Placing images on the world map: a microblog-based enrichment approach

Claudia Hauff[*]
Delft University of Technology
Delft, The Netherlands
c.hauff@tudelft.nl

Geert-Jan Houben
Delft University of Technology
Delft, The Netherlands
g.j.p.m.houben@tudelft.nl

ABSTRACT

Estimating the geographic location of images is a task which has received increasing attention recently. Large numbers of images uploaded to platforms such as Flickr do not contain GPS-based latitude/longitude coordinates. Obtaining such geographic information is beneficial for a variety of applications including travelogues, visual place descriptions and personalized travel recommendations. While most works in this area only exploit an image's textual meta-data (tags, title, etc.) to estimate at what geographic location the image was taken, we consider an additional textual dimension: the image owner's traces on the social Web. Specifically, we hypothesize that information extracted from a person's microblog stream(s) can be utilized to improve the accuracy with which the geographic location of the images is estimated. In this paper, we investigate this hypothesis on the example of Twitter streams and find it to be confirmed. The median error distance in kilometres decreases by up to 67% in comparison to existing state-of-the-art. The best results are achieved when tweets that were posted up to two days before and after an image was taken are considered. Moreover, we also find another type of additional information useful: population density data.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval: Information Search and Retrieval
General Terms: Human Factors, Experimentation
Keywords: image location estimation, image placing

1. INTRODUCTION

Estimating the geographic location of images (also referred to as geotagging, georeferencing or image placement on the world map) is a task which has received increasing attention in recent years. Large numbers of images uploaded

to platforms such as Flickr[1] do not contain GPS-based latitude/longitude coordinates. Placing those images on a world map by obtaining geographic information is not only important to aid people in the browsing and organizing of their personal image archives, but it also plays a role in application scenarios for large geographically tagged image corpora such as the automatic illustration of travelogues [16] and personalized travel recommendations [6, 7, 8].

While GPS-enabled cameras are beginning to reach the mainstream market (where an image's latitude/longitude coordinates are automatically recorded as meta-data), this technology is not yet in widespread use. As an example, consider the test set of images we crawled from the photo sharing platform Flickr to test our hypotheses. Of the images that were taken within the last year, more specifically between November 2010 and September 2011, only twenty percent are geo-tagged- since Flickr allows users to manually add geo-information to images after uploading them, this number can be considered as an upper bound to the use of GPS-enabled devices in our data set.

Thus, there are vast amounts of images which are not geographically tagged. A number of approaches, e.g. [21, 22, 23, 13], have been developed that estimate the geographic location of images based either on the textual meta-data that is available for each image (most often the tags that users assign to images), the visual features derived from the image itself, or a combination of the two. These works exploit the very popular image sharing and organizing platform Flickr which we consider here as well. While to our knowledge the existing approaches only consider information derived directly from elements within Flickr (the image itself or meta-data attached to the image), we look beyond this single platform and investigate if location estimation can be improved when considering traces of the image owner/uploader on other social Web platforms. We consider the exploitation of user traces across platforms an important step forward, as a myriad of social Web platforms exist and users are often active on a number of these.

In this work, we consider microblogs, and here more specifically we focus on the microblogging platform Twitter[2]. We hypothesize that we can improve the accuracy of the geographic location estimation when we not only consider the tags the user has assigned to the image but when we also consider the utterances the user has made on Twitter in approximately the same time frame.

---

[1]Flickr, http://www.flickr.com/
[2]We chose Twitter as it is in wide-spread use around the world.

Imagine a user Alice who for her holidays travelled to North America for a week in the summer. Alice visited Vancouver and Seattle. Alice is also very active on the social Web, having among others a Flickr and a Twitter account which she uses regularly to stay in touch with friends and family. She later uploads the pictures she took during the trip to Flickr. Instead of tagging every single picture with the correct name of the town or monument visited, to safe time, Alice decides to use the same set of tags for all images that describe the entire holiday {*northamerica, seattle, vancouver, holiday*}. During the holiday, Alice regularly tweeted about what she was doing and where she was going, e.g.:
— *"Vancouver this weekend :)"*
— *"Day 1 in Seattle - Been to the Space Needle. Sunny day."*
— *"Sitting in a cafe in Georgetown drinking wine with friends."*
Since posts on Twitter are always time-stamped, we can easily determine which tweets Alice posted around the time of a particular image being taken (information that is readily available for images). We hypothesize that such tweets, where a user mentions place names, particular shops or monuments she has visited or is planning to visit, can improve the accuracy of the location estimation process. Eventually, our goal is the development of a framework, that automatically estimates latitude/longitude of each image, based on the image tags Alice provided *and* the traces Alice left on the social Web. Since various applications, that rely on geographically tagged images are beneficial to the end user (Alice), we expect Alice to provide her online personas and user names willingly.

In this paper, we formulate and investigate the hypothesis that a user's traces on Twitter can increase the accuracy of the image location estimation process. In particular, in this work we make the following contributions:

- We developed an extension to an existing state-of-the-art tag-based approach to location estimation [21] which is based on the language modeling framework for information retrieval. Our extension makes it possible to exploit the information extracted from the user's tweets in a natural way.

- We show that exploiting a user's Twitter stream when estimating the location an image was taken at decreases the median error distance by up to 67%. We also find that the best results are obtained when tweets within two days before and after the image was taken are exploited.

- We find that images, which have fewer than two tags with a geographic scope (see Sec. 3.1.1) assigned to them, benefit the most from our cross-system approach. In our experiments, the median error distance decreases by 40% (57%) for images without tags (with one tag) when the user's Twitter stream is employed as alternative source of information.

- Lastly, we also performed an investigation into the added value of using non-uniform priors in the language modeling approach. We utilized prior information about the population density and the climate. Our results show that a prior based on population density also improves the accuracy of the location estimation.

In the remainder of this paper we first outline previous research (Sec. 2) in the areas of image location estimation and users' motives for using Twitter. The methodology of our approach is presented in Sec. 3. The experimental setup and the data sets used are described in Sec. 4, followed by the results and their discussion in Sec. 5. We conclude the paper in Sec. 6 and outline some avenues for future work.

2. RELATED WORK

Approaches proposed to solve the task of placing images on the world map by determining their latitude and longitude, have been relying on a variety of sources that are based on the image and its meta-data, the user uploading the image or external knowledge bases. Textual features exploited from the image are mostly the assigned tags and the title as well as the description. Visual features derived from the images include a variety of types, such as color histograms or edge histograms [17]. Since photo sharing platforms such as Flickr contain a social network component as well, it is also possible to exploit such information by for example considering the friends of a user or the location of the users commenting on an image. Finally, external knowledge bases, in particular gazetteers (geographic dictionaries), may also be utilized in this task to classify terms as either being geographic in scope or not.

