Towards Extracting Flickr Tag Semantics

Tye Rattenbury, Nathan Good, and Mor Naaman*
Yahoo! Research Berkeley
Berkeley, CA, USA
{tye, ngood, mor}@yahoo-inc.com

ABSTRACT

We address the problem of extracting semantics of tags – short, unstructured text-labels assigned to resources on the Web – based on each tag's metadata patterns. In particular, we describe an approach for extracting place and event semantics for tags that are assigned to photos on Flickr, a popular photo sharing website supporting time and location (latitude/longitude) metadata. The approach can be generalized to other domains where text terms can be extracted and associated with metadata patterns, such as geo-annotated web pages.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

tagging systems, event identification, place identification, word semantics

1. INTRODUCTION

User-supplied "tags", textual labels assigned to content, have been a powerful and useful feature in many social media and Web applications (e.g. Flickr, del.icio.us, Technorati). Tags usually manifest in the form of a freely-chosen, short list of keywords associated by a user to a resource such as a photo, web page, or blog entry. Unlike categories or ontology-based systems, tags result in unstructured knowledge – they have no a-priori semantics. However, it is precisely the unstructured nature of tags that enables their utility. For example, tags are probably easier to enter than picking categories from an ontology; tags allow for greater flexibility and variation; and tags can naturally evolve to reflect emergent properties of the data.

Despite their lack of ontology and a-priori semantics, tags exhibit patterns and trends [2] that allow *some* structured information to be extracted. The ability to assign structure to tags and tag-based data will make tagging systems more useful.

Broadly, we are interested in the problem of identifying patterns in the distribution of tags over some domain; in this work we focus on spatial and temporal patterns. Specifically, we are looking at tags on Flickr [1], a popular photo-sharing web site with support for user-contributed tags and geo-referenced (or, *geotagged*) photos. Based on the temporal and spatial distributions of each tag's usage, we attempt to automatically determine whether a tag corresponds to a place and/or an event. For example, our method should detect that the tag `Bay Bridge` describes a place, and that the tag `WWW2007` is an event. Tag usage distributions are derived from the distributions of photos. Figure 1 shows the spatial and temporal usage distribution for the tag `Hardly Strictly Bluegrass` in the San Francisco Bay Area.

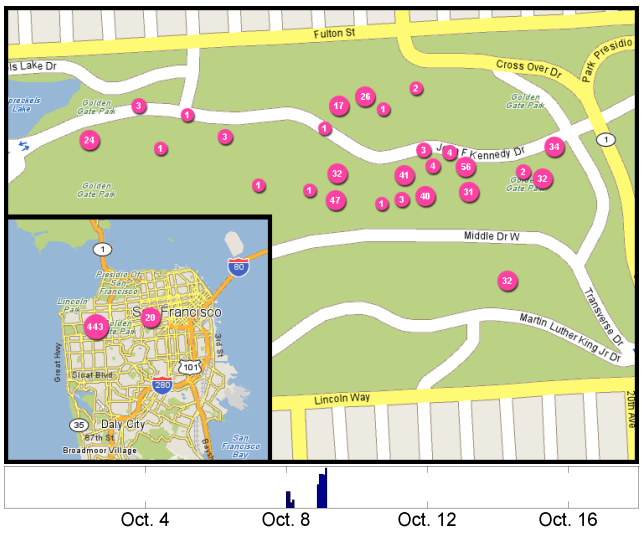

Figure 1: Spatial (top) and temporal (bottom) distributions for the tag `Hardly Strictly Bluegrass` in the San Francisco Bay Area.

Extraction of event and place semantics can assist many different applications in the photo retrieval domain and beyond. Benefits include:

- improved image search through inferred query semantics;
- automated creation of place and event gazetteer data (used to improve web search, for example); and
- automated association of missing location/time metadata to photos, or other resources, based on tags or caption text.

In this work we do not apply our analysis to a specific application, but rather investigate the feasibility of automatically determining place and event semantics from Flickr tags.

*Nathan and Tye are also affiliated with UC Berkeley.

WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
ACM 978-1-59593-654-7/07/0005.

2. GENERAL APPROACH

Our approach relies on the following three assumptions. First, that we have a set of tags whose semantics we are trying to determine. Second, that associated with each tag is a usage distribution over some dimension – e.g. the times when the tag was used. Third, we assume that the semantics we are trying to extract can be defined in relation to the dimension over which the tag's usage is distributed.

We will describe our approach to the extraction of semantics via the notions of events and places. We define *event tags* as tags whose usage distribution is expected to demonstrate significant temporal patterns. Similarly, *place tags* are tags whose usage distribution is expected to demonstrate significant spatial patterns.

