Exploiting Time in Automatic Image Tagging

Philip J. McParlane[1] and Joemon M. Jose

p.mcparlane.1@research.gla.ac.uk & Joemon.Jose@glasgow.ac.uk

Department of Computing Science, The University of Glasgow,
Sir Alwyn Building, University Avenue, Glasgow, G12 8QQ, UK

Abstract. Existing automatic image annotation (AIA) models that depend solely on low-level image features often produce poor results, particularly when annotating real-life collections. Tag co-occurrence has been shown to improve image annotation by identifying additional keywords associated with user-provided keywords. However, existing approaches have treated tag co-occurrence as a static measure over time, thereby ignoring the temporal trends of many tags. The temporal distribution of tags, however, caused by events, seasons, memes, etc. provide a strong source of evidence beyond keywords for AIA. In this paper we propose a temporal tag co-occurrence approach to improve upon the current state-of-the-art automatic image annotation model. By replacing the annotated tags with more temporally significant tags, we achieve statistically significant increases to annotation accuracy on a real-life timestamped image collection from Flickr.

Keywords: automatic image annotation, temporal

1 Introduction

With the amount of multimedia data rapidly increasing, it becomes important to organize this content effectively. To be able to facilitate efficient multimedia retrieval we must first categorize these objects with semantic features, such as keywords[2]. However, unlike traditional text retrieval which can infer topics directly from the distributions of words in a document, multimedia objects provide little or no textual clues. Hence, content annotation with semantically related keywords is therefore necessary before indexing and retrieval can take place. The laborious nature of manual image annotation, however, combined with the need for effective large-scale image search has increased research in the field of automatic image annotation (AIA).

Current state-of-the-art AIA models, however, produce poor results, especially when tested on 'real-world' image collections [1]. Bridging the semantic gap between low-level image features and high-level human concepts is still an

---

[2] For the remainder of this paper we refer to tags and keywords synonymously.

unsolved research problem [2]. In any case, many fundamentally question if there even exists a correlation between these two levels [3]. Much research has focused on looking *beyond the pixel* to incorporate more robust evidence in the annotation process [4, 5]. We propose to explore beyond the visual contents of images in the annotation process by exploiting tag co-occurrence and the time an image is taken; by doing so we can avoid, to an extent, the problems associated with content-based image annotation and tagging.

Since the accuracy of AIA is poor, a number of image sharing websites employ user tagging e.g. Flickr. However, the tagging process is either incomplete or often inaccurate. Automatic tagging techniques are often exploited to improve the quality of annotated tags. Tag co-occurrence has been used by existing tag recommendation [6] and AIA systems [5] to improve performance by discovering additional related tags. Tag co-occurrence for two keywords is defined as the number of documents in which both keywords co-exist; in the field of AIA, these documents are images. The motivation for exploiting tag co-occurrence is that keywords exist in a specific distribution which can be exploited. In the field of timestamped text analysis, a significant body of research has sought to exploit dynamic term distributions, most notably for Topic Detection & Tracking [7] and IR [8]. We observed that user tags show temporal co-occurrence patterns; as such, two keywords which co-occur highly in June may not have the same relationship in December. Figure 1 shows normalised example tag distributions over time from a collated Flickr collection. Strong temporal distributions are seen for seasonal keywords such as `summer` and `winter`, which is expected. Further, tags related to weather cycles also observe a relationship with time. For example, `frost` and `snow` are most prominent during the winter months.

It may be argued, however, that only a restricted set of seasonal and weather related keywords will display such strong temporal distributions in image annotation but actually there are many tags with *implicit* temporality. For example, keywords such as `jet` and `pool` are seen to increase during the summer months i.e. typically when people go on vacation. Similarly, `garden` observes peaks during May through September which is expected due to the increase in outdoor activities in summer.