In [21] the tags assigned to images on Flickr are exploited. A grid is placed over the world map which results in equally sized cells. Each training image (with known location) is assigned to its correct grid cell. For each cell, a language model is created from the tags assigned by the user, and a test image is assigned to the geographic cell that produces the highest probability for generating the image's tag set. In contrast to our approach, the cells are of fixed size, which may not be optimal as some regions (e.g., large cities such as London and Paris) will have a larger density of pictures than relatively remote places. We use dynamically sized grid cells, to account for these differences. To determine whether a tag is geographic in nature, in [21] GeoNames[3] is employed, a large gazetteer of geographic entities. Speficically, the weights of tags that appear in the English part of GeoNames are boosted.

Also based on tags is the approach proposed in [22]. In contrast to [21], the location estimation is performed on different levels of granularity (city granularity, street level granularity, etc.) and the evidence obtained over the different granularities is combined in order to output the best match granularity location estimate. The authors evaluate their approach by creating a data set with 54 distinct locations (cities in Europe) and classifying test images according to the city they were taken in. This setup is more restricted than the one we investigate in this paper: we attempt to place test images as closely to their true geographic location as possible; we are not restricted by a set of classes.

Instead of determining the correct grid cell and returning the latitude/longitude of the cell's center, a text-based two-step approach is proposed in [23]: first, the most likely area is found by a language modeling approach and within the found cell, the best match images are determined by a similarity search. A test image with unknown location is then assigned the location found by interpolating the locations of the most similar images.

An approach that not only exploits textual information but also visual features was proposed in [13]. Here, textual information (tags) is the primary source of information, and visual features are used as fall-back option in instances where

---

[3]GeoNames, http://www.geonames.org/

tags do not provide meaningful information. The approach works in different stages: first, the image tags are evaluated for the occurrence of at least one known geographic location (geographic lookup). If no location is found, PLSA [10] is performed on the tag data of the corpus. A failure here results in the exploitation of visual features which are used as input to a support-vector machine based classifier.

Georeferencing has not only been applied to images or videos. It has also been utilized to determine the geo-location of the subjects of Wikipedia articles (e.g. [25]) and the geo-location of Twitter messages as well as the respective home location of Twitter users [14, 4]. In [4] it was found that for 51% of Twitter users their home location on the city level could be identified within 161 kilometres (100 miles) of their true location. Based on a large training corpus of Twitter message, the authors trained models (term distributions) for a large number of cities in the USA. While this approach yields one location per user, we are concerned in our work with estimating one location per item (in our case an item is an image).

Finally, we also note a number of works that have considered the questions of why people use Twitter and what they tweet about. Java et al. [12] developed a number of tweet categories: daily chatter (most common use of Twitter), shared information and hyperlinks, conversations and news. In [18] the majority of users (80%) were found to focus on themselves in their tweets, while only a minority of users are driven by the sharing of information. Lastly, Zhao et al. [28] conducted interviews asking users about their motivations for using Twitter; several major reasons surfaced including keeping in touch with friends and colleagues and collecting useful information for one's work and spare time. Overall, these studies show that a lot of tweets are concerned with the user herself; we hypothesize that among these user-centred tweets, there are also useful ones for the derivation of the location of images that were taken by the user.

Our method is similar in spirit to the just described tag-based image location estimation approaches; differences are pointed out in detail in Sec. 3. We would like to stress that in this paper our main focus is on the question of whether a user's traces on the social Web platform Twitter can offer valuable information for the image location estimation task. Such cross-system exploitations have recently begun to attract interest in a number of applications, e.g., [1, 3, 11]. In [1] for example, it has been investigated how users use tags on different social Web platforms (Flickr, Twitter and Delicious) in the context of personalized tag and resource recommendation. It was found that a cross-platform approach can lead to a considerable improvement of the recommendation quality.

In a preliminary study [9] we investigated the research question posed here on a very small data set in a coarse setup: we utilized the most recent tweets and most recent images (instead of aligning tweets and images according to publishing dates) in an oracle setting (instead of an automatic setup to combine both information sources) and found first evidence that this combination improves the image location estimation quality.

3. METHODOLOGY

Previous works have shown that geographic location estimation approaches exploiting textual meta-data provided by the image owner (such as image tags, title, etc.), outper-

form approaches that rely mainly on visual features. Here, we investigate two additional sources of information: (i) the image owner's traces on the microblogging platform Twitter, and, (ii) prior information about the world, in particular population density and climate. We investigate two hypotheses with our experiments:

- In cases where little (or no) textual meta-data is available for an image, considering a user's tweets as a source of additional textual information will improve the accuracy of the location estimation.
- Adding world knowledge as prior information will improve the estimation accuracy.

We first outline the baseline approach, which is grounded in the language modeling approach to information retrieval and then we turn to describing our proposed extensions.

3.1 Baseline Approach: World Regions as Language Models

Our baseline approach, which is a state-of-the-art approach we compare our Twitter-based extension against, draws inspiration from a number of recently published works. Following [21], we employ the language modeling approach to information retrieval [19, 15, 27, 26]. Instead of documents and queries though as in retrieval, we deal with world regions and image tags. A language model θ_R, which in the retrieval setting is derived for each document, is derived for each region R of the world. Given a test image $\Im$ with tags $T_{\Im} = \{t_1, ..., t_n\}$ and unknown latitude/longitude, the tags are considered as query terms and the region language models are ranked with respect to the probability $P(\theta_R|T_{\Im})$, which according to the Bayes theorem can be expressed as:

$$P(\theta_R|T_{\Im}) = \frac{P(T_{\Im}|\theta_R)P(\theta_R)}{P(T_{\Im})} \qquad (1)$$

$$\propto P(\theta_R) \prod_{t_i \in T_{\Im}} P(t_i|\theta_R). \qquad (2)$$

This is the standard query likelihood based language modeling setup which assumes term independence. Usually, the prior probability of a region $P(\theta_R)$ is considered to be uniform, that is, each region in the world is equally likely. The language models are multinomial probability distributions over the textual meta-data of all training images that were taken in region R, that is, the textual information of all these images is concatenated and treated analogously to a single document.

Since a maximum likelihood estimate of $P(t_i|\theta_R)$ would result in a zero probability of any region that misses one or more of the tags in $T_{\Im}$, the estimate is usually smoothed with a background language model, generated over all documents in the training corpus. In accordance with [21], we found Dirichlet smoothing [26] to yield the most accurate results when evaluated on our data:

$$P(t_i|\theta_R) = \frac{c(t_i, R) + \mu P(t_i|\theta_C)}{|R| + \mu}. \qquad (3)$$

Here, μ is the smoothing parameter, $c(t_i, R)$ is the count of tag t_i in R and $|R|$ is the length of the region document (derived by concatenating all tags of all training images in the region). The probability $P(t_i|\theta_C)$ is the maximum likelihood probability of tag t_i occurring in the collection lan-

guage model θ_C (derived by concatenating all region documents).

