Placing Flickr Photos on a Map

Pavel Serdyukov *†
Database Group
University of Twente
PO Box 217, 7500 AE
Enschede, The Netherlands
serdyukovpv@cs.utwente.nl

Vanessa Murdock
Yahoo! Research
Diagonal 177
08018 Barcelona, Spain
vmurdock@yahoo-inc.com

Roelof van Zwol
Yahoo! Research
Diagonal 177
08018 Barcelona, Spain
roelof@yahoo-inc.com

ABSTRACT

In this paper we investigate generic methods for placing photos uploaded to Flickr on the World map. As primary input for our methods we use the textual annotations provided by the users to predict the single most probable location where the image was taken. Central to our approach is a language model based entirely on the annotations provided by users. We define extensions to improve over the language model using tag-based smoothing and cell-based smoothing, and leveraging spatial ambiguity. Further we demonstrate how to incorporate GeoNames[1], a large external database of locations. For varying levels of granularity, we are able to place images on a map with at least twice the precision of the state-of-the-art reported in the literature.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]

General Terms

Algorithms, Measurement, Performance, Experimentation

Keywords

image localisation, language models, Flickr

1. INTRODUCTION

Due to the massive production of affordable GPS-enabled cameras and mobile phones [13, 16], location metadata such as *latitude* and *longitude* are automatically associated with the content generated by users. Users have the opportunity to spatially organise and browse their personal media, and photo sharing services are leading the growing enthusiasm for personal location-awareness [22]. Geo-referenced photos can be organised in a browsable taxonomy of major locations or pin-pointed on a map to identify very small regions. Some of the most popular examples are Flickr Places[2] and Google Panoramio.[3]

While in theory every photo can be anchored to the location it was taken, in practice many photos are location agnostic. Furthermore, the majority of Flickr users do not own location-aware cameras. Thus a large proportion of photos uploaded to Flickr contain no location information even when the photo merits localizing. When uploading photos on Flickr users can still geo-tag their photos by dragging the photos to a particular point on the world map. This process is time-consuming and results in less accurate geo-tagging of photos compared to automatically geo-tagged photos from GPS-enabled cameras. When manually geo-tagging photos, Flickr initially suggests the location of the last uploaded photo or simply displays the world map.

The objective of this paper is to provide a more accurate starting point for geo-tagging photos, uploaded on Flickr, using the textual annotations provided by the user. According to recent literature [2, 21] users spend considerable effort to organise their "memory" geographically by describing photos with *tags* related to locations where they were taken. The location specific tags (such as *Torre Agbar* which is only located in Barcelona), and location related tags (such as *elephants* which are related to locations such as zoos, Africa and Asia) provide essential cues as to where a picture was taken. For photos that are location agnostic (such as *dog*), location information may or may not be provided, but it is normally not relevant to the context of the photo.

The literature related to geo-tagging of photos and its use is extensive. In particular the reverse problem of discovering important landmarks and events, given a geographic co-ordinate has been studied extensively [1, 17, 13]. However the problem of placing images on a map using the textual annotations provided by the user has received less attention. While we focus on Flickr as our primary application, our approach can be applied to a wide range of service providers dealing with geo-referenced resources.

In this paper we investigate generic methods for placing photos uploaded in Flickr on the world map. We construct an $m \times n$ grid based on the longitude/latitude co-ordinates of the globe, where each grid cell represents a location. Using a set of images whose locations are known, we place each image in its corresponding grid cell. As a baseline we employ the collective knowledge of Flickr users by es-

*Research performed while the author was an intern at Yahoo! Research.
†Also affiliated with TU Delft, ICT Group
[1]http://www.geonames.org visited May 2009

[2]http://www.flickr.com/places/ visited May 2009
[3]http://www.panoramio.com visited May 2009

SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.

timating a language model from the terms people use to describe images taken at a particular location. We extend this model in several ways, using neighbouring cells under the assumption that "good" locations come from "good" neighbourhoods, and leveraging spatial ambiguity. Finally, we investigate how to incorporate external resources into the model, by boosting the importance of known location tags identified by their presence in GeoNames. We train, develop and evaluate our system using a snapshot of nearly 400,000 geo-tagged photos from Flickr with textual annotations. We predict the single most likely location of a photo in terms of accuracy at different levels of spatial granularity.

The remainder of this paper is organised as follows. The related research on spatial data mining of user generated content and web pages is reviewed in the next section. In Section 3 we explain our approach for location representation. In Section 4 we introduce our geographic location identification framework based on language models using a bag-of-words approach, and our improvements over the baseline language model. A description of our dataset and evaluation measures can be found in Section 5. The results are discussed in Section 6. Finally, Section 7 summarizes our findings and outlines directions for future research.