This paper offers a novel approach for harnessing the time an image is taken in the image annotation process. Our methodology only takes a list of annotations for an image (i.e. confidence scores are not required) as input to our annotation improvement model, meaning our approach could be applied as a post processing stage to *any* image annotation model. In this paper we observe which time intervals (time of the day, days of the week, months, etc.) are the most significant in photography by evaluating our approach on a real life image collection. Our approach could be employed in a number of other fields which exploit tag co-occurrence such as: tag recommendation systems as used on social bookmarking websites [9], query expansion [10], event detection [11] and personalised IR [12].

This paper is organised as follows. In Section 2 we present related work in the field of automatic image annotation and temporal IR. Section 3 describes the methodology behind our temporal co-occurrence based approach. In Section
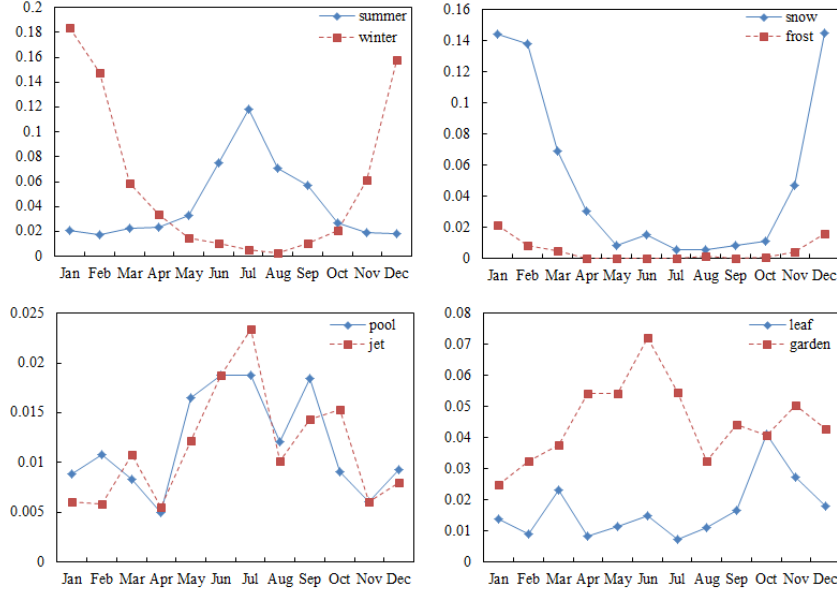
Fig. 1. Tag distributions over time from our Flickr Collection

4 we discuss our experimental setup. Finally, Section 5 presents the results of our experiments and Section 6 concludes and discusses avenues for future work.

2   Related Work

The problem of image classification is often treated as a cross media modelling problem where we try to map low-level features in vector format to high-level textual concepts. Duygulu *et al.* [13] treated the problem of image annotation using a machine translation approach where images are segmented into small regions; keywords were then mapped based on a number of image features. In 2003, Joen *et al.* [14] adopted the cross lingual language model of Lavrenko *et al.* [15], Cross-Media Relevance Models (CMRM), to predict the probability of generating a word given blobs in an image in the training set. The model assumed that regions in an image can be described by a small vocabulary of blobs, which were created from image features using clustering techniques. Lavrenko *et al.* [16] then proposed the Continuous-space Relevance Model (CRM) which generalised the previous CMRM to model highly dimensional continuous features without clustering and quantization. Bag of Visual Words (BOVW) has gained much interest in the field; Carneiro *et al.* [17] proposed a Gaussian mixture model using the bag of local features approach for class conditional dependencies.

More recently, Makadia *et al.* showed that all of the previously stated models could be outperformed by adopting a K-nearest neighbour approach trained on

simple colour and texture image features [18]. In a similar experiment, Athanasakos *et al.* showed that these approaches were also out-performed by using an SVM approach trained on global features [1]. We have chosen to implement the approach by Makadia *et al.* as a baseline, due to its simplicity and performance against other state-of-the-arts (SOTA).

