution that enables more effective and efficient keyword search, we provide the following contributions of this demonstration: (1) we use the Bayes rule map the image annotations into metric space of probability distribution; (2) the definition of a refining-based formalism to connected keyword queries; (3) proposed QF-P algorithm defines a sophisticated verification process to avoid the matching of irrelevant annotations; and (4) the development of an efficient and suitable search model to help users make the most of this novel keyword query algorithm.

Organization: Section 2 reviews the popular retrieval algorithms and several related works. Section 3 describes our proposed framework and presents the QF-P algorithm for integrating annotation; also, we

discuss the Geometric-Center approach for multi-keywords query. Section 4 discusses the implementation of the algorithms, and gives an experimental evaluation. Section 5 gives conclusions and discusses prospects for future improvements.

2. Related Work

The problem of image retrieval has attracted much research interest over the last few years, with the goal to search for similar images in huge image collections via queries. The Content-Based Image Retrieval (CBIR) system is an attractive research area that was proposed to allow users to retrieve relevant images in an effective and efficient pattern. In [9], the authors proposed a CBIR system that evaluates the similarity of each image in its data store to a query image in terms of color and textural characteristics, and it then returns the images within a desired range of similarity. However, the retrieval results of CBIR accomplished by studies thus far are unsatisfactory.

Keyword-based search solutions have been proposed for dealing with different kinds of data. Existing work so far focuses on the efficient processing of keyword queries or effective ranking of results. Searching results on image data requires finding matches for the keyword as well as considering annotations in the data connecting them, which represent final answers covering all query keyword(s). The forward module implements the method described in [10] for discovering the top-k configurations associated with the user keyword queries. The process is modeled by means of a HMM that contains a state for each database element. Therefore, by modeling in this way the search process, the emission probability distribution describes the likelihood for a keyword to be “generated” by a specific state, while the transition probability distribution describes the likelihood for two keywords to be associated with adjacent states. The emission probabilities are computed for each keyword and for each database attribute by applying the search function over full text indexes provided by the DBMS. In [11], we studied how to optimize search of structured product entities (represented by specifications) with keyword queries such as “cheap gaming laptop”. They propose a novel probabilistic entity retrieval model based on query generation, where the entities would be ranked for a given keyword query based on the likelihood that a user who likes an entity would pose the query.

In many cases, the task of extracting relationships from large textual collection, such as the Web, has been explored in different ways. The word-to-word correlation model [12] is explored to refine more candidate annotations. The image annotation solved two problems: (1) find one accurate keyword for a query image (consider again the Figure 1); (2) given one keyword, and find complementary annotations to describe the details of this image. The keyword join approach (KJ) [13] materializes paths in the index and only joins them online. KJ was shown to be faster than BDS but also employs a larger index. More powerful annotation model based on language parsing have been used as well [14,15]. These approaches have been able to describe images “in the wild”, but they are heavily hand-designed and rigid when it comes to text generation.

Recent studies that have attempted to use multi-label learning algorithms to solve the image annotation problem focus on mining the label relationship to achieve better annotation performance. In [16], the authors propose an image retagging scheme that is formulated as a multiple graph-based multi-label learning problem, which simultaneously explores the visual content of the images, semantic correlation of the tags as well as the prior information provided by users. Different from the classical single graph-based multi-label learning algorithms, the proposed algorithm propagates the information

of each tag along an individual tag-specific similarity graph, which reflects the particular relationship among the images with respect to the specific tag, and the propagations of different tags simultaneously interact with each other in a collaborative way with an extra tag similarity graph. In [17], AGGREGO search is present to offer a novel keyword-based query solution for end users in order to retrieve precise answers from semantic data sources. AGGREGO search suggests grammatical connectors from natural languages during the query formulation step in order to specify the meaning of each keyword, thus leading to a complete and explicit definition of the intent of the search. AGGREGO search is half-way between natural language queries and classical keyword queries. An AGGREGO search query is composed of keywords referencing classes and instances of an ontology but also includes connectors referencing properties that clearly express how these concepts have to be linked. A large body of work has addressed the problem of ranking annotations for a given image [18,19]. They are based on the idea of co-embedding of images and annotations in the same vector space. For an image query, annotations are retrieved which lie close to the image in the embedding space. Most closely, neural networks are used to co-embedding images and annotations together [20] or even image crops and sub-descriptions [21] but do not attempt to generate novel annotations. Kiros *et al.* [22] use a neural net, but a feedforward one, to predict the next annotation given the image and previous annotations. A recent work by Mao *et al.* [23] uses a recurrent neural network for the same prediction tasks, which randomly initialize word embedding layers and learn them from the training data.

3. Model Architecture

3.1. Image Retrieval Architecture

CBIR has attracted more studies in the field of informatics. CBIR facilitates techniques for effective indexing and retrieval of images by features, such as color, texture, shape, movement, and contours. We propose a content-based image retrieval framework, as shown in Figure 2.