Having identified the most likely region θ_R given test image $\Im$ is only the first step, as such regions often cover hundreds of kilometres and simply assigning the center of the region as estimated latitude/longitude to image $\Im$ is not sufficient. Thus, similar to [23], in a second step we only consider the training images occurring within R. A language model is generated for each of these training images, which are then ranked according to their probability of generating $T_\Im$. The latitude/longitude of the top ranked training image within R is assigned to test image $\Im$.

In contrast to [21], we do not partition the world map into regions (or cells) of fixed size. Instead, we found cells of varying size to yield more accurate results: starting with a grid cell that spans the entire world map (if viewed as a graph, this cell is the root node), the training items are added to the cell one at a time. Once the number of items in a cell exceeds the set limit ℓ_{split}, the cell is split into four equally sized cells, each covering a quarter of the original cell (four children nodes are added) and the training items are re-distributed to these cells. To avoid too many splits in areas where large amounts of training data are available, a cell may not be split any further if its latitude/longitude range reaches a lower limit ℓ_{lat_lng}. This process yields cells of small size in areas where the training data is dense, and cells of large size in areas where the training data is sparse (e.g. regions covering oceans).

If a test image $\Im$ contains no textual elements (or all terms were removed during the filtering steps described in Sec. 3.1.1), the terms in the user location are used instead, which has been proposed in [5]: if a user does not tag an uploaded item with its location, it may have been taken at the user's home location. In contrast to [5], we add the user location terms to $\Im$, instead of relying on an external resource to convert the user location to geographic coordinates which is also often inaccurate due to the ambiguity of place names. Finally, if none of these steps yield a non-empty set $T_\Im$, the test image is assigned the latitude/longitude coordinates of the most frequently occurring location in the training data.

3.1.1 User-Based and Geographic Spread Filtering

In exploratory experiments, we found two strategies that filter terms from the test images' tag sets $T_\Im$ to lead to more robust and considerably more accurate estimations. Firstly, terms are removed from $T_\Im$ that were utilized as tags by less than U users in the training corpus, in line with [22], and secondly, terms are removed that appear in many different geographic areas in the training data. Whether a term is likely to have a geographic scope can either be determined by matching the term against a geographical dictionary such as GeoNames as done in [21] or by considering how localized the term occurs in the training data. We follow the latter approach here. For example, while in our training data the term "sydney" occurs primarily in one particular region (the area containing the location of Sydney, Australia), the term "bowling" appears in many different regions. This observation leads to a simple yet effective filtering method: given a term t_i whose spread is to be determined, a grid is placed over the world map (1 degree latitude/longitude range per cell) and the number of training items in the cell that contain t_i are recorded. Neighbouring grid cells with a non-zero count are merged (to avoid penalizing geographic terms that

cover a wide area) and the number of non-zero connected components are determined. This score is normalized by the maximum count. Thus, the smaller the score, the more localized t_i occurs in the training data. While yielding comparable results, our approach is considerably simpler than the χ^2 feature selection based geo-term filtering [23], which determines the geographic score for the tags in each cell separately.

3.2 Adding Additional Knowledge

We consider the language model approach with user-based and geographic spread filtering as our baseline and attempt to improve upon it by relying on additional sources of information. When considering Equation 1, additional information can be added in this framework in three ways:

- **The set of terms** $T_\Im$: The bag-of-words that describe an image can be extended by not only including the tags that are assigned to image $\Im$, but by also including terms from the user's traces on the social Web, in particular the user's Twitter stream.

- **The prior probability** $P(\theta_R)$: Instead of a uniform prior probability, we include knowledge about the world and, for instance, assign a higher prior probability to regions of high population density or regions of moderate climate.

- **The region language model** θ_R: The language model of a region R can be expanded by adding additional information, for example, information extracted from Wikipedia.

In the work presented here, we investigate the first two aspects, expanding the term set used for image location estimation and adding informative priors. We leave the third aspect for future work. Discussed next are the usage of terms extracted from a user's tweets and the usage of informative priors.

3.2.1 Terms Extracted from Tweets

The hypothesis we investigate in this work is that traces a user leaves on Twitter can aid us in estimating the geographic location of the user's images on Flickr. Location estimation based on an image's textual meta-data is based on the assumption that a user who uploads an image or video also spends some time on adding tags, possibly also a title and a description. Not every user though has the time or the patience to do this. In particular in instances where a user adds no or just very little extra information, we expect the user's traces on Twitter to be valuable.

On Twitter, users post short messages (tweets) with up to 140 characters about anything they choose. They can be followed by other users and themselves follow users in order to receive their tweets. Tweets can be directed (at *@user*) and tweets can contain hashtags (*#sigir2012* or *#portland*)[4]. The microblogging service Twitter was chosen due to its popularity and wide-spread use today. Given the date $\mathcal{D}$ (day, month, year) a test image $\Im$ was taken at, we extract the tweets by the user who posted the image to Flickr. We consider only those tweets by the user that were posted within d days of $\mathcal{D}$. After the removal of URLs and user names (*@user*), we are left with a bag-of-words $\mathcal{W}_\Im = \{w_1, ..., w_n\}$.

---

[4]We also note, that Twitter allows tweets to be geo-tagged directly, though this feature is rarely enabled by users and only available when using Twitter from certain devices. In [4] only 0.42% of tweets were found to be geo-tagged.

The world regions are no longer only ranked according to their probability of generating the image tags (Equation 2), but also according to their probability of generating the twitter terms. In our implementation, we interpolate both scores for the final ranking of each world region. Formally:

$$P(\theta_R|T_{\Im}, \mathcal{W}_{\Im}) \propto P(\theta_R) \times \tag{4}$$

$$\left(\lambda \prod_{t_i \in T_{\Im}} P(t_i|\theta_R) + (1-\lambda) \prod_{w_i \in \mathcal{W}_{\Im}} P(w_i|\theta_R) \right) \tag{5}$$

Both, $P(t_i|\theta_R)$ and $P(w_i|\theta_R)$ are determined according to Equation 3. This interpolation approach naturally includes the original image-tag only approach when the parameter λ is set to $\lambda = 1.0$. Conversely, at $\lambda = 0.0$, the image tags of an image are ignored an only the terms extracted from the user's tweets are used to estimate $P(\theta_R|T_{\Im}, \mathcal{W}_{\Im})$.

We chose this interpolation approach due to its simplicity. More elaborate combinations of image tags and tweet terms will be investigated in future work. Note that the term filtering steps (Sec. 3.1.1), if employed, are now not only performed on terms $T_{\Im}$ but also on $\mathcal{W}_{\Im}$. This means effectively, that only a small number of terms that originate from a user's Twitter stream will be included in $\mathcal{W}_{\Im}$.

3.2.2 Region Priors

Instead of a uniform prior, we investigate two informative priors, one based on population information and one based on climate information.