Following research in text based IR [9–12], tag co-occurrence has been used as a secondary source of evidence in tag recommendation systems [6] and image annotation models [19, 5]. Sigurbjornsson *et al.* proposed a tag recommendation strategy to support users annotating photos on Flickr [6]. The relationships between tags were exploited to suggest highly co-occurring tags. Llorente *et al.* incorporated tag co-occurrence in their annotation model which formulated the problem of image annotation as that of direct image retrieval [19]. Novelty was achieved by not only exploring the dependencies between words and their semantic context, but also between visual features and words.

Temporality has previously been studied and exploited in both information seeking and retrieval systems. Despite this, its implication on automatic image annotation has not yet been explored. Klieinberg *et al.* [20] developed a framework for modelling periodic bursts of keywords in a corpus with hierarchical structure using an infinite-state automaton. More recently, Leskovec *et al.* [21] performed a large-scale study of "memes" diffusing throughout news media as a result of temporal rhythms. As a result, a mathematical model was provided for analysing the temporal variation in the context of news. We propose to exploit such temporal trends of tags in a tag co-occurrence model, for the improvement of existing AIA approaches.

3   Temporal Tag Co-occurrence

Given the wide popularity of ubiquitous tagging systems (e.g. Flickr), we propose an automatic tagging process which combines automatic image annotation methods, tag co-occurrence and temporality. We hypothesise that image tagging accuracy will increase by considering *when* an image is taken, in the computation of co-occurrence measures, due to the temporal distributions of tags in image collections, e.g. see Figure 1.

3.1   Problem Statement

Let $I = \{i_1, ..., i_m\}$ denote an image collection, where m is the number of images in the image set. We denote t as a tag and $T = \{t_1, \ldots, t_n\}$ our vocabulary, where n is the number of keywords in our collection. m^t denotes the number of images tagged with t. Every image contains a number of tags in its ground truth, of which we attempt to predict k using an automatic image annotation model. G_x and A_x denote these lists of tags in the ground truth and annotation set respectively, for image x. Of these k predictions, we aim to replace the annotations which are *most likely* to be incorrect, with new tags which co-occur highly with the other tags in the annotation set. The novelty of our approach

is achieved by including the *time a photograph is taken* (extracted from camera meta-data) in the computation of tag co-occurrence measures, as is explained in detail in the following section.

3.2   Tag Co-occurrence Matrices

Firstly, we define a number of co-occurrence matrices with dimensions $m \times n$: 1 global, static co-occurrence matrix and 5 which capture different temporal intervals:

- C^{Global}: denotes the global co-occurrence matrix where an entry (i,j) counts the number of images tags t_i and t_j co-occur in the *entire* training set.
- $C^{Time(h)}$: denotes the co-occurrence matrix where an entry (i,j) counts the number of images, taken at a given time in the day, in which tags t_i and t_j co-occur. h takes 4 different values $\{1, 2 \ldots 4\}$, denoting morning (0500 to 1159), afternoon (1200 to 1759), evening (1800 to 2359) and night (0000 to 0459) respectively. For example, $C_{ij}^{Time(2)}$ denotes the number of images taken in the afternoon in which tags i and j co-occur.
- $C^{Day(d)}$: denotes the co-occurrence matrix where an entry (i,j) counts the number of images, taken on a given day of the week, in which tags t_i and t_j co-occur. d takes 7 different values $\{1, 2 \ldots 7\}$, denoting Sunday to Saturday respectively.
- $C^{Week(w)}$: denotes the co-occurrence matrix where an entry (i,j) counts the number of images, taken in a given week in the year, in which tags t_i and t_j co-occur. w takes 53 different values $\{1, 2 \ldots 53\}$, each denoting a week in the year.
- $C^{Month(c)}$: denotes the co-occurrence matrix where an entry (i,j) counts the number of images, taken in a given month in the year, in which tags t_i and t_j co-occur. c takes 12 different values $\{1, 2 \ldots 12\}$, with each denoting a particular month.
- $C^{Year(y)}$: denotes the co-occurrence matrix where an entry (i,j) counts the number of images taken in a given year, in which tags t_i and t_j co-occur. y takes the value of any year after 2005, up until 2012. The years before 2006 are ignored due to the sparsity of data in this period.