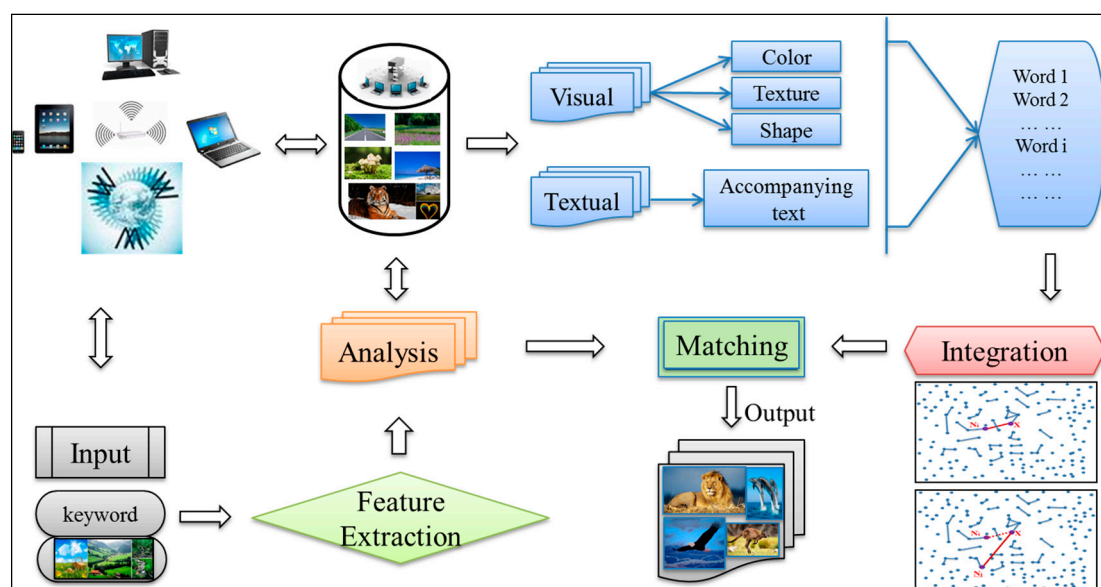


Figure 2. Proposed content-based image retrieval framework.

Figure 2 shows a typical prototype of the content-based image retrieval architecture. In this scenario, users interact with the system to retrieve images that meet their queries. In the indexing phase, the image database is organized and managed. In the retrieval phase, a way for interaction between user and system is provided. This scenario combines the text-based and content-based approaches that consist of two major aspects, visual feature extraction and multi-label indexing engine. The user interface supplies a basic textbox for textual image query, and typically provides an image browser interface for non-textual image query (image). This system builds a text or visual index to search among image collections. Each image in the database can be represented by the association with visual features and semantic annotations in a multi-modal label. First, feature extraction uses principal component analysis to select the most appropriate visual features for prediction and removes the irrelevant and redundant features. Then, each image will be assigned with one or multiple annotations from a predefined label set.

Annotations play an important role in search engines, information extraction, and for the auto-production of linked data. More relevant annotations would allow users to further exploit the indexing and retrieval architecture of image search-engines to improve image search. Therefore, one of the next research challenges deals with the annotation integration. It requires a great deal of intelligent processing of this data because high quality data is mixed up with low quality noisy text. However, the main problem of annotation is users' subjective preferences, which means that a different user may have different perceptions of the same image. Therefore, to effectively and correctly find user desired images from large-scale image collections is not an easy task. The major problem is that the images are usually not annotated using semantic descriptors. Because image data is typically associated with multi-labels, each image could contain various objects and, therefore, could be associated with a set of labels. By exploiting a large collection of image annotations, the basic idea is to identify annotations which are frequently detected together and to derive their type(s) of relationship based on the semantic correlation. The verification step includes machine learning techniques to compute the type of relationship between two annotations.

3.2. Problem

The problem of image annotation has gained interest more recently. Realistically, more and more words have been annotated to images, and irrelevant images have been connected to each other. Figure 3 clearly shows the image-to-image and word-to-word correlation model in the real-world dataset Flickr. Images were connected to each other via to the interrelated and diverse annotations (the leftmost figure). The rightmost figure shows an example of annotations; the nearest annotations can be found in the learned embedding space. Indeed, an image belonging to the category "nature" is more likely annotated with words "sky" and "trees" than words "car" and "vehicle". Subsets labels are correlated, as many annotations are semantically close to each other. Therefore, this correlation among subsets can help predict the keywords of query examples. The main problems of multi-label indexing consist of two components. First, it is not clear how much annotation is sufficient for describing the image. Second, it is not clear what the best subset of objects to annotate is. Unlike spoken language and text documents, the semantic annotations of images are usually unavailable. Thus, we focus on how to integrate annotation in order to specify the meaning of each image annotation, thus leading to the most representative annotations of the intent of a keyword search. The goal of our approach is to translate a user-defined keyword query into refined annotations. Indeed, an annotation must capture not only the objects contained in an image, but it also must express how these images relate to each other as well as the

annotations they are involved in (Figure 3). This task is significantly harder, for example, than the well-studied image classification or image objects recognition tasks, which have been a main focus in the computer vision community [24]. Annotations embed into a dense word representation, the semantically relevant annotations can be found by calculating the Euclidean distance between two dense words in metric space.

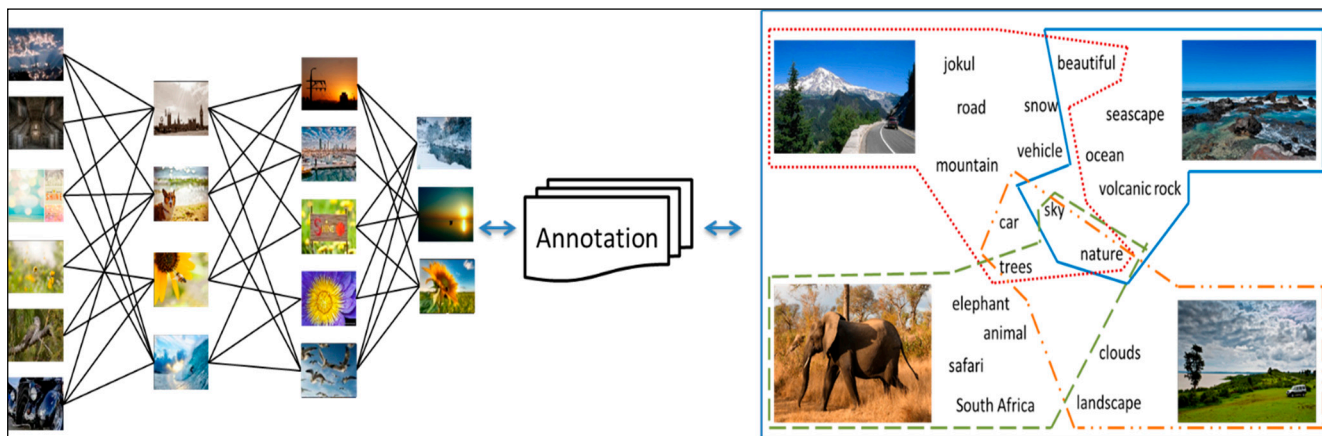


Figure 3. Example of correlative images.