One approach to identifying tags that correspond to events and places is to detect bursts of usage in space or time – i.e., if the tag demonstrates a strong spatial or temporal burst of usage, then it is likely a place or an event, respectively. We tested two standard burst detection methods. The first, *Naïve Scan*, was used to detect important query terms in web query logs [4]. The second, *Spatial Scan*, is used by epidemiologists to detect disease outbreaks [3]. The primary issue with these approaches is that while bursts are important, there is no check performed to ensure that only one burst has occurred. Specifically, these methods do not perform well when the data is sparse and contains multiple bursts (see, for example, the spatial and temporal distributions for the tag `Hardly Strictly Bluegrass` in Figure 1).

To handle the issue of multiple bursts, we developed a novel method, *Scale-structure Identification* (or SSI). This method measures how similar the data is to a single cluster at multiple scales.[1] For example, the tag `Hardly Strictly Bluegrass` appears as a single strong cluster at the city scale; but appears as multiple clusters at a neighborhood scale (see Figure 1). SSI works by: (1) clustering the usage distribution for a tag at multiple scales; (2) measuring the dispersion of usage occurrences among the clusters by calculating the information entropy; and (3) summing the entropy calculations at each scale to produce a single score.

3. EXPERIMENTS

To test each method's ability to identify place and event tags, we chose to focus on 49,896 Flickr photographs taken in the San Francisco Bay Area. From these photos we found 803 tags that were used at least 25 times and by at least 2 people. We compare the results of our automatic approaches to a hand-labeled, ground truth – generated by a human judge who examined a subset of each tag's associated photos and captions. Photo and caption content enabled the human judge to generalize, correct, and interpolate inaccurate and sparse data.

With a ground truth data set, we can measure the effectiveness of the automatic approaches by calculating precision and recall. We define precision and recall for event identification (the definitions for place identification are analogous). Each of our methods classifies the list of tags as an event tag or not. Given this classification, precision is the percentage of tags correctly labeled as event tags – i.e. precision is the number of tags correctly classified as event tags divided by the total number of tags classified as event tags.

[1]SSI handles periodic events by treating time as cyclical instead of linear.

	P-R area	Max F1	Min CE	
Naïve Scan	0.4455	0.5279	0.2914	PLACE
Spatial Scan	0.6028	0.5907	0.2441	
Scale-structure	0.7034	0.6655	0.1930	
Naïve Scan	0.3291	0.3636	0.1009	EVENT
Spatial Scan	0.4130	0.4811	0.1034	
Scale-structure	0.6420	0.6533	0.0648	

Table 1: Precision-Recall Area, Maximum F1, and Minimum CE values for the various methods.

Recall is the percentage of all event tags (from the ground truth data) that are correctly classified as event tags.

By varying the classification threshold associated with each method, we can cover all possible recall values. From the precision and recall measurements for each method, we can compute a number of standard scores: (1) the area under the precision-recall curve (P-R area), (2) the maximum value of the F1 statistic (Max F1), a metric that balances precision and recall, and (3) the minimum total classification error (Min CE). Results for the three methods, Naïve Scan, Spatial Scan, and SSI, are shown in Table 1. SSI clearly outperforms the standard burst detection methods on these metrics.

Errors produced by SSI have simple explanations. First, the majority of false positives and false negatives for place identification were the result of sparse data. For example, tags like `drunk` and `sail` were incorrectly classified as places while tags like `UCSF` and `Mission District` were incorrectly classified as not being places. Likewise the false positives for event identification were often due to sparse data. False negative event tags were also caused by bad data – noisy as opposed to sparse. For example, tags like `thanksgiving` and `October` were incorrectly classified as not being events.

4. FUTURE WORK

The experiments presented in this paper correspond to data from the San Francisco Bay Area. We would like to extend our methods to the entire world, which will require some specification of "regions of interest". For example, the tag `carnival` may be event-like around Rio de Janeiro, but elsewhere in the world it is less likely to exhibit event-like usage patterns. We plan to explore how to generate, store, and disambiguate tag semantics for different regions throughout the world. Additionally, we will look at extending the metadata features used, beyond location and time, to extract semantics other than place and event.

5. REFERENCES

[1] Flickr.com. http://www.flickr.com.

[2] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *Proc. Multimedia*, p. 89–98. ACM Press, 2006.

[3] M. Kulldorff. Spatial scan statistics: models, calculations, and applications. In *Scan Statistics and Applications*, p. 303–322, 1999.

[4] M. Vlachos, C. Meek, Z. Vagena, and D. Gunopulos. Identifying similarities, periodicities and bursts for online search queries. In *Proc. SIGMOD*, p. 131–142. ACM Press, 2004.