For every co-occurrence matrix, we normalise using two steps. First, all diagonal values are set to zero as they dominate the matrix and are useless as they do not offer information about the *co-occurrence* of tags. Secondly, all values in a given column are normalized by dividing by the maximum value in the column. Therefore, the maximum value in all matrices is 1. This method has previously been used in an image tag recommendation system with much success [22].

3.3   Tag Removal and Promotion

The problem of improving annotations made by an existing image annotation system is modelled into two stages, first, the tags which are *most likely* to be incorrect in the annotations are removed. Therefore, for each tag in the annotation set $t \in A_x$, we compute the following score:

$$remove(t, A_x) = idf(t) \sum_{j \in A_x} C_{ij} \qquad (1)$$

where $idf(t)$ is the inverse document frequency for tag t defined as $idf(t) = log(m/m^t)$. The likelihood of a tag being a correct annotation is therefore computed as the product of its inverse popularity and the co-occurrence with the other annotations. The tag with the lowest score is removed and is replaced by a new tag from $\hat{t} \in T$ which is computed as the sum of co-occurrences with each of the tags in the annotation ($t \in A_x$):

$$addition(\hat{t}, A_x) = \sum_{j \in A_x} C_{ij} \qquad (2)$$

By adopting this simple methodology, building on previous work [22], we are able to offer a generic solution which can be applied to any annotation model as the input is simply a list of annotations for a given image.

Our model proposes to incorporate time in the computation of tag co-occurrence measures. For a given image, this is achieved by replacing the co-occurrence matrix in Equation 2 with any of the temporal based co-occurrence matrices defined in 3.1, for the interval the image was taken in. By switching to a temporal tag co-occurrence matrix, we offer tags which are not only semantically related, but are also temporally relevant.

3.4   Content Based Annotation

Previous work [18, 1] has shown that SVM or KNN models trained on global image features perform better than other, more sophisticated, AIA models. Our proposed approach, therefore, builds on top of the current state-of-the-art automatic image annotation model [18]. The algorithm models the problem of image annotation as that of image retrieval using a K-Nearest Neighbour approach. Seven features are extracted from images, three colour histograms with 16 bins in each channel (in RGB, HSV and LAB colour space), two texture descriptors (using HAAR and Gabor filters) and two quantized versions of the texture features.

Given a test image, the model transfers 5 tags from the nearest image in the training collection. If the nearest neighbour contains less than 5 tags in its ground truth however, the remaining tags are made up from the top tags in neighbourhood, based on the co-occurrence and frequency of the tags. The distance between images is calculated as an equal combination of distances for each feature, measured using the L-1 distance and KL-divergence measure. The features and the tag transfer process are described in detail in Section 3 of [18].

4   Experiments and Data

4.1   Data Set

For our experiments, we used a subset (7000 training images and 500 test images) of a large image collection (2 million images) crawled from Flickr[1]. A subset was used due to the computationally expensive nature of the brute force KNN approach adopted by the annotation model as defined in [18]. Further, we created our own collection for two reasons. Firstly, many image collections, such as Corel5k [13], do not contain image timestamps. Secondly, for those collections that do, for example IAPR TC-12 [18], we cannot ensure that the images are selected without temporal bias. The collection was therefore collected as follows:

– **Step 1:** Initially a number of *nouns*, which were used to query the Flickr API, were taken from the lexical database, WordNet [23]. The 2000 most popular nouns (which were categorised into the following categories: `animal, artifact, body, food, plant, substance`), as ordered by their term frequency in WordNet, were selected. These nouns are representative of popular concepts which can be used to query Flickr. The diversity of these tags is highlighted in Figure 2, a word cloud representing the top 100 nouns used to query Flickr.
– **Step 2:** A Flickr search was executed for each noun; the results were ordered by decreasing relevance. The top 2000 images from each search were then used in our collection. Using 2000 images on a wide number of nouns achieves a general, balanced collection with sufficient training data per topic.