In this paper, we propose an integration of probabilistic model to refine annotation A_R^* from images. Recent advances in statistical machine translation have shown that, given a powerful sequence model, it is possible to achieve high-quality results by directly maximizing the probability of the correct translation given an input keyword in an “end-to-end” fashion. We propose to directly maximize the probability of the correct annotation given the image by using the following formulation:

$$k^* = \arg \max_k \sum_{(g, A_R^*)} \log p(A_R^* | g, A_s) \tag{1}$$

where k is the parameters of our model, g is an image, and A_s is the existing annotations.

3.3. Probabilistic Model

DEFINITION 1. Image database are collected as a directed form $D(G, C, A_s)$. Each image g is assigned to image category $C = \{c_1, c_2, \dots, c_n\}$ according to its characteristics. Given a set of images $G = \{g_1, g_2, \dots, g_n\}$ and the relations Ω between them, each image g in G is characterized by its annotations A_s . For $\forall g$, let $A_s = \{a_1, a_2, \dots, a_n\}$ denote a collection of annotations mentions in g . For each c_i , the order of the frequency annotations for A_s to be defined as $A_F = \{f_0, f_1, \dots, f_m\}$. Each image query q is characterized by its annotation Q . The objective of our probabilistic model is to determine the referent images in G of the annotations in A_s .

To capture the relationships in the annotation, we will make use of Ω which states the relationships between categories of images. The most common constraints supported for concepts are:

- ✧ Jointness constraints $c_m \wedge c_n \subseteq \oplus$, stating that g is an image of category c_m and also an image of category c_n .
- ✧ Disjointness constraints $c_m \wedge c_n \subseteq \emptyset$, stating that category c_m and c_n have an empty intersection.

The above constraints are supported by the relationships of annotations. We exploit the statistics of the extracted probabilistic model groundings. Based on that, we define the probability $P(a|c)$ that category c in G refers to image annotation a

$$P(a|c) = \frac{\text{count}_{\text{correlation}}(a,c)}{\sum_{a_i \in c} \text{count}_{\text{correlation}}(a_i,c)} \quad (2)$$

where $\text{count}_{\text{correlation}}(a, c)$ denote the number of correlation using c as primary unit pointing to a as destination and c is the set of images that have the correlative images.

In addition, we define the probability $P(a_i \in A)$ that the annotation a_i in an image is an annotation name as

$$P(a_i \in A) = \frac{\text{count}_{\text{correlation}}(a_i)}{\text{count}_{\text{total}}(a_i)} \quad (3)$$

where $\text{count}_{\text{correlation}}(a_i)$ denotes the number of annotations that is assigned to c and $\text{count}_{\text{total}}(a_i)$ denotes the number of annotations where appears in G .

Subsequently, we map A_s into the metric space to harvest expressive words from existing image annotations. We would like to present model that start from the keyword Q of image as a input query, and is trained to maximize the likelihood of producing a target sequence of annotations $A^*_R = \{a^*_0, \dots, a^*_m\}$, where denote by a^*_0 a special start annotation (the most frequently word) and by a^*_m a special stop annotation, and each annotation a^*_i come from a given dictionary, that describes the image adequately. We model the relevance of an annotation a which serves as a keyword with conditional probability $P(c_i|a)$ and $P(f_i|a)$, which can be interpreted as the probability of that keyword search by the user after we observe the user's query q . With Bayes rule, we have:

$$P(c_i | a) = \frac{P(c_i) \cdot P(a | c_i)}{\sum_{j=1}^n P(c_j) \cdot P(a | c_j)} \quad (4)$$

$$P(f_i | a) = \frac{P(f_i) \cdot P(a | f_i)}{\sum_{j=1}^n P(f_j) \cdot P(a | f_j)} \quad (5)$$

The functions only depend on two component probabilities. The first is $P(c_i)$, which is the likelihood that we would observe query q if g is indeed relevant. The second is $P(f_i)$, which denotes the probability density of annotation. This conditional probability can capture how well image g matches the keyword in query q that if user desires image g , the user would likely search a query matching the annotations of image g . $P(f_i|a)$ is the probability that the user who desires image g would include a preference for the keyword query. In the proposed model, relevance is primarily based on the probability that the user interested in an image will search using the query. The model attempts to simulate the query formulation process of a user and recognize annotations for the keyword of a query.

We set these two probabilities as abscissa and ordinate, each $a \in A_s$ can be mapped into a point $p(P(c_i|a), P(f_i|a))$ in metric space. Then, we present the Query-Frequency Pair (QF-P) algorithm to integrate the annotations. Notice that in the acronym, “ Q ” is the keyword of query, “ F ” is the most frequently annotation F_A in c_i which “ Q ” occurred most frequently. The distance d between Q and F_A is used to detect the relevant subset S for integrating annotation. For $\forall a_i$,

$$\begin{cases} a_i \in S, & \text{if } d(a_i, Q) \leq d \text{ or } d(a_i, F_A) \leq d \\ a_i \notin S, & \text{otherwise} \end{cases} \quad (6)$$

We can give a certain threshold value (as the number of points) of parameter k to limit the number of object (point) in S . If the number of subset k' is smaller than k , selecting the f_0' in S as the new F_A to further detect relevant subset. When the $k' \geq k$, $S = \{f_0^*, f_1^*, \dots, f_k^*\}$. There is a special subset S_s in which the annotations a_i satisfied $d(a_i, Q) \leq d$ and $d(a_i, F_A) \leq d$. Finally, we can use relevant subset to integrate annotation for image retrieval.