In the case of the population based prior, which we base on the population density of a region, we hypothesize that images are more likely to be taken in areas of high population density than in areas of low population density. This intuitively makes sense, we would expect that people take more images in and around New York City than in the Great Victoria Desert of Australia[5], simply because a greater number of people live or visit the former than the latter. Given a set of places (towns, cities) $\wp = \{p_1, p_2, .., p_m\}$ with their geographic location and number of inhabitants (population pop), for each region R, we sum up the population of all places that are situated within R as well as of those places that are situated within a distance of 100km (according to the Haversine formula hs) of the region's center R_c. To avoid undue influence of major cities (with millions of inhabitants), we use the square root of the population. This yields the following population prior (Z is the normalization constant):

$$P_{pop}(\theta_R) = \frac{1}{Z} \times \left(1 + \sum_{\substack{p \in R \lor \\ hs(p, R_c) \leq 100km}} \sqrt{pop_p} \right).$$

As a second prior, we exploit climate data. Our assumption is that it is more likely that people live in or travel to regions with a moderate climate than regions of extreme weather conditions. We derived the climate prior by collecting a number of geo-tagged images and recording the average temperature in the region where the image was taken at the

#**Training data** (in total)	3,195,474
#**Training data** with acc. ≥ 11	2,974,635
#**Terms** (data with acc. ≥ 11)	21,557,411
#**Unique terms** (data with acc. ≥ 11)	1,095,561

Table 1: Overview of training data.

month the image was taken. We then fitted a Weibull distribution[24][6] to the data; this resulted in a prior which, given the climate (av. temperature) in a region R at month m returns the prior probability of an image being taken in R.

4. EXPERIMENTAL SETUP

In Sec. 4.1 we first describe the training data and the setting of the parameters used to derive the language models of the world regions. Then, in Sec. 4.2 we introduce our methodology for deriving the test images and the Twitter messages belonging to the users who uploaded these images.

4.1 Training Data

We rely on a publicly available Flickr data set for training: the MediaEval 2011 Placing Task data set [20]. The MediaEval Placing Task data set is a benchmark that tackles the same question we address in this work, namely the estimation of the geographic location of an uploaded Flickr item. It consists of approximately 3.2 million Flickr images and 10,000 Flickr videos. Meta-data is provided for all training images and videos, including their geo-location as well as the accuracy with which they were geo-tagged. There are 16 accuracy levels, ranging from 1 (world level) to 16 (street level). We utilize all training items with accuracy 11 (city level) or higher. We indexed[7] all available tags (available for the images and videos) and title terms (videos only). Each image/video is considered to be a document and the tags and title terms are considered to be the document's content.

In preliminary experiments we found that an index which is neither stemmed nor has stopwords removed yields the best results. Table 1 provides the basic statistics of the index created from the training data.

4.2 Test Data

Although the MediaEval benchmark also provides a test set (a set of Flickr items for which to estimate their latitude/longitude), we created our own for two reasons:

- It is very difficult to determine for arbitrary Flickr users whether they have a Twitter account unless they provide it in their Flickr user profile.

- To investigate our approach, we need to consider a user's tweets for a substantial period of time as the average user does not take photos every day. Due to the limits imposed by the Twitter API, in our investigation we cannot for an arbitrary Twitter user retrieve all the user's tweets at once, only the most recent ones are available.

Taking the above limitations into account we derived our test data by considering an existing long-term Twitter data set (TWD) and finding matching Flickr accounts. In previous work [2] the TWD data set was created by monitoring

---

[5]On February 11, 2012, a search on http://flickr.com with the query "New York City" returned 3,672,535 results while the query "Great Victoria Desert" returned 115 images.

[6]We chose this distribution as empirically it yielded the best fit to the data.

[7]For indexing and retrieval we utilized the Lemur Toolkit, http://lemurproject.org/.

the tweets posted by $21,102$ users over a period of eleven months (November 2010 to September 2011). Having access to such a long term profile of users suits our purposes very well. The question then remains for us, how to determine the Flickr accounts of these users (if they have one).

We approached this issue in two ways: (i) by performing a crawl of the social Web aggregator site FriendFeed and (ii) by manually assessing a subsets of the tweets posted by these users.

First, we conducted a crawl of FriendFeed[8], a social Web aggregator website, which allows users to create a unified feed of the entries on the social Web across a multitude of platforms including Twitter and Flickr. We started the extraction process with one highly connected FriendFeed user and crawled the profiles of all his subscribers and subscriptions. This process was conducted recursively, until no further profiles were discovered. In total, we found $444,226$ profiles with at least one public entry in the feed. Of those users, 14.69% have listed in their profiles (among others) at least one Twitter and one Flickr account. We matched these discovered FriendFeed users with a Twitter and Flickr account against the users in the TWD data set and found 131 FriendFeed profiles that listed one of the TWD Twitter accounts with at least 10 public tweets and 5 or more publicly accessible images in their Flickr stream.

As a second strategy, we consider the tweets of the TWD data set, accumulated within the first five month of monitoring and extracted all tweets containing the string "flickr.com". Our intuition is, that Twitter users who tweet a link to Flickr, are likely to tweet about their own Flickr account. Since this is an overly strong assumption, one of the paper's authors judged for each tweeted link whether it indeed links to the user's Flickr account. This could either be inferred from the text (e.g.,"Check out the pics of my recent trip to France") or from the similarity of the Twitter profile and the Flickr profile (same avatar image, the same or similar user name, etc.). Among the tweets we found $3,260$ with a "flickr.com" string[9]. After manually investigating them, we found a total of 135 users in TWD that had tweeted their Flickr account and had posted at least 10 tweets and 5 images to Flickr.

The profiles found through these two approaches have little overlap. In total, our test data set consists of 252 users of whom we have recorded their Twitter as well as Flickr account. For these 252 users, we have collected a total of $1,892,168$ tweets over the eleven month period. From the users' Flickr accounts we collected a total of $153,773$ images, after filtering out batch uploads[10]. Not all images were taken within the eleven month time frame of our TWD data set and not all images contain latitude/longitude values. Table 2 shows the split of the Flickr image data: there are $7,477$ images which were taken within the time frame and have a latitude/longitude of sufficient accuracy. Since for our evaluation we can only rely as ground truth on images for which we have a geo-location, we used these 7477 images as our test set of images.

#Non-batch images	153,773
of those:	
#Images within Nov '10-Sept '11	27,879
#Images with latitude/longitude	30,951
#Images within Nov '10-Sept '11 & latitude/longitude	7,477

Table 2: Flickr data set.