Fig. 2. Top 100 Nouns used to query Flickr

A number of cleaning stages were then taken out on the collection to make it useful for image annotation. This was taken out due to the large amount of

---

[1] This data is available at `http://www.dcs.gla.ac.uk/~philip/`

noise present in online image collections [24]; leaving noisy tags makes automatic annotation difficult due to the large number of concepts. In our collection, we consider tags with insufficient training data per tag (i.e. tags which are used in very few images) as noise (as tags which exist in only a few images are useless for the purposes on image annotation). Further, some tags initially appear valid due to their occurrence in a number of images, however, many tags in this set are only used by a single user, often as a result of batch tagging (where one can automatically annotate multiple images with a predefined tag set). Finally, Flickr adds a number of automatic tags to images, which are awards, often based on user interest for a particular image; these tags are deemed useless for image annotation. We therefore cleaned the collection using the following approach:

- **Removing Flickr Automatic Tags:** Initially the automatic Flickr tags were removed manually from the 1000 most frequently occurring keywords. Tags were deemed irrelevant if they were camera meta data (e.g. `d60`), a Flickr award (e.g. `excellentphotographeraward`) or were the names of a Flickr group or task (e.g. `5photosaday`).
- **Remove sparsely annotated images:** Those images which contained less than 4 tags were removed.
- **Removing frequent tags from few users:** Those tags which were used by less than 150 users were removed.

After this cleaning stage, we selected a *random* subset of 7000 images for training and 500 images for testing. Any tags which existed in less than 100 images in the subset were also removed. The resulting collection contained 120 tags with an average of 4.2 tags per image. Images are taken between January 1922 and October 2012, with 97% of the images taken between January 2006 and October 2012.

By taking a random subset of well annotated images from a large collection which is sampled over a long period of time, we obtain a test collection with a good user/topic distribution which isn't temporally bias (i.e. where the temporal window is defined).

4.2   Experiments and Procedure

We initially annotate every image in the test set with 5 tags, based on image contents, using the K-Nearest Neighbour approach described in 3.4. This technique of suggesting 5 tags has been used by previous work [18]. For each of the 5 annotated tags, we compute the tag likelihood score, $remove(t, A_x)$. The tag which produces the lowest score is deemed least likely to be a correct annotation and is therefore removed. Tag co-occurrence is then computed for all tags in the vocabulary $t \in T$ with the four remaining annotations in the given image as, described in Equation 2. For each tag t, the co-occurrence with the tags in A_x are summed together. The tag which co-occurs highest is used as a new annotation and is therefore added to the image's annotation list. In our experiments, the co-occurrence matrix used to compute co-occurrence scores is changed between our baseline (which ignores temporality) and the 5 temporally significant

co-occurrence matrices, as well as a combination of these as described in the following section.

We consider the tags provided by users as the ground truth for a given image. Therefore, our model attempts to correctly identify these tags. For each of the approaches, we compute the standard measures as used in automatic image annotation, *precision*, *recall* and the *number of words recalled* (i.e. number of tags which are annotated correctly *in at least 1* image).

4.3   Annotation Models

We compare annotation accuracy of our 8 different methods:

- **KNN (Content)**: Our annotation baseline is the model described in Section 3.4 which annotates images using a K-Nearest Neighbour method trained on a number of texture and colour based image features. As previously described, for each image, 5 annotations are suggested.
- **Baseline (Content + Co-occurrence)**: Our co-occurrence baseline improves suggestions made by KNN by replacing deemed incorrect annotations with new tags which co-occur highly with the other tags in the ground truth as described in Section 3.1. Temporality is *ignored*, however, in the computation of co-occurrence measures.
- **Hour, Day, Week, Month, Year (Content + Co-occurrence)**: We incorporate time into the annotation process by calling on tag co-occurrence measures from images taken within a given time window. For each approach, we call upon the relevant matrix as described in Section 3.2.
- **All (Content + Co-occurrence)**: We also combine the 5 temporal approaches previously described by equally adding the tag co-occurrence measures from each of temporal co-occurrence matrices, when computing tag co-occurrence for a given tag. By doing so we are able to combine the hourly, daily, weekly, monthly and yearly cycles of tags co-occurrences.