For a new test instance, the pre-processing first using query to find F_A and then makes the QF-P algorithm to detect relevant subsets for integrating annotation. For the retrieval task, the index only depends on integrated annotation A^*_R . In testing, the general trend is that the most relevant annotations exist in S_s . Assume that k is big enough for a large scale database: we propose a hierarchical design for integrated annotation index structure (Figure 4).

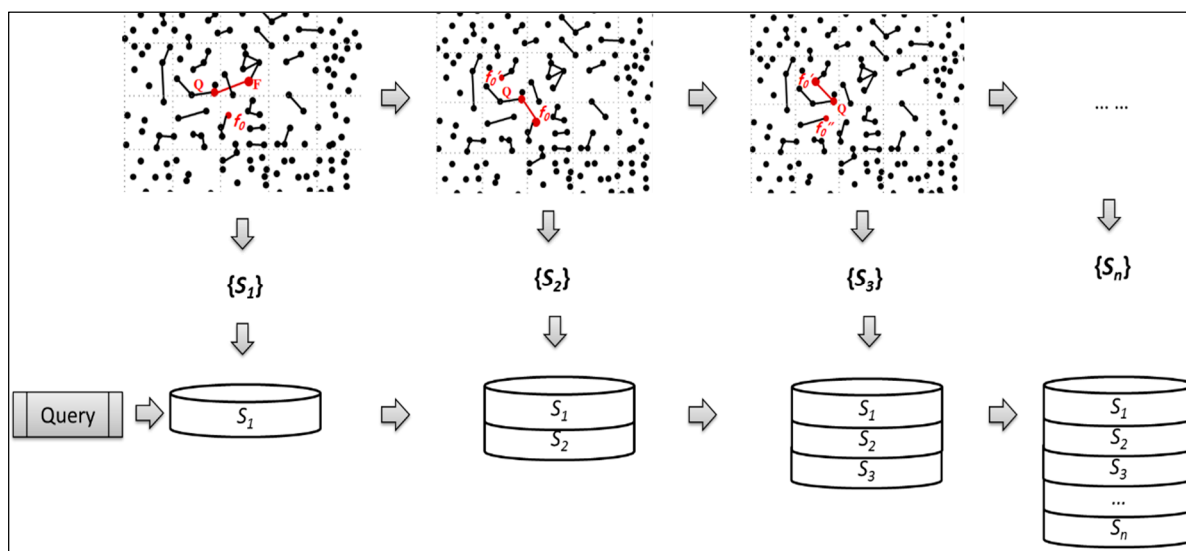


Figure 4. Hierarchical design for integration.

3.4. Query-Frequency Pair Algorithm

The detailed QF-P algorithm is shown in Figure 5. QF-P algorithm is derived via the traditional k -nearest neighbor algorithm. The input to the procedure is $A_s = \{a_1, a_2, \dots, a_n\}$. Also, $Q \in A_s, F_A \in A_s$. The pair (Q, F_A) provides distance to limit detection range of relevant subset S . The given threshold k is used to limit the maximum number of words in S . In fact, our proposed algorithm implied two major meanings; “hierarchical” means use of hierarchical model to classify the original features in order to express different degrees of correlations. If the number of words in S is $k' < k$, then we chose the most frequent word f_0' ($f_0' \neq F_A$) as new F_A to construct a new pair (Q, f_0') , and then, repeated the detection procedure. Otherwise, the relevant subset $S = \{f_0^*, f_1^*, \dots, f_k^*\}$ is a set sorted by frequency. This algorithm can find the best combination between the visual and textual features, as the annotations, and it exploits the high-quality representative correlation between annotations and the image.

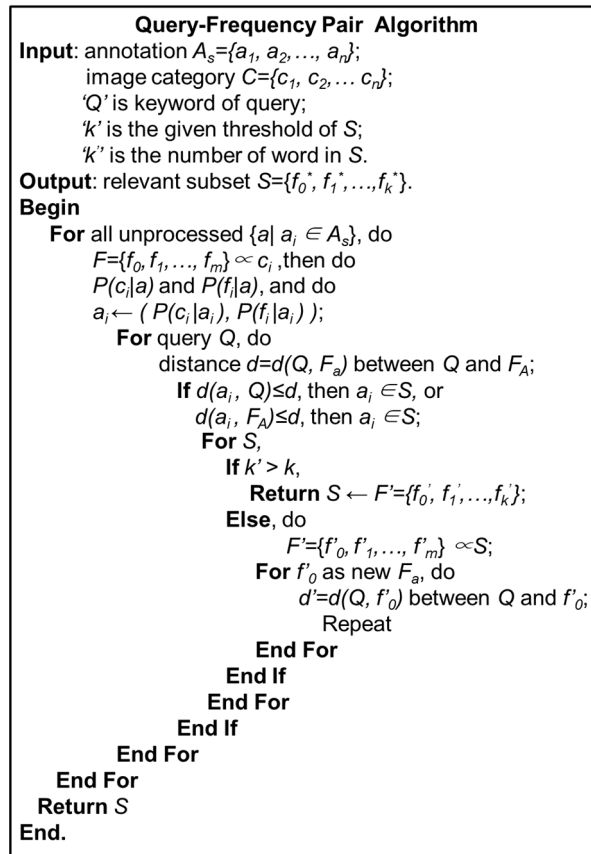


Figure 5. Query-Frequency Pair algorithm.