4.3 Parameter Settings & Runs

The following subsections describe the parameter settings and runs we investigate. We tuned the parameters for our experiments in a grid-search manner on the $5,347$ test items that were provided for the MediaEval 2011 benchmark. There is no overlap to our test set. We found the following parameter settings to yield the best results with respect to the median distance error metric: indexing without stopword removal and without stemming, Dirichlet smoothing with $\mu = 10000$, $U = 2$ (filtering terms used by less than two users in the training data), $\theta_{geo} = 0.1$ (filtering terms with a geo-spread score above 0.1), $\ell_{split} = 5000$ and $\ell_{lat\_lng} = 0.01$. Apart from the geo-filtering, these settings were fixed in all experiments reported in the result section. For the sake of showing the influence of geographic scope filtering, we present results for both types: with geo-filtering and without geo-filtering.

In this setup, there are a total of 1786 non-empty world regions. The maximum extent in terms of latitude and longitude are 22.5 and 45.0 degrees respectively, in areas of the world map where the training data is extremely sparse. The most frequently occurring location in the training data (2834 items) was found to be at latitude/longitude $40.7/-73.9$ (New York City, USA).

Baselines. We report the results for two runs: $BaseLine_{geo}$ and $BaseLine$. The first is our primary baseline (geographic spread filtering is employed), while in the latter case, the terms in $T_{\Im}$ are not geo-filtered. In both cases, the mixture parameter is $\lambda = 1$ (Equation 4), i.e. only the image tags ($T_{\Im}$) are utilized to estimate the images' location.

Terms Extracted from Tweets. For the Twitter-based approach (Sec 3.2.1), we extract all tweets from a given user's Twitter stream that were posted $\pm d$ days within the day the test image was taken on. For $d = 0$, we only consider tweets taken at the same day as the image's creation date and we label that run *SameDay*. We also tested $d = \{1, 2, 3, 5, 10, 20, .., 50\}$. Due to space constraints we only report the results for $d = \{2, 10, 20, 50\}$ (labelled $\pm$ *2 Days* for $d = 2$ and so on). Recall, that for instance $d = 2$ means that tweets of a total of five days are considered: the two days before the image was taken, the image creation date day itself and the two following days.

The extracted tweets are pre-processed: re-tweets are removed (we consider them as useful for inferring something about a user's interests but not about a user's travels or actions), URLs and *@names* are stripped. Apart from these steps, no further processing is performed. In particular, to align the Twitter-based $\mathcal{W}_{\Im}$ with our training data index, we performed neither stemming nor did we remove stopwords. If a user posts no tweets within the d days before/after the image's creation date, $\mathcal{W}_{\Im}$ will be empty.

	λ	Geo-Filter	1km	10km	50km	100km	1000km	ME km	#Better	#Worse	#Same
BaseLine$_{geo}$	1.0	yes	7.24%	34.96%	48.64%	52.67%	61.35%	60.97	–	–	–
BaseLine	1.0	no	7.46%	27.87%	34.87%	36.32%	42.29%	2513.15	2521 $\Delta_{\overline{km}}=-1715$	3017 $\Delta_{\overline{km}}=+4941$	1939
Informative Priors											
Population (Pop)	1.0	yes	7.10%	34.72%	48.40%	52.59%	70.43%	62.17	1015 $\Delta_{\overline{km}}=-4800$	252 $\Delta_{\overline{km}}=+248$	6210
Population	1.0	no	7.65%	28.25%	35.27%	36.55%	44.66%	2217.60	2692 $\Delta_{\overline{km}}=-1986$	3051 $\Delta_{\overline{km}}=+5276$	1734
Climate	1.0	yes	7.25%	34.98%	48.54%	52.54%	61.16%	60.97	460 $\Delta_{\overline{km}}=-2045$	48 $\Delta_{\overline{km}}=+4783$	6969
Terms Extracted from Tweets Only											
±2 Days	0.0	yes	4.25%	16.91%	25.21%	28.97%	41.78%	1974.25	2360 $\Delta_{\overline{km}}=-4318$	4446 $\Delta_{\overline{km}}=+5402$	671
±2 Days	0.0	no	1.82%	4.47%	6.42%	8.40%	23.63%	5717.71	1557 $\Delta_{\overline{km}}=-3573$	5611 $\Delta_{\overline{km}}=+5836$	309
Interpolation: Image Tags and Terms Extracted from Tweets											
SameDay	0.8	yes	9.21%	37.68%	54.09%	59.32%	70.27%	22.04	2064 $\Delta_{\overline{km}}=-4180$	742 $\Delta_{\overline{km}}=+2089$	4671
±2 Days	0.8	yes	9.03%	38.16%	54.66%	59.93%	71.18%	20.13	2180 $\Delta_{\overline{km}}=-4055$	848 $\Delta_{\overline{km}}=+1914$	4449
±10 Days	0.8	yes	8.53%	36.83%	53.98%	59.26%	70.74%	23.36	2355 $\Delta_{\overline{km}}=-3678$	1143 $\Delta_{\overline{km}}=+1967$	3979
±20 Days	0.8	yes	8.31%	36.72%	53.64%	58.84%	70.75%	27.34	2469 $\Delta_{\overline{km}}=-3681$	1353 $\Delta_{\overline{km}}=+2001$	3655
±50 Days	0.8	yes	8.48%	36.39%	51.97%	56.91%	68.02%	33.97	2464 $\Delta_{\overline{km}}=-3368$	1695 $\Delta_{\overline{km}}=+2340$	3320
Combining Informative Priors with the Interpolation Approach											
±2 Days+Pop	0.8	yes	9.00%	37.87%	54.59%	59.62%	75.98%	20.92	2520 $\Delta_{\overline{km}}=-4245$	959 $\Delta_{\overline{km}}=+1756$	3998

Table 3: Overview of the results across different setups. The column *Geo-Filter* indicates whether the experiment was run on the geo-filtered term set only (yes) or on all terms (no) from $T_{\Im}$ and/or $\mathcal{W}_{\Im}$. Column 2 contains the interpolation parameter λ. Three evaluation measures are reported: (i) the percentage of test images within $\{1, 10, 50, 100, 1000\}$km of the ground truth location in columns 4-8, (ii) the median error (ME) in kilometres in column 9, and, (iii) the number of test images that were located with better, worse or the same accuracy (columns 10-12) as *BaseLine$_{geo}$*. For the test images with better/worse accuracy than *BaseLine$_{geo}$*, we provide the average change in error distance in kilometres $\Delta_{\overline{km}}$ - it is negative when the error decreases and positive when the error increases. In total, the test set consists of 7477 images. In gray, improvements over *BaseLine$_{geo}$* are shown for columns 4-9 (note that in column 9, a smaller error distance indicates a better performance).

Two types of runs are investigated, those with mixture parameter $\lambda = 0$ (only information from the tweets is utilized to estimate the image's location) and those with $0 < \lambda < 1$. For the bulk of the interpolation experiments we set $\lambda = 0.8$.