2. RELATED WORK

Crandall et al [6] propose a system to place images on a map with a combination of textual and visual features, using a corpus of 20 million images crawled from Flickr. In spirit their work is similar to ours, but they limit their task to deciding which of ten landmarks in a given city is the subject of an image, whereas in our system the location of the image is completely unrestricted. They build a classifier for each of the ten landmarks in the city where the image was taken. For each of the ten classifiers the positive examples are images of a given landmark, and the negative examples are images from the other landmarks. The images are represented by vectors of features of the tags, and visual keywords derived from a vector quantization of the SIFT descriptors. Our images are represented only by the tags used to describe them, without incorporating any visual features. Furthermore, while they investigate a location granularity of 100 kilometers and 100 meters, we investigate multiple granularities. We do not assume any prior knowledge about the city or country the image was taken in, and such information may or may not be present in the tag sets. It is not clear how to scale the system described in Crandall et al. to place an image at any point on the globe, without prior knowledge of the city or country the image was from.

The work of Hayes and Efros [8] is also related to the work described in this paper. Using Flickr images, they propose visual features to predict the geographic location using a nearest-neighbour classification method. They report geolocating 16% of test images within 200 km. Their data is limited to a sub-set of Flickr images tagged with at least one name of a country, continent, densely populated city or popular tourist site and not tagged with specific non-geographic tags such as "birthday" or "concert". By contrast, our approach is knowledge-free, highly scalable and not limited to photos that are known to contain locations in the textual annotations.

The remaining related literature falls into three areas: spatial data mining of user-generated content, finding the geographical focus of Web pages, and toponym resolution.

2.1 Spatial mining of user-generated content

Working with blog data, Mei et al. [12] present methods for finding latent semantic topics and their distribution over locations (states or countries) and Wang et al. [25] propose a *Location-Aware Topic Model* based on Latent Dirichlet Allocation. The works are similar to ours in that they attempt to discern whether a topic is location-related, however blog data has a considerably richer semantic representation than the tags associated with images on Flickr, which may be only two or three terms.

Web queries are more similar to tags than blogs, in the sense that queries are two or three content terms representing much larger concepts. Backstrom et al. [4] propose a method to measure the geo-specificity of a query, using the level of dispersion around the location of the query's highest frequency. With a similar goal in mind, Zhuang et al. [28] calculate the inverse correlation of a query's click distribution over locations with their populations. Vadrevu et al. [24] use the probability of co-occurrence of a query term with place names from each region to determine queries that might be related to a given region.

In relation to the location identification of images, Ahern et al. [1] propose a method for detection of Flickr tags exhibiting spatial patterns. They find dense areas using geodesic distances between images, and rank all tags in these areas with a $tf.idf$-based feature selection measure to select the most representative location-related tags. The focus of their research is on selecting tags, rather than localizing images. In later papers they propose a method for detection of tags that correspond to local events [17]. Naaman et al. [14] look at the other side of the coin, recommending tags to the user, given a known location for an image.

2.2 Finding geographical focus of web pages

The task of finding a geographical focus of a web page was first proposed by Ding et al. [7]. Their approach was two-fold: finding locations of web pages with hyper-links to the analysed page and detection and disambiguation of toponyms in its content. Follow-on work by Amitay et al. [3] and Zong et al. [29] relied on propagating the confidence weights of found toponyms up to the root of the gazetteer taxonomy to find the most probable common ascendants (e.g. finding a *country* for several *cities* mentioned).

2.3 Toponym resolution in text

The research on finding geographical focus of text described in the previous section is almost entirely based on finding and resolving toponyms. For our task of placing images on a map, an obvious solution would be to simply resolve the toponyms in the Flickr tags. This would allow us to identify locations that are mentioned in the tags, but does not allow us to infer locations from tags that are not found in gazetteers or other resources. We briefly describe the state-of-the-art in toponym resolution. A complete and detailed description of toponym resolution heuristics is made by Leidner [10], but most approaches are derived from the following ideas.

Using location priors. Even without any context it is possible to make an intelligent guess about the most probable referent for a toponym by just considering the prior probabilities of places. For example, places with larger populations, or more frequent mentions in text are more likely candidates.

Search for disambiguators. It is assumed that each place has a list of disambiguators, such as its neighbours in gazetteer hierarchy, which resolve it if found in the proximity of the place name mention. For example, the country of *France* and the state of *Texas* are both disambiguators for the city of *Paris*.

Spatial minimality. If several toponyms are mentioned in the text without disambiguators, then those places are selected as referents that minimise the *minimum bounding rectangle* containing them or the sum of their pair-wise distances from each other. For example, if *Moscow* and *Helsinki* appear together in the text, then *Moscow, Russia* is selected instead of *Moscow, Idaho, US* since it is thousands of kilometers closer to Helsinki, and Helsinki is unambiguous.

Ranking locations by the probability of generating a tag set implies leveraging the above-mentioned approaches. Both spatial minimality and disambiguator-based methods are implicit in the language modeling approach because places are more likely to appear together with their disambiguators in tags and tag sets from the same location. A significant number of users enlist disambiguators for place names in their tag sets by mentioning not only a city, but also a country, for example. However, as demonstrated in Figure 1, in addition to toponyms, other region-specific terms traditionally unnoticed by gazetteers can serve as disambiguators (e.g. *hermitage* for *St. Petersburg, Russia* and *pelican* for *St. Petersburg, Florida*). Using additional population and popularity based priors for locations is also not necessary due to the origin of tags representing locations: popular and highly populated locations get more image uploads and hence higher related tag counts.