Actually, the QF-P algorithm is also a greedy algorithm that is derived via the traditional *k*-nearest neighbors' algorithm. We use a tree structure to present this method, and the tree would continue to grow until a threshold *k* is satisfied (see Figure 6). The more correlative annotations will have higher probabilities detected in the *S*. According to different degrees of a relevant subset, we establish hierarchical levels to classify the subset for retrieving the information from the database. In reference to the relevant subset *S*₁, for the matching annotations, we index the matching images. Otherwise, go to the next *S*₂, *S*₃ continuously. It is an efficient way to improve the retrieval speed and ensure the accuracy of the results. For the retrieval task (as the algorithm shown in Figure 7), we rank the images based on prioritized recall with the query keyword and output the top ranked ones.

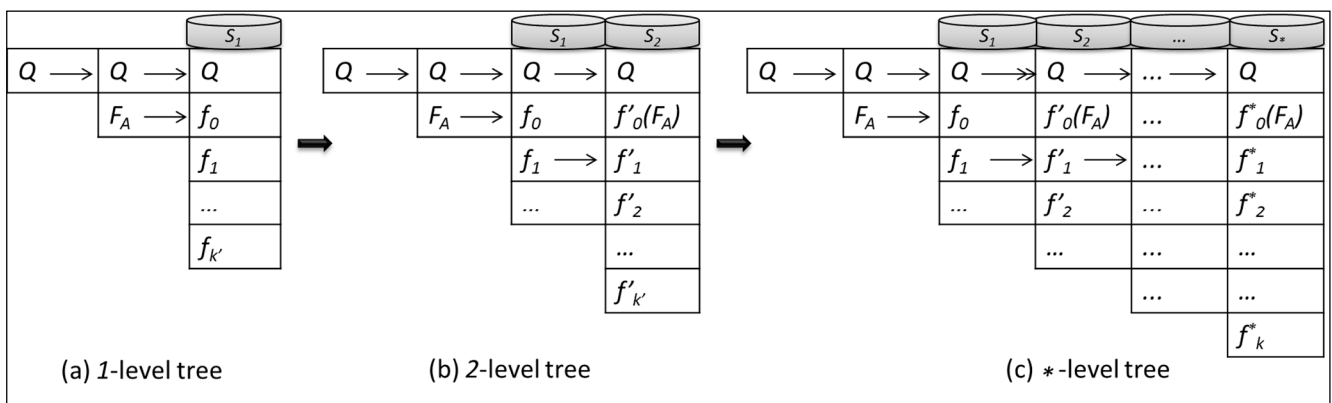


Figure 6. The tree-growth of QF-P algorithm.

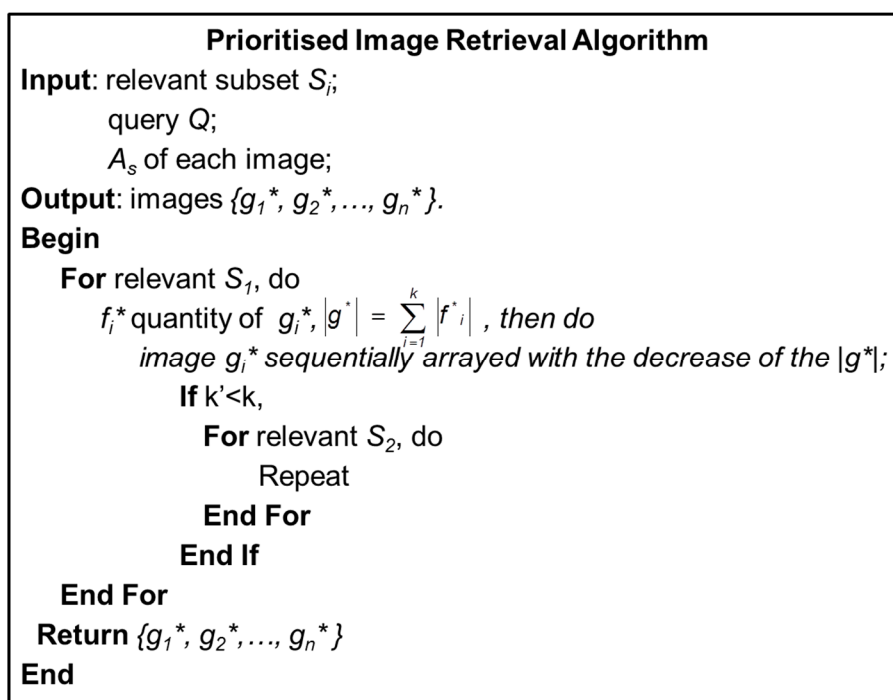


Figure 7. The prioritized retrieval algorithm.

3.5. Geometric-Center Approach

Generally, if you try to use multiple hot keywords as a text entry to find related images, it is possible to refine using keyword annotations. Therefore, we propose another method of automatically assigning relevant multi-keywords to a user specified image which could greatly improve the retrieval accuracy, and fast response is also required. “Multi-keywords” in a text search involves some information by uploaders being manually supported and, in visual searches, some hidden image information under the conditions of invisibility is also considered as a multi-keywords problem. So, the multi-keywords problem is common to both text and visual searches.

Semantic is a “feeling” expression of images. Sometimes, it is not sufficient to explain an image, but rather, users express the affective semantics of the images using some manual words. Some background information is equally important to express the images, which always contain many hidden relationships. Unlike QF-P which is proposed for integrating annotations, research on multi-keywords’ retrieval based on multiple relations information retrieval, as far as possible, allows relevant keywords to converge to a user specified image, in order to get a probable minimum subset which contains the major and the most correlative annotations. Then, based on these selected and classified annotations, we retrieved the relevant images in the metric space. We called this approach the Geometric-Center algorithm.