Region Priors. For the population based region prior, labelled *Population*, we utilize population data available from GeoNames[11]. It contains an overview of the latitude, longitude and population numbers for a total of $113,816$ towns and cities across the world (places with more than $1,000$ inhabitants). For the climate based region prior (labelled *Climate*), we collected $100,000$ geo-tagged images from Flickr (no overlap with our training/test data) and determined the average temperature in the region on the month the image was taken by relying on global historical climate data available from the National Climatic Data Center[12]. It provides data from 7278 weather stations around the world. The majority of those images were taken at temperatures between

$10°$ and $28°$ Celsius. The lowest and highest temperature found were $-38°$ and $37°$ Celsius respectively. Since the Weibull distribution is only defined for positive values, we shifted the temperatures by $+50$ which resulted in a maximum likelihood fit of the Weibull distribution with parameters $\lambda = 68$ and $k = 9.578$.

4.4 Evaluation Measures

The accuracy of estimating the geographic location of images is commonly evaluated by reporting the percentage of test items whose estimated location is within a distance of x kilometres (km) according to the Haversine formula from the ground truth geographic location of the image. As in [21, 23], we report the accuracy for different ranges of x: $\{1, 10, 50, 100, 1000\}$km. We also report the median error distance (ME) in kilometres, that is, the error distance that no more than half of the test items exceed.

Additionally, we introduce two more measurements: to offer a comparison to the primary baseline *BaseLine$_{geo}$* which we attempt to improve upon, we report the number of test images for which the adapted approach performed better (i.e. the error distance decreased), worse (the error distance

[11]The data is available at http://download.geonames.org/export/dump/cities1000.zip, accessed 09/2011.
[12]The data is available at ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/, accessed 09/2011.

increased) or exactly the same as the baseline. We also determine the average change in error distance in kilometres ($\Delta_{\overline{km}}$) for the test items that perform better or worse than $BaseLine_{geo}$. A positive change indicates that the error distance *increases* on average by $\Delta_{\overline{km}}$ kilometres, while a negative change indicates that the error distance *decreases* on average by $\Delta_{\overline{km}}$ kilometres.

5. RESULTS

We first discuss the main results and then turn to investigation particular parameters and particular subsets of the test images.

The main results of our experiments are reported in Table 3. For each run, the interpolation parameter is given and it is indicated whether geographic spread based filtering was employed. Shown in gray are improvements over $BaseLine_{geo}$ for the distance cut-off evaluation measures and the median error.

To give an example of how to interpret the numbers, we compare the two baselines now in detail: both, $BaseLine$ and $BaseLine_{geo}$ neither include informative priors nor information derived from tweets (thus, $\lambda = 1$). When comparing the results across the various distance cut-offs, the influence of the geo-spread filtering on the accuracy becomes visible. While at the 1km cut-off there is little difference between the two runs, at the 50km cut-off already, $BaseLine_{geo}$ outperforms $BaseLine$ considerably with 48.6% accuracy compared to 34.9% of $BaseLine$. The results are similar for the median error distance in kilometres: while for $BaseLine_{geo}$, the median error is 61km, the median error increases to 2513km for $BaseLine$. However, not all test images are estimated with a greater accuracy when geo-filtering is employed. In fact, for 2521 test images $BaseLine$ yields a more accurate estimate than $BaseLine_{geo}$. Conversely, for 3017 test images, $BaseLine_{geo}$ yields better estimation than $BaseLine$. These numbers do not explain the big difference observed in median error though; for this reason, we also report the average change in error distance $\Delta_{\overline{km}}$. For the 2521 test images, $BaseLine$ outperforms $BaseLine_{geo}$, the error decreases on average by 1715km, whereas for the 3017 test images where $BaseLine$ performs worse than $BaseLine_{geo}$, the error increases by 4941km on average.

Informative Priors. Next, let us consider the impact of the informative priors in Table 3. Since these priors are determined over regions (which may span hundreds of kilometres), we do not expect an impact in accuracy in the 1-100km error distance cut-offs. If we however consider the accuracy within 1000km, the population based prior improves considerably over the baseline: while $BaseLine_{geo}$ achieves 61.4% accuracy, adding a population based prior yields 70.4% accuracy. For the 1015 test images that yield a decreased error with this prior, the error distance decreases on average by 4800km, while for the 252 test images that yield a larger error in this setup, the error distances increases on average by only 248km. The climate prior on the other hand has very little effect overall. One reason could be, that monthly average temperatures do not not provide a signal that is fine-grained enough.

Terms Extracted From Tweets. Let us now turn to the main research question of this paper: can the accuracy of estimating the geographic location of images be improved when taking the user's utterances on Twitter in the same time period into account? We first discuss a set of experiments where the information provided in the image tags was ignored and only the terms extracted from the user's Twitter stream is exploited, i.e., $\lambda = 0$. We report the results for $d = 2$, as this setting resulted in the highest accuracy in our experiments. We present the results both for runs with and without geo-filtering employed. When geo-filtering is not applied to $\mathcal{W}_{\Im}$, the accuracy of the estimation is extremely poor with a median error of 5718km. When geo-filtering is employed, the results improve, though the median error is still high at 1974km compared to $BaseLine_{geo}$. These results can be explained by the fact that users do not exclusively tweet about their personal experiences and actions but may also comment on news. We note, that again, there are test images which benefit from relying on Twitter information instead of image tags: 2360 test images have a lower error distance when using Twitter terms compared to $BaseLine_{geo}$.

While relying on Twitter alone to estimate an image's location overall does not yield an improved accuracy, the information contained in a user's Twitter stream is very helpful when combined with the textual meta-data obtained from the image. Consider the results that interpolate both sources with $\lambda = 0.8$. Across a range of time periods, the accuracy in all distance cut-offs increases and the median error decreases with respect to $BaseLine_{geo}$. Overall, the best setting is to consider tweets that were posted up to two days before and after the time the image was taken at: at the 1km cut-off, the accuracy improves from 7.2% to 9%, at 50km the accuracy improves from 48.6% to 54.7% and at the 100km cut-off the accuracy increases again from 52.7% to 59.9%. The median error on the other hand decreases sharply from 61km to 20.1km. Including tweets from a wider time interval (more than two days before and after the image was taken), decreases the accuracy again, though the results are still outperforming $BaseLine_{geo}$. When considering the split of the test images into better/worse/same as $BaseLine_{geo}$, we notice that the inclusion of a wider set of tweets increases the number of test images that either benefit or are harmed by the inclusion of the Twitter data. This is to be expected, since not all images had tweets that were posted within d days of the time the image was taken.

Combining Priors and Twitter Terms. Finally, we consider a combination of the well performing population density based prior and the best performing Twitter setup ($d = 2$, $\lambda = 0.8$). The result of this combination is shown in the last row in Table 3. Recall, that we can only expect the prior to influence the accuracy at the 1000km cut-off. The results show that indeed at the 1000km cut-off using the informative prior increases the accuracy, from 71.2% (run ± 2 Days) to 76%.