5   Results and Discussion

The following sections detail the impact of including time in the computation of tag co-occurrence measures. Table 1 displays the overall results, comparing our 8 methods.

By incorporating the time an image is taken, we are able to produce accuracy increases of up to 5% over our static co-occurrence baseline (the "static co-oc" column). The greatest increases to accuracy are achieved when we consider the day in a week and year an image is taken in, highlighting the tag cycles in weekly life as well as the yearly trends/events. Setting the time interval to 1 week gave a slight reduction in performance, however, which may be explained due to a week having less rigid cycles than that of a single day or the seasons in a year.

Figure 3 shows an example image from our test collection as annotated by our different models; a number of key points are captured in this image. First

	Contents	*Static*	*Temporal Tag Co-occurrence*					
	KNN	**Co-oc**	**Hour**	**Day**	**Week**	**Month**	**Year**	**All**
Precision	0.0703	0.06921	0.07169	**0.0728**	0.0693	0.07198	0.0725	0.0715
Recall	0.0804	0.0829	0.0865	0.0863	0.0820	0.0826	**0.0875**	0.0841
Words Recalled	72	71	71	71	**75**	74	74	71
F-1 Measure	0.0750	0.0755	0.0784	0.0790	0.0751	0.0769	**0.0794**	0.0773
+/- Over Baseline	N/A	N/A	+3.8%	+4.6%	-0.5%	+1.9%**	**+5.2%**	+2.4%*

Table 1. Full Results: Comparison of annotation based purely on image contents (KNN), image contents and tag co-occurrence combined and image contents and temporal tag co-occurrence. The bottom row denotes the percentage improvement/reduction when comparing the given *temporal approach* against the *static co-occurrence approach* i.e. the percentage improvement posed by temporality. The best performing models for each measure displayed in bold. Paired t-test statistical significance comparing our experimental approach against the baseline are denoted as * being $p < 0.05$ and ** being $p < 0.01$.

of all, the incompleteness of the ground truth makes it particularly difficult to annotate; as a human, it would be impossible to predict the ground truth exactly, as the tag `usa` appears to have no relevance to the image. Secondly, the KNN annotation model makes a number of predictions which could potentially be interpreted as correct i.e. `nature`, `spring`. These factors perhaps explain to an extent, the low performance of the annotation model as seen in Table 1. Secondly, our approach successfully removes the tag with (arguably) the least relevance to the image: `art`. Finally, the co-occurrence baseline then offers the incorrect tag, `green`, whereas our temporal approach successfully identifies the correct tag, `flower`, due to its high co-occurrence with the other annotations in May. We would expect the number of photographs tagged with the tag `flower` to increase during this period, which is captured by our model.

Fig. 3. Example Image: Comparison of annotations

GT   [color, flower, usa]
KNN *(0)*   [art, blackandwhite, grass, nature, spring]
Baseline *(0)*   [**green**, blackandwhite, grass, nature, spring]
Temp Co-oc *(1)* [**flower**, blackandwhite, grass, nature, spring]

Image taken on May 14, 2009

Figure 4 shows tag clouds generated during a given period. These were generated by subtracting the fraction of images taken in a given time interval for a given tag from the fraction of images in which it occurs over the entire collection. Keywords with obvious temporal relationships such as `spring` and `autumn` can be clearly observed, however, it is the tags with implicit temporal relationships such as `bar`, `festival`, `restaurant` and `nightclub` which are of most interest as it emphasises that temporality is present beyond weather and seasonal cycles.