For instance, in the 2-keyword query situation, Q_1 and Q_2 are the keywords for query. It is easy to search the Geometric Center (GC) between these two keywords $GC = (Q_1, Q_2)/2$. In the 3-keywords query situation, the $GC = (Q_1, Q_2, Q_3)/3$. By this way, the GC of M -keywords query is $GC = (Q_1, \dots, Q_M)/M$. all major keywords and the most related objects in this relevant subset should be combined and the images that can be indexed and retrieved should be discussed. The detailed algorithm is shown in Figure 8.

THEOREM 1. Suppose a certain M-keywords query $Q[M] = \{Q_1, Q_2, Q_3, \dots, Q_M\}$, we construct a M-level relevant subset S_M is consistent with the distance $d(GC, Q_i)$ from a proximal end to a distal end. The GC of $Q[M]$ is defined as:

$$GC = \sum_{i=1}^M \frac{Q_i}{M} \quad (7)$$

THEOREM 2. Consider the distance $d_i(GC, Q_i)$ from GC to Q_i according to $d_1 \leq d_2 \leq \dots \leq d_M$. For i-level S, $S_i = \{a_i | a_i \in A_s, d(GC, a_i) \leq d_i\}$. If $k' < k$, then proceed to the next level. If the k is big enough, the (M + j)-level detecting distance is

$$d_{(M+j)} = \sum_{i=1}^M d_i \quad (8)$$

Geometric-Center Algorithm

Input: 'Q[M]' is a query of M-keywords;
'd_i' is the distance between GC and Q_i;

Output: relevant subset S.

Begin

For all keywords of query Q[M], do

$GC = \sum_{i=1}^M \frac{Q_i}{M}$;

For $\forall Q_i$, do

distance d_i between GC and Q_i , satisfied
 $d_1 \leq d_2 \leq \dots \leq d_M$;

For $\forall a_i \in A_s$,

If $d(a_i, GC) \leq d_1$, then $a_i \in S$;

If $k' \geq k$,

Return $S \leftarrow F = \{f_0, f_1, \dots, f_k\}$;

Elseif, do $d(a_i, GC) \leq d_2$;

Repeat

Else $d = \sum_{i=1}^n d_i$, then do $d(a_i, GC) \leq d$;

End If

End If

End If

End For

End For

End For

Return S

End.

Figure 8. The Geometric-Center algorithm.

4. Experimental Section

We test our approach on the SBU dataset [25] which consists of images and sentences in English describing these images. SBU consists of descriptions given by image owners when they uploaded them to Flickr. The most reliable evaluation metric is to ask for raters to give a subjective score on the usefulness of each description given to images. For this metric, we selected a basic comparative dataset

from a recent study on image annotation to evaluate the proposed algorithm. We reserved 5000 images as a test set, and split them into 15 CDs based on different categories.

In the experiments, we observe that token rewriting helps to find more relevant keyword elements and, thus, improve the quality of the final keyword search answers for dirty queries. We present an experimental evaluation of the techniques proposed in the paper. For evaluation, we construct a query set using the following procedure. The goals of the experimental study include:

- ✓ To evaluate the quality of the semantic labels (unit, scale and year) and semantic matches discovered by our collective inference approach.
- ✓ To evaluate the impact of the discovered and match labels on image queries.
- ✓ To evaluate the efficiency of QF-P processing.

We manually selected a random category of images. To illustrate the intuitiveness and relevance of this auto-completion system, an instance of the QF-P algorithm will be made available for the demonstration to query data coming from Flickr. For retrieval tasks, we report a recall accuracy curve with respect to the percentage of retrieved S for image retrieval. In our case, we can use concrete cardinality estimates since we are interested only in the precision and recall for QF-P algorithm in the SBU database. More precisely, we define the precision and recall as:

$$\text{Precision, } P = | S_r \cap S_c | / | S_c | \quad (9)$$

$$\text{Recall, } R = | S_r \cap S_c | / | S_r | \quad (10)$$

where S_r is the number of relevant annotations in S , S_c is the number of correlation annotations.

We test different value of parameter k in turn, for $k = 1, 3, 5, 7, 10$, and 15 , respectively. Figure 9 shows the average precision and recall in relevant subset S that we randomly retrieve images with query. In this experiment, P curve is relatively stable except $k = 1$. With the increase of k , more irrelevant annotations are also included in S . However, for P there was no significant change for image retrieval, but for R , more and more relevant annotations are refined in the relevant subset S .

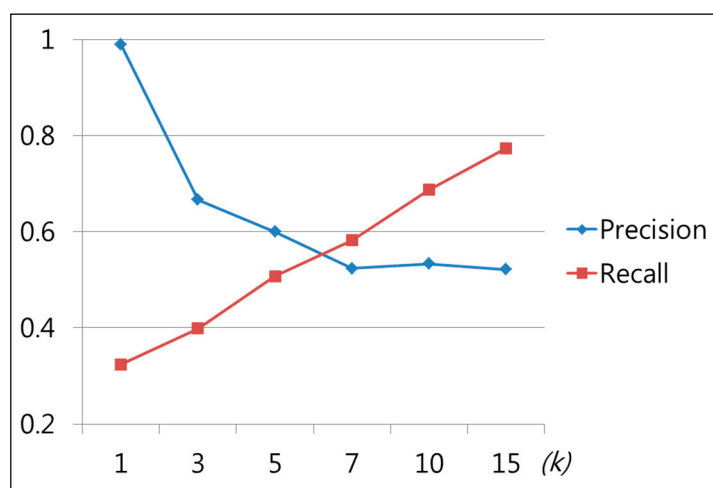


Figure 9. Precision vs. Recall.