In Figure 1 we present an overview of the total number of *unique* terms used to estimate the locations of the 7477 test images. In case of $BaseLine$, for instance, all 5600 unique terms found as tags within the test images are utilized, whereas in the case of $BaseLine_{geo}$, 466 unique terms remain after employing geo-filtering. Shown is also the development of including Twitter terms from various days: the vast majority of terms are considered to be non-geographic and thus they do not play a role. The number of geographic terms for instance for ± 2 *Days* is 1983, compared with $41,101$ unique terms that are not geographic in scope (this high number is the result of neither employing stemming nor stopwording).
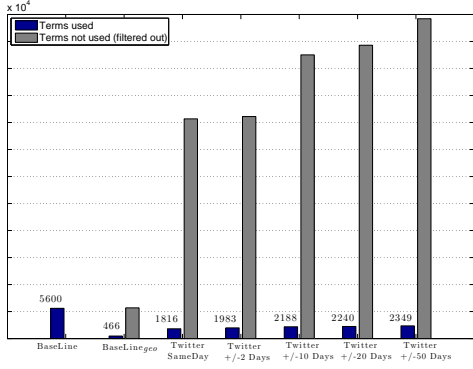
Figure 1: Number of *unique* terms appearing in the tags of the 7477 test images.

5.1 Interpolation Parameter λ

In the results just described, we had fixed $\lambda = 0.8$ for the interpolation experiments. We chose this value as it gave robust results across a range of settings. Now, we only consider the best performing setup ($\pm$ *2 Days*) and plot the results when λ is varied from 0 (no Twitter information is used) to 1 (no image information is used) in steps of 0.1. In Figure 2 we show the development of the accuracy for 1km and 50km. Please note that the two y-axes are on different scales: on the left (right), the results for the 1km (50km) distance cut-off are shown. The horizontal lines indicate the accuracies of $BaseLine_{geo}$ ($\lambda = 1$) for both cut-offs.
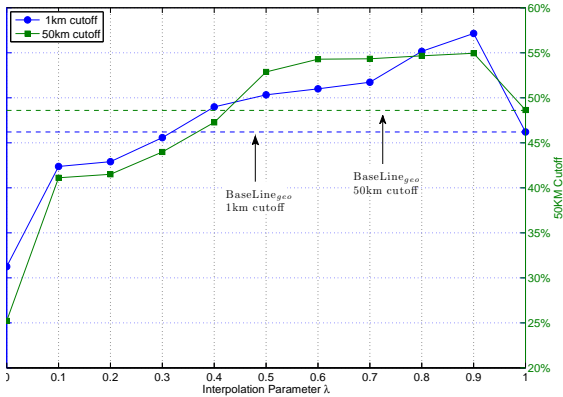

Figure 2: Interpolation parameter λ varied from $[0, 1]$ in steps of 0.1. Results are shown for ± 2 *Days* with $\lambda = 0.8$ and geo-filtering employed. Note, the y-axes are on different scales.

The increase in accuracy is similar for both evaluation measures, at $\lambda > 0.3$ (1km cut-off) and at $\lambda > 0.4$ (50km cut-off), the interpolation approach outperforms the baseline. For this particular setup, $\lambda = 0.9$ yields the most accurate results with 55% of the test images being placed within 50km of their true location (compared to 49% in $BaseLine_{geo}$).

5.2 What Images Benefit from $\mathcal{W}_\Im$?

Finally, we turn to the question of what images do (not) benefit from adding information from the user's Twitter stream. We consider two dimensions: (i) the number of tags
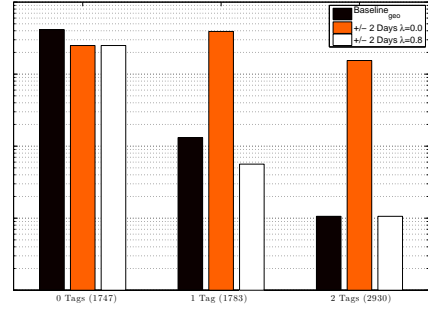

Figure 3: Test set split according to the number of tags $T_\Im$ *after* employing geo-filtering. Compared are $BaseLine_{geo}$ ($\lambda = 1$), ± 2 *Days* with $\lambda = 0$ and ± 2 *Days* with $\lambda = 0.8$. In brackets, the size of each partition is shown.

in $T_\Im$ after applying geographic scope filtering, and, (ii) the distance of the image from the user's home location.

Number of Tags. To investigate the impact of the number of tags on the estimation accuracy, we plotted the median error in kilometres for different settings in Figure 3 for the three cases of a test image having 0, 1 or 2 tags in $T_\Im$. Please note, that the plot is on a logarithmic scale (y-axis) and that we consider the size of $T_\Im$ *after* having employed geographic scope filtering.

Let us first focus on $BaseLine_{geo}$. As can be expected, if no tags are present, the median error distance is highest (with 4142km), as either the most frequent location in the training data is assigned to the test image or, if available, the user's home location is added to $T_\Im$. Since the given home location may be geographically ambiguous (.e.g.,"Perth" may refer to a city in Australia or the UK), underspecified (e.g.,"USA") or plainly inaccurate (e.g.,"Mars"), there is often little additional value in it. If one tag is available, the median error distance decreases to 131km, at $|T_\Im| = 2$, the median error has fallen to 11km. The interpolation approach ± 2 *Days* with $\lambda = 0.8$ improves over $BaseLine_{geo}$ if $T_\Im$ contains either zero or one tag only. At zero tags, the median error distance decreases by 40%, for the test images with one tag, the median error decreases by 57%.

For the cases where $|T_\Im| \geq 3$, adding information from a user's Twitter stream does not yield any more improvement in accuracy. This confirms our original hypothesis: extracting terms from a user's tweets is valuable for images which contain less than three tags (after geographic filtering).

Distance to Home Location. In [4] it was found that a user's home location can be estimated based on the user's tweets. In our experiments we found that even if tweets from a large number of days (± 50) around the test images are used to extract terms for $\mathcal{W}_\Im$, the results still improve over $BaseLine_{geo}$. One potential explanation is that this Twitter based information enables us to mostly estimate those images, that are close to a user's home location with greater accuracy. To evaluate this hypothesis we conducted the following analysis. We manually transcribed/disambiguated the home location that the users who contributed the test images gave on their Flickr profile and description fields and converted them into latitude/longitude. For 4515 of the test images, we could determine the corresponding user's home lo-
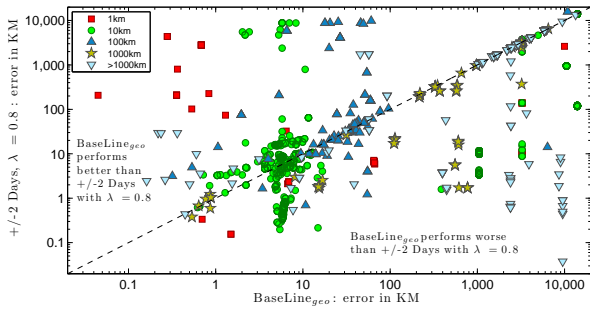
Figure 4: Scatter plot with test data split according to the distance, the images were taken at with respect to the corresponding user's home location. Each marker stands for one image.

cation with city level accuracy. We then removed all images where $BaseLine_{geo}$ and the run ± 2 Days ($\lambda = 0.8$) yield the same error distance. The remaining 1330 images were partitioned according to how far away from the user's home location they were taken: within $\{1, 10, 100, 1000, > 1000\}$km of the user's home location.

