World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections

Shane Ahern, Mor Naaman, Rahul Nair, Jeannie Yang
Yahoo! Research Berkeley
Berkeley, CA, USA
{sahern, mor, rnair, jeanniey}@yahoo-inc.com

ABSTRACT

The availability of map interfaces and location-aware devices makes a growing amount of unstructured, geo-referenced information available on the Web. This type of information can be valuable not only for browsing, finding and making sense of individual items, but also in aggregate form to help understand data trends and features. In particular, over twenty million geo-referenced photos are now available on Flickr, a photo-sharing website – the first major collection of its kind. These photos are often associated with user-entered unstructured text labels (i.e., tags). We show how we analyze the tags associated with the geo-referenced Flickr images to generate aggregate knowledge in the form of "representative tags" for arbitrary areas in the world. We use these tags to create a visualization tool, World Explorer, that can help expose the content of the data, using a map interface to display the derived tags and the original photo items. We perform a qualitative evaluation of World Explorer that outlines the visualization's benefits in browsing this type of content. We provide insights regarding the aggregate versus individual-item requirements in browsing digital geo-referenced material.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: General; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

algorithms, human factors

Keywords

geo-referenced data, photos, tagging systems, visualization

1. INTRODUCTION

The amount of geographically annotated material available on the Web is constantly growing. Such geo-referenced

materials include geographically-annotated web pages, blogs, and other types of data served via the GeoRSS format[1], as well as geo-referenced digital photographs. The information associated with these geo-referenced resources is relatively unstructured, and usually includes some type of text, be it labels, web page text, or tags. Traditionally, the digital libraries and information retrieval communities have looked at methods of efficiently indexing, retrieving, ranking and browsing documents in geographic data collections [10, 21]. However, the existence of location metadata also allows the generation of interesting location-driven *aggregate* knowledge from these unstructured text-based collections: when enough information is available, systems can identify useful location-driven trends and patterns in the text data.

More than any other resource, geo-referenced photographs are commonplace on the Web today. Our focus in this work is Flickr [7], a popular photo-sharing website. Flickr employs various means for its users to geo-reference (or, "geo-tag") their photographs – most prominently, Flickr supplies a map interface through which users can "drag" their photos to the map locations where the photos were taken. In addition, many photos are geotagged using GPS logs or location-aware devices (for a review of techniques for geo-referencing photos, see [22]). We expect the rate of addition of geo-tagged photos to keep increasing as location-aware devices become more available. Other than location, Flickr allows users to associate photos with multiple tags: unstructured textual labels that are freely entered, in most cases by the user who has taken the image [14]. These tags, together with the location information assigned to the photo, are the basis for the analysis in this work.

Our sample application considers all geo-referenced photos on Flickr and generates an aggregate representation that allows navigation, exploration and understanding of the underlying data (and the world). The idea is similar to a 1976 social psychology work by Stanley Milgram, who had asked his subjects to list places of interest in Paris [15]. Milgram then aggregated the results, effectively creating a "psychological map" of Paris with landmark names appearing in a larger font according to the number of subjects who mentioned each landmark. In effect, we similarly create "Psychological Maps" of arbitrary areas in world, based on implicit user input via their Flickr photos. Our analysis demonstrates key ideas in location-based information analysis, and serves as input to a visualization of the aggregate data in an application we call *World Explorer*. The World Explorer application (Figure 1) enables users to easily explore and effi-

---

[1]http://www.georss.org/

Figure 1: A screenshot from the World Explorer visualization, showing parts of San Francsico; the user highlighted the tag Golden Gate Bridge (partially off-screen) to bring up photos annotated with that tag, from that specific area.

ciently browse through this large-scale geo-referenced photo collection, in a manner that improves rather than degrades with the addition of more photos. While our application is tied to Flickr and photographic data, a similar approach can be used to analyze and visualize other types of geo-referenced resources in large-scale digital libraries.

The data analysis in our system is based on multi-level clustering and TF-IDF-based (term frequency, inverse document frequency) scoring of tags. The visualization exposes, for each map region and zoom level, the high-scoring tags for the generated clusters; these tags are shown as text over the map area where each cluster occurs. In addition, the tags are connected to the corresponding photographs such that when a user points the mouse over a tag, photographs associated with that tag, from that map area, appear in a dedicated pane. Thus, a user can get an idea of what is available in the visible map region via the tags, and explore the available individual items via photographs.

There are, of course, challenges in analyzing and visualizing such unstructured user-contributed data. Issues of noise (e.g. photos with tags that are not relevant to the location) and errors (e.g., photos that are geotagged incorrectly) abound in the Flickr data; the algorithm should be able to handle these in a graceful manner. There are also considerations of scale, especially as the amount of data increases, and the requirements of an on-line, interactive visualization require real-time response for arbitrary map bounding boxes.