E-Measure (E): Definition for image annotation, relevant annotation and correlation annotation in annotation integration leads to the following definition for E measure.

$$E = 2 | (R_n + C_n) | / | A_s | \tag{11}$$

where E value was used as the performance measure. R_n is the number of annotation in the relevant subset, C_n is defined as the number of correlative annotations. We randomly selected six image categories by choosing one random image from each CD, using the QF-P algorithm to refine the annotation. The test E values are shown in Figure 10. The average of E for test image is not much different from others.

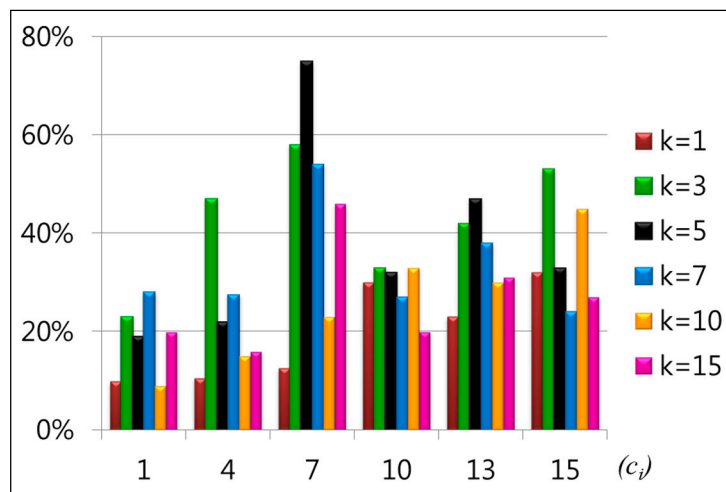


Figure 10. E value of random image in each category.

For a query, were these detected correlative annotations consistent with the top- k most frequent words? Figure 11 clearly shows the results for percentage of detected correlative annotations in the top- k of each category. For the results of the three test categories, we obtained similar results. When $k = 10$, we observe the highest proportion of correlative annotations, almost 50%. However, for $k = 15$, the proportion drops to nearly 40%. It is because more uncorrelated annotations will be detected in the relevant subset. Therefore, k is not necessarily that the bigger, the better, providing the appropriate value for k can be used to improve the image retrieval. In our experiments, $k = 10$ is a good choice.

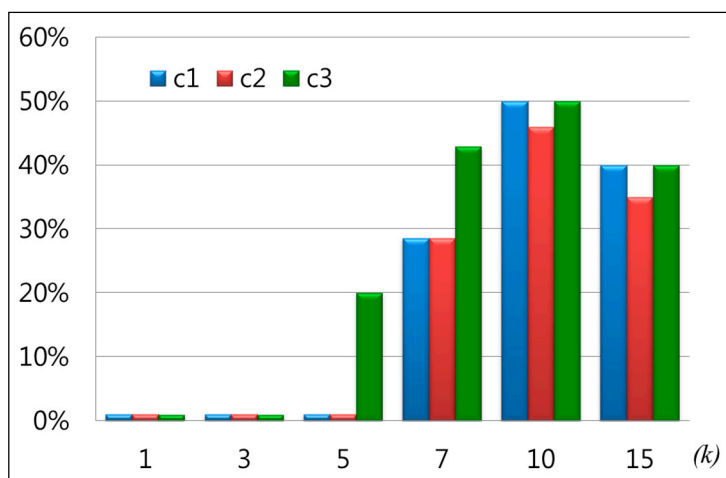


Figure 11. Correlative annotations in top- k of each category.

The ultimate goal of our approach is to effectively retrieve images based on a query. For a query, the top ranking results of image retrieval are shown in Figure 12. Our approach detects the relevant annotations and removes redundant annotations for image retrieval. Since we randomly assign each image to i^{th} category in our experimental, our approach would be to achieve better performance when datasets have a proper image categorization. The experimental results prove that our proposed method is an efficient and effective refinement method for image annotation. In summary, annotation integration has a clear positive effect on the precision of a keyword search, while still preserving high recall when the number of results is not too large.




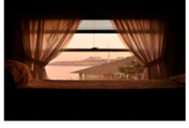












Query	Results			
house	 	 	 	 
water	 	 	 	 

Figure 12. Top ranking results of image retrieval.

5. Conclusions and Future Work

In this paper, we proposed the integration of a probabilistic model which aims at discovering the relationships among image annotations for retrieval. We presented the Query-Frequency Pair (QF-P) algorithm to integrate image annotations for exploring word-to-word relevance to refine more candidate annotations for images. QF-P algorithm focuses on mining the representative image keywords to improve image search result quality. We also discussed a Geometric-Center model for multi-keywords query to study how to assign relevant annotations, which could greatly improve the retrieval accuracy. QF-P is a convergent method, aiming to obtain relevant subsets by query and the most relevant hidden keywords in a related database. The combination of these two kinds of methods is effective for improving query satisfaction based on keywords or tags for various levels of classification.

In the future, we plan to mine and refine more relevant semantic annotations, and bridge more effective connections between image content features and semantic concepts. Thus, we can perform more extensive experiments to enhance our proposed approaches for image retrieval systems. Also, we

will compare other image annotation algorithms in order to make large-scale image annotations more executable and effective.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No.2013R1A2A2A01068923), by Export Promotion Technology Development Program, Ministry of Agriculture, Food and Rural Affairs (No.114083-3) and by the science and technology plan projects of Jilin City, China (No.201464059).