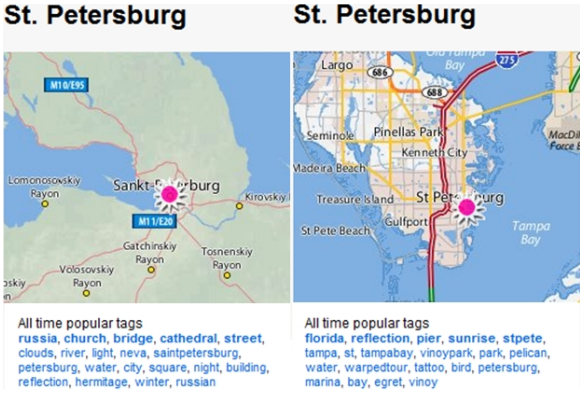

Figure 1: Tags for places with ambiguous names, generated with Flickr Places: http://www.flickr.com/places

3. REPRESENTING LOCATIONS ON THE MAP

This section explains how geo-tagged photos are associated with a particular grid-cell on a map an how a textual representations of locations is build and organised in a graph structure to model their spatial and semantic relations.

3.1 Locations as bags-of-tags

For each photo in our collection we have the following sources of information: a *FlickrID*, a *geographic co-ordinate*, and a *set of tags*. The first task is to map each geographic co-ordinate to a location on the map. For that purpose we place a grid over the world map, which allows us to define a location as a cell on this grid, described by a pair of universal geographical co-ordinates (UGC). We can then investigate the accuracy of our system given various levels of spatial granularity of the grid. To universally represent locations we bind them to a cell using their latitudes and longitudes, considering 0 - 2 digits of the decimal part (UGC are represented in Flickr as decimals). For example, each pair of co-ordinates, ignoring the decimal part and considering only *degree* units, defines a unique *location*: an approximate rectangle, or a cell of the earth grid, with longitude side of about 111 kilometers long and variable latitude side. The length of latitude side varies from 0 kilometers at the poles (but 60 starting from inhabited places) to 111 kilometers at the equator [23]. In this paper we consider locations (cells) of roughly 1, 5, 10, 50 and 100 kilometers long over latitude.

Alternatively, one can tie a geographic location to a known semantic location as defined in Gazetteers. Locations are named and either defined by a bounding box (GeoPlanet[4]), or just by a point on the map (GeoNames). In the first case, bounding boxes are often overlapping and do not cover all regions where people may appear. In the second case, it is unclear how to define boundaries in an unsupervised way. Moreover, developing methods dependent on specific symbolic representation of geographical information may lead to difficulties in maintaining such a framework, including the need to rebuild models with every new update of the underlying topology. At the same time, when the most probable geographic location of an image is determined, mapping this location onto a specific geographical ontology (*reverse geocoding*) is almost straightforward.

Each photo has associated one or more textual tags which are used to derive a language model that represents a location. In addition to tags, users may annotate their photos with titles and descriptions. Although titles and descriptions are potentially useful in combination with tags, we have chosen to limit ourselves to a tags-only representation, as they are the most compact and concise descriptions of a photo. We assume that the order of tags is not important for placing images on a map and adopt a bag-of-tags approach to sample representative tags for a given location. In order to preserve the unique semantics of each tag set we do not apply any stemming or stop-word filtering. However, we use the standard tag normalisation automatically provided by Flickr: all terms in compound tags are concatenated and all special characters are stripped.

3.2 Locations as a graph

In web retrieval it is common to assume various relations among documents defined by hyper-links. In our case we consider that the grid structure underlying our collection of locations implies a spatial relationship. Based on these observations, we represent all locations in an undirected graph, where the link between a pair of locations (grid cells) exists only if they are situated close enough on the grid. For the sake of simplicity, we use cell-based distance and consider that any cell has 8 cells situated within 1-cell distance from it, 24 cells situated within 2-cell distance etc. Those locations that are found within a predefined distance can be then linked and hence considered to be *neighbours*. In our case

[4]http://developer.yahoo.com/geo/ visited May 2009

representing locations as *pseudo-documents* implies not only spatial, but also semantic similarity. This fact should not be neglected when localising a tag set: it is easy to expect that linked locations will have high probability to be represented by similar tags and that locations relevant to a classified image will also be close in the graph.