In Figure 4 we show a scatter plot for these 1330 images: the error in kilometres at $BaseLine_{geo}$ vs. the error at ± 2 Days ($\lambda = 0.8$). The images belonging to each partition are marked with different markers/colors. Many images were taken within 100km of the respective users' home location. While for these types of images, adding information from a user's Twitter stream sometimes improves and sometimes decreases the error, a clearer pictures emerges for images that were taken a greater distance from the user's home location: the the vast majority of images improve when utilizing information extracted from Twitter.

6. CONCLUSIONS

In this paper, we have presented an approach to image location estimation that exploits a user's traces on the social Web across platforms. We hypothesized that information extracted from a person's Twitter stream can be utilized to improve the accuracy of the image location estimation. Our investigation was motivated by the fact that people often tweet about themselves, and it is likely that they post messages which refer to their activities or travels. Furthermore, we investigated the usage of informative priors, based on population data and on climate data. We developed an extension to an existing state-of-the-art tag-based approach to location estimation [21] which is based on the language modeling framework for information retrieval. Our extension makes it possible to exploit the information extracted from the user's tweets in a natural way. In this work, we presented a number of findings:

- Exploiting a person's Twitter stream when estimating the location of an image decreases the median error distance by up to 67% in comparison to a state-of-the-art approach. The best results are obtained when tweets within two days before and after the image was taken are exploited.

- Images, that were assigned less than two tags after geographic scope filtering (Sec. 3.1.1), benefit the most from terms extracted from Twitter. Specifically, we showed a

40% (57%) decrease in median error distance for images without tags (with a single tag).

- We also investigated the value of informative priors in the language modeling approach. The population density prior improves the accuracy of the location estimation.

These results leave a lot of potential for future work. One possibility is to not only consider the traces a user leaves on Twitter, but also on other social Web platform such as Facebook or LinkedIn. Another area of future work is to classify users' tweets as either being about personal experiences and activities or about local vs. global news: tweets discussing global news (e.g. *"Japan announced meltdown yesterday; situation grim."*) are often adding noise for our purposes, while tweets discussing local news (e.g. *"Here in Toronto the police made multiple arrests today"*) are useful as they can serve as indicators of the user's home location. Lastly, we also consider an investigation into the similarities/differences between (estimated) travel patterns of different users and user groups an interesting avenue of further work.

7. REFERENCES

[1] F. Abel, S. Araújo, Q. Gao, and G.-J. Houben. Analyzing cross-system user modeling on the social web. In *ICWE '11*, pages 28–43, 2011.

[2] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on Twitter for personalized news recommendations. In *UMAP '11*, pages 1–12, 2011.

[3] F. Abel, N. Henze, E. Herder, and D. Krause. Interweaving public user profiles on the web. In *UMAP '10*, pages 16–27, 2010.

[4] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *CIKM '10*, pages 759–768, 2010.

[5] J. Choi, A. Janin, and G. Friedland. The 2010 ICSI Video Location Estimation System. In *MediaEval 2010 Workshop*, 2010.

[6] M. Clements, P. Serdyukov, A. P. de Vries, and M. J. T. Reinders. Finding wormholes with flickr geotags. In *ECIR '10*, pages 658–661, 2010.

[7] M. Clements, P. Serdyukov, A. P. de Vries, and M. J. T. Reinders. Using flickr geotags to predict user travel behaviour. In *SIGIR '10*, pages 851–852, 2010.

[8] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu. Automatic construction of travel itineraries using social breadcrumbs. In *HT '10*, pages 35–44, 2010.

[9] C. Hauff and G.-J. Houben. Geo-location estimation of flickr images: Social web based enrichment. In *ECIR '12*, pages 85–96, 2012.

[10] T. Hofmann. Probabilistic latent semantic analysis. In *UAI âĂŹ99*, pages 289–296, 1999.

[11] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff. Identifying users across social tagging systems. In *ICWSM '11*, 2011.

[12] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *9th WebKDD / 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis*, pages 56–65, 2007.

[13] P. Kelm, S. Schmiedeke, and T. Sikora. Multi-modal, multi-resource methods for placing flickr videos on the map. In *ICMR '11*, pages 52:1–52:8, 2011.

[14] S. Kinsella, V. Murdock, and N. OHare. "i'm eating a sandwich in glasgow": Modeling locations with tweets. In *SMUC '11*, pages 759–768, 2011.

[15] V. Lavrenko and W. B. Croft. Relevance based language models. In *SIGIR '01*, pages 120–127, 2001.

[16] X. Lu, Y. Pang, Q. Hao, and L. Zhang. Visualizing textual travelogue with location-relevant images. In *LBSN '09*, pages 65–68, 2009.

[17] M. Lux and S. A. Chatzichristofis. Lire: lucene image retrieval: an extensible java cbir library. In *MM '08*, pages 1085–1088, 2008.

[18] M. Naaman, J. Boase, and C.-H. Lai. Is it really about me?: message content in social awareness streams. In *CSCW '10*, pages 189–192, 2010.

[19] J. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR '98*, pages 275–281, 1998.

[20] A. Rae, V. Murdock, P. Serdyukov, and P. Kelm. Working Notes for the Placing Task at MediaEval 2011. In *MediaEval 2011 Workshop*, 2011.

[21] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *SIGIR '09*, pages 484–491, 2009.

[22] O. Van Laere, S. Schockaert, and B. Dhoedt. Combining multi-resolution evidence for georeferencing Flickr images. In *SUM '10*, pages 347–360, 2010.

[23] O. Van Laere, S. Schockaert, and B. Dhoedt. Finding locations of flickr resources using language models and similarity search. In *ICMR '11*, pages 48:1–48:8, 2011.

[24] W. Weibull. A statistical distribution function of wide applicability. *J. Appl. Mech.-Trans. ASME*, 18(3):293–297, 1951.

[25] B. P. Wing and J. Baldridge. Simple supervised document geolocation with geodesic grids. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, pages 955–964, 2011.

[26] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *SIGIR '01*, pages 334–342, 2001.

[27] C. Zhai and J. Lafferty. Two-stage language models for information retrieval. In *SIGIR '02*, pages 49–56, 2002.

[28] D. Zhao and M. B. Rosson. How and why people twitter: the role that micro-blogging plays in informal communication at work. In *GROUP '09*, pages 243–252, 2009.