Author Contributions

All authors contributed extensively to the work presented in this paper. Tie Hua Zhou developed the concept and drafted the manuscript, Ling Wang and Tie Hua Zhou designed and performed experiments and analysed data, the revisions and case-study of this paper were led by Keun Ho Ryu.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kennedy, L.S.; Chang, S.F.; Kozintsev, I.V. To search or to label? Predicting the performance of search-based automatic image classifiers. In Proceedings of the 8th ACM Workshop on Multimedia Information Retrieval, Santa Barbara, CA, USA, 23–27 October 2006; ACM: New York, NY, USA, 2006; pp. 249–258.
2. Chua, T.S.; Tang, J.H.; Hong, R.C.; Li, H.J.; Luo, Z.P.; Zheng, Y.T. NUS-WIDE: A real-world web image database from National University of Singapore. In Proceeding of the ACM Conference on Image and Video Retrieval, Santorini, Greece, 8–10 July 2009; ACM: New York, NY, USA, 2009; doi:10.1145/1646396.1646452.
3. Ames, M.; Naaman, M. Why we tag: Motivations for annotation in mobile and online media. In Proceedings of the SIGCHI Conference on Human factors in computing Systems, San Jose, CA, USA, 30 April–3 May 2007; ACM: New York, NY, USA, 2007; pp. 971–980.
4. Singh, M.; Curran, E.; Cunningham, P. *Active Learning for Multi-Label Image Annotation. Technical Report UCD-CSI-2009-01*; University College Dublin: Dublin, Ireland, 2009.
5. Hanbury, A. A survey of methods for image annotation. *J. Vis. Lang. Comput.* **2008**, *19*, 617–627.
6. Chen, L.Y.; Ortona, S.; Orsi, G.; Benedikt, M. Aggregating semantic annotators. *J. VLDB Endow.* **2013**, *6*, 1486–1497.
7. Takhirov, N.; Duchateau, F.; Aalberg, T.; Solvberg, I.T. KIEV: A tool for extracting semantic relations from the World Wide Web. In Proceedings of the Conference on Extending Database Technology, Athens, Greece, 24–28 March 2014; pp. 632–635.
8. Zhang, L.; Tran, T.; Rettinger, A. Probabilistic query rewriting for efficient and effective keyword search on graph data. *J. VLDB Endow.* **2013**, *6*, 1642–1653.

9. Vassilieva, N.S. Content-based image retrieval methods. *J. Programm. Comput. Softw.* **2009**, *35*, 158–180.
10. Bergamaschi, S.; Guerra, F.; Rota, S.; Velegrakis, Y. A hidden markov model approach to keyword-based search over relational databases. In Proceedings of the 30th Conference on Conceptual Modeling, Brussels, Belgium, 31 October–3 November 2011; Springer-Verlag: Berlin, Germany, 2011; pp. 411–420.
11. Duan, H.Z.; Zhai, C.X.; Cheng, J.X.; Gattani, A. Supporting Keyword Search in Product Database: A Probabilistic Approach. *J. VLDB Endow.* **2013**, *6*, 1786–1797.
12. Dalvi, N.; Kumar, R.; Soliman, M. Automatic wrappers for large scale web extraction. *J. VLDB Endow.* **2011**, *4*, 219–230.
13. Ladwig, G.; Tran, T. Index structures and top-k join algorithms for native keyword search databases. In Proceedings of the 20th ACM Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; ACM: New York, NY, USA, 2011; pp. 1505–1514.
14. Elliott, D.; Keller, F. Image description using visual dependency representations. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1292–1302.
15. Kuznetsova, P.; Ordonez, V.; Berg, T.; Choi, Y.J. Treetalk: Composition and compression of trees for image descriptions. *J. Trans. Assoc. Comput. Linguist.* **2014**, *2*, 351–362.
16. Liu, D.; Yan, S.C.; Hua, X.S.; Zhang, H.J. Image retagging using collaborative tag propagation. *J. IEEE Trans. Multimedia* **2011**, *13*, 702–712.
17. Smits, G.; Pivert, O.; Jaudoin, H.; Paulus, F. AGGREGO SEARCH: Interactive Keyword Query Construction. In Proceedings of the Conference on Extending Data Base Technology, Athens, Greece, 24–28 March 2014; pp. 636–639.
18. Hodosh, M.; Young, P.; Hockenmaier, J. Framing image description as a ranking task: Data, models and evaluation metrics. *J. Artif. Intell. Res.* **2013**, *47*, 853–899.
19. Gong, Y.C.; Wang, L.W.; Hodosh, M.; Hockenmaier, J.; Lazebnik, S. Improving image-sentence embeddings using large weakly annotated photo collections. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 529–545.
20. Socher, R.; Karpathy, A.; Le, Q.V.; Manning, C.D.; Ng, A.Y. Grounded compositional semantics for finding and describing images with sentences. *J. Trans. Assoc. Comput. Linguist.* **2014**, *2*, 207–218.
21. Karpathy, A.; Joulin, A.; Li, F.F. Deep fragment embeddings for bidirectional image sentence mapping. In Proceedings of the Conference on Neural Information Processing Systems Foundation, Montreal, PQ, Canada, 8–11 December 2014; pp. 1889–1897.
22. Kiros, R.; Salakhutdinov, R.; Zemel, R. Multimodal neural language models. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 595–603.
23. Mao, J.H.; Xu, W.; Yang, Y.; Wang, J.; Yuille, A.L. Explain images with multimodal recurrent neural networks. In Proceedings of the NIPS 2014 Deep Learning and Representation Learning Workshop, Montreal, PQ, Canada, 12–13 December 2014.
24. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.H.; Karpathy, A.; Khosla, A.; Bernstein, M.; *et al.* ImageNet Large Scale Visual Recognition Challenge. **2014**, arXiv:1409.0575.

25. Ordonez, V.; Kulkarni, G.; Berg, T.L. Im2Text: Describing Images Using 1 Million Captioned Photographs. In Proceedings of the Annual Conference on Neural Information Processing Systems, Granada, Spain, 12–14 December 2011; pp. 1143–1151.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).