4. MODELING LOCATIONS

In this section we describe the baseline approach for determining the location of a photo, given a set of tags. By estimating a language model through analysis of the terms people use to describe images taken at a particular location, we can predict the most likely position where this photo was taken. In other words, we are interested in obtaining a ranking list of locations L, which is ordered by the descending probability for a given tag set T belonging to an image is taken within the bounds of L:

$$P(L|T) = \frac{P(T|L)P(L)}{P(T)} \tag{1}$$

A location in our framework is represented by a multinomial probability distribution over the vocabulary of tags. Since we do not have any prior information about locations and tags that would otherwise influence the ranking, we consider that $P(L)$ is distributed uniformly and $P(T)$ does not influence the ranking. The locations are then ranked by the probability to generate the tag set of the image. Assuming that each tag t_i in the tag set T is generated independently, the tag set likelihood can be expressed as:

$$P(T|L) = \prod_{i=1}^{|T|} P(t_i|L) \tag{2}$$

$$P(t|L) = \frac{|L|}{|L|+\lambda} P(t|L)_{ML} + \frac{\lambda}{|L|+\lambda} P(t|G)_{ML} \tag{3}$$

where $P(t|L)_{ML}$ and $P(t|G)_{ML}$ are maximum likelihood estimates of tag generation probabilities for the location and for the global language models, $|L|$ is the size of the location L in tags and λ is the parameter of Dirichlet smoothing [27]. Dirichlet smoothing outperformed other kinds of smoothing in our preliminary experiments, seemingly due to its capability to decrease the influence of model lengths on ranking: other smoothing models imply the preference for smaller models, while in our case we have a large number of "tiny" models (i.e. locations containing only a few images) due to sparseness of our data. This becomes even more critical, when the granularity of the earth grid is small.

4.1 Tag-based smoothing with neighbours

The motivation to smooth from neigbourhoods of locations comes from the need to overcome data sparseness and from understanding that some tags indicate an area that exceeds the bounds of specific location. For example, even if we use very large 100 km cells, some tags specify a country or continent, which may be larger than 100 km. Moreover, some geographical objects and related tags can be situated in several locations due to the way the grid is placed on the earth surface (for example Rio de Janeiro is situated in 4 neighboring 100 km cells).

The smoothing of document class models with models for broader categories is known to be effective for hierarchical document classification [11]. We employ a similar strategy, with a slightly smaller boundary, as we use a static grid where cell size is pre-determined.

The first way to use spatial neighbourhood is to consider that each tag found within a specific location is generated by either the location's language model, or by language models of neighbouring locations:

$$P(t|L) = \mu \frac{|L| \cdot P(t|L)_{ML}}{|L|+\lambda} + (1-\mu) P(t|NB(L)) + \frac{\lambda \cdot P(t|G)_{ML}}{|L|+\lambda} \tag{4}$$

$$P(t|NB(L)) = \sum_{L' \in NB(L)} \frac{|L'|}{|L'|+\lambda} \frac{P(t|L')_{ML}}{(2d+1)^2 - 1} \tag{5}$$

where NB(L) consists of all locations included in the neighbourhood of location L, d is the minimal distance (in grid cells) between locations to be connected in our earth grid graph and μ is the smoothing coefficient on the probability that the term is generated from the initial location's language model.

4.2 Smoothing cell relevance probabilities

It is reasonable to assume that "good" locations come from "good" neighbourhoods. This means that some relevance should be propagated through the links between close locations. Similar techniques have shown themselves to be effective for the web retrieval [20] or expert finding [19]. While relevance propagation on document and entity graphs is traditionally modeled with random walks [20], we do not expect very distant nodes to have high influence on the relevance of the specific location. For these reasons and also because of computational efficiency requirements, we apply a simple *weighted in-degree* approach: the probability to generate the tag set of a certain location is augmented with the probabilities of neighbouring locations:

$$P(T|L) = \alpha P(T|L) + (1-\alpha) \sum_{L' \in NB(L)} \frac{P(T|L')}{(2d+1)^2 - 1} \tag{6}$$

Note that we are still able to include indirectly adjacent locations by setting parameter $d > 1$:

So far we have regarded our grid graph as undirected, which means that probabilities from all neighboring locations are used in equation 6. It is known from document retrieval [18, 20] that it is more efficient to propagate relevance in the hyper-link graph in the direction of more relevant documents. We propose to propagate relevance only from those locations that have lower scores than the location to be smoothed. The motivation is that it is safer to support those documents that have already enough probability to be relevant, than to make highly relevant documents support poor ones. Speaking of location retrieval, we may think of the following similar motivation. In the cases when smoothing from nearby locations helps, it succeeds not to select the best location within a certain neighborhood, which is already efficiently selected by the initial retrieval step, but to select the right "global winner" among those "local winners" from different parts of the globe. Thus, relevance propagation in the direction of "local losers" is not motivated. In graph-related terms, we make our graph query-dependent: edges between cells become directed (from lower scored to higher scored cells) and hence not all of them are used for

calculating weighted in-degree. We test propagation on both undirected and directed graph models in our experiments.

4.3 Boosting geo-related tags

It is known that users often annotate images with tags that can be recognised as location-specific from a first glance: names of places (e.g. cities or countries), points-of-interest (e.g. monuments, stadiums, hotels, or bars) or events known to happen in certain locations (e.g. festivals, sport competitions). While the geo-specificity of these tags is captured by our models, it is possible to conclude that some of these tags should be more popular near certain locations even without analysis of their spatial distribution. We introduce preliminary knowledge about tags into our models using a simple boosting approach, similar to the one recently used by Cao et al. to boost expansion terms [5]:

$$P(t|L)_{ML} = P(t|L)_{ML}(1 + \beta P(Loc|t))/Z \quad (7)$$

where $P(Loc|t)$ is a probability of the tag t to be location-specific, β is a boosting coefficient and Z is a normalisation coefficient.