Fig. 4. Word Clouds as Generated from the Overall Collection. Top: Night, Morning, Afternoon, Evening. Bottom: Spring, Summer, Autumn, Winter

6   Conclusion & Future Work

In this paper we built upon an existing state-of-the-art automatic image annotation model by removing deemed irrelevant tags and by offering new suggestions based on their co-occurrence with the other tags in the annotation set. In particular, we incorporated the time an image was taken in the computation of co-occurrence measures. We tested our approach on an image collection crawled from Flickr. The analysis of the results showed that by incorporating time in tag co-occurrence measures we (statistically significantly) outperformed a non temporal baseline in annotation accuracy. Further analysis showed that both short term temporal distributions (i.e. during a single day) as well as long term temporal trends (i.e. over a year) can be used to increase annotation accuracy.

As future work we will explore the temporal aspect of tag co-occurrence in image tag recommendation models. Finally, we will consider more elaborative models for the combination of temporal data with other data sources, such as geographical information.

References

1. Athanasakos, K., Stathopoulos, V., Jose, J.M.: A framework for evaluating automatic image annotation algorithms. In: ECIR '10 Milton Keynes, UK. (2010)
2. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. **22**(12) (dec 2000) 1349 –1380
3. Santini, S., Gupta, A., Jain, R.: Emergent semantics through interaction in image databases. **13**(3) (6 2001) 337 –351
4. Monaghan, F., O'Sullivan, D.: Leveraging ontologies, context and social networks to automate photo annotation. SAMT'07, Springer-Verlag 252–255
5. Li, W., Sun, M.: Automatic image annotation based on wordnet and hierarchical ensembles. CICLing'06, Berlin, Heidelberg, Springer-Verlag (2006) 417–428
6. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. WWW '08, New York, NY, USA, ACM (2008) 327–336
7. Allan, J.: Topic detection and tracking, Norwell, MA, USA, Kluwer Academic Publishers (2002) 1–16
8. Whiting, S., Moshfeghi, Y., Jose, J.M.: Exploring term temporality for pseudo-relevance feedback. SIGIR '11, New York, NY, USA, ACM (2011) 1245–1246
9. Cattuto, C., Benz, D., Hotho, A., Stumme, G.: Semantic grounding of tag relatedness in social bookmarking systems. ISWC '08, Berlin (2008) 615–631
10. Jin, S., Lin, H., Su, S.: Query expansion based on folksonomy tag co-occurrence analysis. (aug. 2009) 300 –305
11. Yao, J., Cui, B., Huang, Y., Zhou, Y.: Data engineering (icde), 2010 ieee 26th international conference on. (march 2010) 780 –783
12. Byde, A., Cayzer, S.: Personalized tag recommendations via tagging and content-based similarity metrics. New York (2) (2007)
13. Duygulu, P., Barnard, K., De Freitas, J., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. ECCV '02
14. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. SIGIR '03, NY, USA 119–126
15. Lavrenko, V., Choquette, M., Croft, W.B.: Cross-lingual relevance models. SIGIR '02, New York, NY, USA, ACM (2002) 175–182
16. Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: IN NIPS, MIT Press (2003)
17. Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Transactions **29** (2007)
18. Makadia, A., Pavlovic, V., Kumar, S.: Baselines for image annotation. Int. J. Comput. Vision **90**(1) (October 2010) 88–105
19. Llorente, A., Manmatha, R., Rüger, S.: Image retrieval using markov random fields and global image features. CIVR '10, NY, USA, ACM (2010) 243–250
20. Kleinberg, J.: Bursty and hierarchical structure in streams. KDD '02, New York, NY, USA, ACM (2002) 91–101
21. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the dynamics of the news cycle. KDD '09, New York, NY, USA, ACM (2009) 497–506
22. Garg, N., Weber, I.: Personalized, interactive tag recommendation for flickr. In: RecSys. (2008) 67–74
23. Miller, G.A.: Wordnet: A lexical database for english. Communications of the ACM **38** (1995) 39–41
24. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: WWW. (2008) 327–336