For the research described in this paper, we use the list of toponyms limited to English names of populated locations, which is taken from the GeoNames database to decide whether the tag is location-specific. For all tags that are in this list the $P(Loc|t)$ equals 1.0 and otherwise equals 0. More sophisticated approaches to tag classification such as described by Overell et al. [15] also fit our boosting approach. Although the suggested method depends on external sources, we can assume that most popular gazetteers more or less agree on official names of populated places, but might differ significantly in how they define location centroids, bounding boxes and geographic ontologies.

4.4 Spatial ambiguity-aware smoothing

It is obvious that some tags are specific for more than one location: either because their scope exceeds the bounds of a single cell, or due to their ambiguity (for example *bath* and *Bath, UK* or because they are instances that are typically spotted at a few specific locations, such as *Elephants* [26]. It is intuitive to trust highly spatially ambiguous tags less than tags that have a single geographical focus. Since we know the co-ordinates of all tag instances in the training data, we are able to characterise spatial ambiguity of a tag by the standard deviation of its latitudes and longitudes $\sigma_{lat}, \sigma_{lon}$. To include this factor into our model we let the smoothing coefficient λ in Equation 3 be tag-specific and proportional to the ambiguity of a tag:

$$\lambda(t) = \lambda + \gamma(\sigma_{lat}(t) + \sigma_{lon}(t)) \quad (8)$$

where γ is a weight coefficient to control the influence of ambiguity level on smoothing. The individually generated probabilities of ambiguous tags will be less decisive for finding the most probable location for a tag set. This is especially important to prevent over-boosting the ambiguous toponyms.

5. EXPERIMENTAL SETUP

In this section we describe the experimental setup for predicting the locations where an image was taken. We will first discuss the Flickr data set, followed by the evaluation measures adopted.

5.1 The Flickr data set

We randomly sampled a set of 397,000 geo-tagged Flickr images with the associated tags. We applied a filter to remove the effect of bulk uploads by users. Users can decide to apply the same set of tags set of a large number of photos which can be taken at the same or different locations. With the filter in place, we ensure that at the smallest spatial granularity, there is at most one photo per user with the same tag set. This reduces the set of photos to 140,000 which is still larger than data sets used recently for data mining in Flickr [17, 9].

To better understand the data, we geo-referenced all images with the GeoNames gazetteer. Figure 2 (left) indicates that a third of photos originate from the U.S., and an equal number of photos relate to Europe. Overall, the collection contains photos from about 180 different countries which makes the data set more representative compared to data used recently, which focus on the U.S. [9].

We partition the data into three parts: data for building location models, data for parameter tuning, and test data. User-specific tags play a decisive role for finding locations (what would be a re-finding of one of the previous locations of the user). Despite that we expect the results to be better for the users that have their own data in the collection already, it was more important to find out whether we are able to make good predictions for unseen users and their tag sets. Therefore, to separate the data we just sorted the initial image set by unique user ids and considered roughly 120,000 (≈85%) images to build models, 10,000 (≈7%) to tune parameters and 10,000 (≈7%) to test our methods.

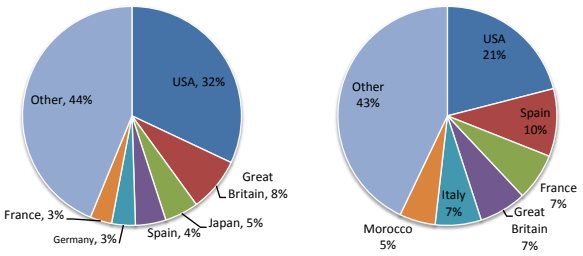

Figure 2: Images over countries: all images (left) and mapped correctly within 100 km (right)

.

5.2 Evaluation measures

The main metric that we use for the evaluation and for tuning parameters on training data is location accuracy (**Acc**), which calculates the percentage of correct predictions over all test examples. However, since we consider *location recommendation* as a likely task to benefit from our location prediction techniques, we also analyse additional measures of prediction quality:

Mean Reciprocal Rank (MRR) measures the ability of the system to find the actual location of a photo among its top recommendations.

Parent Accuracy (PAcc) determines whether the predicted location belongs to the same parent with the correct location (for instance, 100 km cells are parents for 50 km cells, 50 km cells - for 10 km cells, etc.).

Division	Acc	MRR	Acc@1	Acc@2	Acc@3	PAcc
1km	0.067	0.073	0.125	0.152	0.170	0.122
5km	0.140	0.155	0.226	0.248	0.259	0.177
10km	0.181	0.197	0.261	0.278	0.291	0.247
50km	0.256	0.277	0.332	0.354	0.378	0.289
100km	0.288	0.309	0.370	0.410	0.435	-

Table 1: Performance of the baseline LM method

Accuracy within K cells (Acc@K) computes whether the actual location is within a K-cell distance from the predicted location.

For many tag sets (e.g. *dog* or *birthday*) it is difficult, if not impossible, to predict the location of a photo due to the generic nature of the tag set. Furthermore, very generic tag sets are less likely to have an implicit geo-intent. However, when the user does have an implicit geo-intent, they expect the system to be able to properly determine the area where the photo was taken. Given that we are only working with a small sub-set of all geo-tagged photos in Flickr, we expect to suffer from data sparseness when building the language models for the smaller grid cells. The context-free nature of our language model allows us to scale to the full set of geo-tagged photos, which is likely to result in significant improvement in terms of accuracy.

6. RESULTS

We evaluated the following methods and their combinations: baseline language model **LM** (Section 4), tag-based smoothing **TS** (Section 4.1), cell-based smoothing **CS**, cell-based smoothing with score propagation in the direction of higher relevance **CSR** (Section 4.2), toponym based boosting **TB** (Section 4.3) and ambiguity-aware tag specific smoothing **AS** (Section 4.4).

All parameters for the tested methods (see Equations 3, 4, 6, 7, 8) are optimised on the held-out data using line search strategy, maximising accuracy (**Acc** in the table). After optimising λ for the baseline retrieval model, the other parameters are optimised independently.

Table 1 details the performance of our methods for the different evaluation measures at five different levels of spatial granularity (1, 5, 10, 50, and 100 km). Focusing first on the results for our baseline **LM** method, we observe that the accuracy increases, when increasing the grid size, from 0.067 to 0.288, which is consistent with our expectations. Additional performance improvement is observed when analysing the relaxed accuracy measures to include the direct neighbours of the predicted location.

Focusing on the effect of the three neighbourhood smoothing extensions **TS**, **CS** and **CSR**, we find some marginal improvements, with the CSR method outperforming the other two smoothing extensions independent of the chosen grid size, as shown in Table 2 for the grid size of 1, 10, and 100 km. Smoothing was only done with the immediate neighbours ($d = 1$ in Equations 5, 6), using larger neighbourhoods did not turn out to be beneficial.

Finally, we have tested different combinations of **LM**, **TB**, **AS** and **CSR** methods. Table 3 shows the baseline and the best performing methods from Table 2 for clarity. We first observe that all methods improve over baseline **LM** for all measures. Second, among all improvements applied alone, **TB** method shows the best performance. The information in the GeoNames knowledge base added a small

Method	Acc	MRR	Acc@1	Acc@2	Acc@3	PAcc
	1 km					
LM	0.067	0.073	0.125	0.152	0.170	0.122
+TS	0.068	0.074	0.128	0.160	0.180	0.129
+CS	0.066	0.073	0.13	0.158	0.179	0.126
+CSR	**0.070**	**0.075**	**0.141**	**0.176**	**0.197**	**0.140**
	10 km					
LM	0.181	0.197	0.261	0.278	0.291	0.247
+TS	0.181	0.197	0.260	0.278	0.291	0.245
+CS	0.183	0.195	0.266	0.285	0.297	0.252
+CSR	**0.187**	**0.201**	**0.271**	**0.288**	**0.301**	**0.255**
	100 km					
LM	0.288	0.309	0.370	0.410	0.435	-
+TS	0.290	0.311	0.371	0.409	0.437	-
+CS	0.289	0.310	0.387	0.430	0.456	-
+CSR	**0.296**	**0.314**	**0.390**	**0.443**	**0.470**	-

Table 2: Performance of neighbourhood smoothing

boost in performance, but the information gleaned from such a knowledge base is already inherent in the language model itself. However, the combinations of two methods produce even better results and the maximum performance is reached by using all three methods together (except for **Acc** and **MRR** measures for 1 km grid division). To sum up, the proposed techniques equally improve the performance of **Acc** and **MRR** measures for all cell sizes and are especially effective for 1 km cells according to the rest of measures (up to 31% improvement). The results for accuracy-based measures were not tested for statistical significance because they have a binary outcome (correct or incorrect). The final improvements for MRR are significant at $p < 0.001$ for 1 km and 10 km cells, and at $p < 0.005$ for 100 km cells for the paired t-test.

To study the potential of the proposed improvements, we also measured highly relaxed accuracies with **K** up to 50. Classification into regions of larger bounds is principally easier and, as shown on Figure 4, the performance of the baseline method increases accordingly. Despite that fact, the benefit from the advanced methods stays the same which probably means that they avoid especially coarse errors (from improperly predicted continents to countries).

We conducted an error analysis of our best method to know the boundaries of its performance. There are two main sources of errors: those caused by sparsity and noisiness of the location models and those arising from ambiguity and incompleteness of the tag sets used for mapping. In the first case, the right location is either not represented in the data (from 70% test cases for 1 km to 7% for 100 km cells), or poorly represented with tags specific for this location only (e.g. containing no toponyms).

There are several types of images that are difficult to localize in the second case: (1) images with tags specific to too many locations (e.g. *beach coast rocks lovers*; (2) images with toponyms, but with no tags disambiguating them (e.g. *michigan cats dogs*); (3) images with a tag falsely indicating the reference to a location (e.g. *madrid toronto* taken in Madrid, but mapped to Toronto, or *paris hilton* picturing a poster in New York); (4) images containing a tag specific to a region larger than a chosen grid cell size (e.g. *alaska snow* for 100 km cells or *montmartre paris* for 1 km cells). We suppose that first three types of errors can be eliminated in the future by taking some contextual or user-specific evidence: for instance, tags of recently uploaded images or the

location of user IP. Highly ambiguous tagsets may be successfully mapped by relying on the history of user locations, since such tagsets might be location-specific on the personal level (for example, people celebrating their *birthday* in their home location).

Cases like the one shown in Figure 3 may be resolved with additional image content analysis. Errors resulting from images containing tags specific to a larger region are more difficult to avoid. This is due to the fact that some parts of large locations are richer in region-specific tags because they are more popular among users. However, the suggestion of the most popular part of it instead of the region as a whole or its centroid is a sensible strategy conceivably correlating with user satisfaction. It is interesting to notice that the dependence of mapping performance on the quality of test tag sets is reflected in distribution of accurately mapped images over countries. As we see on Figure 2 (right), photos taken in very popular tourist destinations, such as France, Italy or Morocco, are represented better among correct mappings than in the entire data set, seemingly because tourists almost always describe their photos with location specific tags.

Figure 3: Images with *palma* tag falsely mapped near Palma de Mallorca, Spain

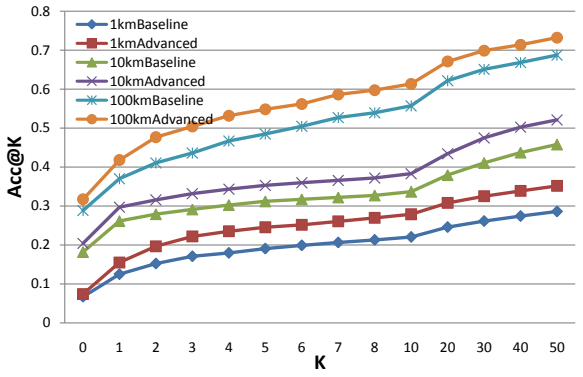

Figure 4: Accuracies at different distances K

7. CONCLUSIONS

In this paper we presented and evaluated generic methods for automatically placing photos uploaded in Flickr on the world map. We show that we can effectively estimate a language model through analysis of the terms people use to describe images taken at a particular location. This provides us with an extensible baseline, for which we have shown that we can further increase the accuracy of our predictions by incorporating ambiguity-aware smoothing, cell-based smoothing with score propagation in the direction of highly relevant neighbours, and using an external knowledge base.

There are several ways in which we would like to extend this work. First, it is necessary to automatically define an appropriate grid division for a tag set. It is important to minimise interactions between users and the system by showing a map view at the optimal zoom level: probably covering more than one cell, if there are several relevant cells near each other. Second, it seems promising to study the utility of additional evidence coming from a user profile, uploads history, social network or IP address. Finally, images used to build location models can be distinguished by using common (e.g. noise ratio) or Flickr-specific (number of views, interestingness) quality measures.

8. ACKNOWLEDGMENTS

We would like to thank Djoerd Hiemstra for fruitful discussions and his comments on the first versions of the paper. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreements (nr. 231507 and nr. 215453).

9. REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. Yang. World Explorer: Visualizing aggregate data from unstructured text in geo-referenced collections. In *JCDL '07*, 2007.

[2] M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07*, pages 971–980, New York, NY, USA, 2007. ACM.

[3] E. Amitay, N. Har'El, R. Sivan, and A. Soffer. Web-a-where: geotagging web content. In *SIGIR '04*, pages 273–280, New York, NY, USA, 2004. ACM.

[4] L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak. Spatial variation in search engine queries. In *WWW '08*, 2008.

[5] G. Cao, J.-Y. Nie, J. Gao, and S. Robertson. Selecting good expansion terms for pseudo-relevance feedback. In *SIGIR '08*, pages 243–250, New York, NY, USA, 2008. ACM.

[6] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *proceedings of the 18th International World Wide Web Conference (WWW 2009)*, Madrid, Spain, April 2009.

[7] J. Ding, L. Gravano, and N. Shivakumar. Computing geographical scopes of web resources. In *VLDB '00*, pages 545–556, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[8] J. Hays and A. A. Efros. im2gps: estimating geographic information from a single image. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[9] L. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *WWW '08*, 2008.

[10] J. Leidner. *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. PhD thesis, University of Edinburgh, 2007.

[11] A. McCallum, R. Rosenfeld, T. M. Mitchell, and A. Y. Ng. Improving text classification by shrinkage in a hierarchy of classes. In *ICML '98*, pages 359–367, San

Method	Acc	MRR	Acc@1	Acc@2	Acc@3	PAcc
	1 km					
LM	0.067	0.073	0.125	0.152	0.170	0.122
+CSR	0.070	0.075	0.141	0.176	0.197	0.140
+TB	**0.074** (+10%)	**0.078** (+7%)	0.141	0.175	0.198	0.142
+TB+CSR	0.074	0.078	0.147	0.188	0.212	0.148
+AS	0.061	0.068	0.132	0.167	0.186	0.124
+AS+TB	0.070	0.074	0.143	0.183	0.207	0.139
+AS+CSR	0.062	0.068	0.143	0.181	0.202	0.136
+AS+TB+CSR	0.069	0.073	**0.155** (+24%)	**0.197** (+30%)	**0.222** (+31%)	**0.149** (+22%)
	10 km					
LM	0.181	0.197	0.261	0.278	0.291	0.247
+CSR	0.187	0.201	0.271	0.288	0.301	0.255
+TB	0.198	0.209	0.283	0.303	0.316	0.269
+TB+CSR	0.198	0.210	0.286	0.305	0.319	0.269
+AS	0.190	0.205	0.275	0.292	0.306	0.260
+AS+TB	0.204	0.213	0.295	0.314	0.329	0.279
+AS+CSR	0.194	0.206	0.285	0.303	0.317	0.267
+AS+TB+CSR	**0.204** (+13%)	**0.213** (+8%)	**0.297** (+14%)	**0.316** (+14%)	**0.332**(+14%)	**0.280**(+13%)
	100 km					
LM	0.288	0.309	0.370	0.410	0.435	-
+CSR	0.296	0.314	0.390	0.443	0.470	-
+TB	0.306	0.322	0.398	0.446	0.475	-
+TB+CSR	0.310	0.324	0.409	0.465	0.494	-
+AS	0.302	0.321	0.386	0.427	0.452	-
+AS+TB	0.314	0.328	0.406	0.453	0.481	-
+AS+CSR	0.309	0.324	0.405	0.461	0.488	-
+AS+TB+CSR	**0.317** (+10%)	**0.329**(+6%)	**0.418** (+13%)	**0.477** (+16%)	**0.504** (+16%)	-

Table 3: Performance of combinations of methods

Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[12] Q. Mei, C. Liu, H. Su, and C. Zhai. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In *WWW '06*, 2006.

[13] M. Naaman, S. Harada, Q. Wang, H. Garcia-Molina, and A. Paepcke. Context data in geo-referenced digital photo collections. In *MULTIMEDIA '04*, pages 196–203, New York, NY, USA, 2004. ACM.

[14] M. Naaman, A. Paepcke, and H. Garcia-Molina. From where to what: metadata sharing for digital photographs with geographic coordinates. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 196–217, 2003.

[15] S. Overell, B. Sigurbjornsson, and R. van Zwol. Classifying tags using open content resources. In *WSDM '09*, 2009.

[16] J. Raper, G. Gartner, H. Karimi, and C. Rizos. Applications of location-based services: A selected review. *Journal of Location Based Services*, 1(2), 2007.

[17] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *SIGIR '07*, 2007.

[18] M. Richardson and P. Domingos. The intelligent surfer: Probabilistic combination of link and content information in pagerank. In *NIPS '01: Advances in Neural Information Processing Systems*, 2001.

[19] P. Serdyukov, H. Rode, and D. Hiemstra. Modeling multi-step relevance propagation for expert finding. In *CIKM '08*, pages 1133–1142, New York, NY, USA, 2008. ACM.

[20] A. Shakery and C. Zhai. A probabilistic relevance propagation model for hypertext retrieval. In *CIKM '06*, pages 550–558, New York, NY, USA, 2006. ACM Press.

[21] B. Sigurbjornsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *proceedings of the 17th International World Wide Web Conference (WWW 2008)*, Beijing, China, April 2008.

[22] C. Torniai, S. Battle, and S. Cayzer. Sharing, discovering and browsing geotagged pictures on the web. Technical report, Digital Media Systems Laboratory, HP Laboratories Bristol, 2007.

[23] K. Toyama, R. Logan, and A. Roseway. Geographic location tags on digital images. In *MULTIMEDIA '03*, pages 156–166, New York, NY, USA, 2003. ACM.

[24] S. Vadrevu, Y. Zhang, B. Tseng, G. Sun, and X. Li. Identifying regional sensitive queries in web search. In *Proceedings of WWW '08*, 2008.

[25] C. Wang, J. Wang, X. Xie, and W.-Y. Ma. Mining geographic knowledge using location aware topic model. In *GIR '07*, 2007.

[26] K. Weinberger, M. Slaney, and R. van Zwol. Resolving tag ambiguity. In *Proceedings of the 16th International ACM Conference on Multimedia (MM 2008)*, Vancouver, Canada, November 2008.

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

[28] Z. Zhuang, C. Brunk, and C. L. Giles. Modeling and visualizing geosensitive queries based on user clicks. In *LocWeb '08*, 2008.

[29] W. Zong, D. Wu, A. Sun, E.-P. Lim, and D. H.-L. Goh. On assigning place names to geography related web pages. In *JCDL '05*, pages 354–362, New York, NY, USA, 2005. ACM.