To summarize, the contributions of this work are:

- An approach for deriving meaningful data from unstructured text associated with geo-referenced collections.
- A sample application that derives such information from Flickr geotagged images.
- A visualization technique for large-scale geo-referenced photo collections, that allows automatically-derived and effective world exploration via photos and maps.

We begin by discussing the related work; a formal definition of our data model and analysis follows in sections 3 and 4. We describe the visualization in Section 5 and our evaluation in Section 6.

2. RELATED WORK

Two recent efforts have used visualization techniques and computation similar to the ones we employ, albeit in different settings. In [23], the authors used TF-IDF to compute and extract meaningful words to represent email correspondence in a personal email archive. More specifically, the "Themail" application described in that work extracted words to represent each month of the collection owner's correspondence with specific contacts. Such a scheme allows the user to get an overview of their correspondence with a contact, and brings out major themes that were discussed in different times. In the work of Dubinko et al. [5], a TF-IDF approach was used to visualize tags in Flickr over time, or rather, extract the most representative tags for arbitrary time periods. The authors' approach did not consider the location metadata and map-based visualization requirements; potentially, a combination of location and time metadata could generate a location-specific, dynamic display of Flickr data.

Since 2001, a number of different efforts have examined applications geographic location information associated with photos [3]. In [22], the authors describe WWMX, a map-based system for browsing a global collection of geo-referenced photos. Several similar map-based photo browsing systems appeared on the Web in the last few years;[2] Flickr now also employs a map-based interface to access photos.[3] All of those systems face the problem of clutter in the map interface: as the number of photos available in each location

---

[2]current systems include whoophy.com and mappr.com
[3]http://flickr.com/map

grow, the full set of images cannot possibly be shown on the map at once. While some systems default to showing the most recent photos, both Flickr and WWMX try to handle clutter by consolidating multiple photograph markers into a single marker according to the distance between markers. The system described in this work can potentially alleviate the clutter issue caused by the need to expose individual photos' location.

Similar to our work, work in both [4, 16] considers patterns and distributions of textual terms that are associated with geo-referenced digital photos, and uses them to generate tag suggestions for new photographs. These systems consider ad-hoc queries and do not attempt to visualize or analyze the underlying data.

Of course, the GIS and Digital Libraries communities have been active in investigating geo-referenced data access and visualization techniques (see [9, 10, 21, 24] and more recently [11, 12, 13]). Our focus in this work is somewhat different, showing that textual unstructured data can be analyzed for meaningful patterns in a geographic information system.

More semantically-rich location analysis problems have been studied in the domain of web-based information retrieval or specifically in the field of "GeoIR"; for example, two groups of researchers [1, 17] looked into extracting the scope of geographic terms or entities based on co-occurring text and derived latitude-longitude information, a goal somewhat similar to our visualization goals. In fact, our analysis can use these methods to assign location information to items where the lat/long data is not available. Similarly, we have shown elsewhere [18] that place and event semantics of Flickr tags can be automatically derived from the time- and location-based patterns of photos associated with that tag.

3. DATA MODEL

We begin with a formal definition of our dataset and research problem. Notice that our data model and heuristics refer to elements from the domain of geotagged photographs and their associated tags. However, both the data model and the heuristics can be easily applied or modified to other domains mentioned above, such as geo-referenced blog posts and web pages.

Our dataset consists of three basic elements: photos, tags and users. We define the set of photos as $\mathbb{P} \triangleq \{p\}$, where p is a tuple $(\theta_p, \ell_p, \alpha_p, u_p)$ containing a unique photo ID, θ_p; the photo's capture location, represented by latitude and longitude, ℓ_p; the location accuracy, α_p, and the ID of the user that contributed the photo, u_p. The location ℓ_p generally refers to the location where the photo p was taken, but sometimes marks the location of the photographed object. The parameter α_p indicates the granularity in which the location data was captured and can range from exact coordinates to "country-level", allowing for fuzzy or uncertain locations for photos.

The second element in our dataset is the set of tags associated with each photo. We use the variable x to denote a tag. Each photo p can have multiple tags associated with it; we use $\mathbb{X}_p$ to denote this set of tags. The set of all tags over all photos is defined as: $\mathbb{X} \triangleq \cup_{p \in \mathbb{P}} \mathbb{X}_p$. We can use the equivalent notation to denote the set of tags that appear in any subset $\mathbb{P}_S \subseteq \mathbb{P}$ of the photo set as $\mathbb{X}_S$. For convenience, we define the subset of photos associated with a specific tag

as: $\mathbb{P}_x \triangleq \{p \in \mathbb{P} \mid x \in \mathbb{X}_p\}$. Accordingly, photos with the tag x in a subset $\mathbb{P}_S$ of $\mathbb{P}$ are denoted $\mathbb{P}_{S,x} \triangleq \{\mathbb{P}_S \cap \mathbb{P}_x\}$.

The third element is users, the set of which we denote by the letter $\mathbb{U} \triangleq \{u_p\}$. Equivalently, we use $\mathbb{U}_S \triangleq \{u_p \mid p \in \mathbb{P}_S\}$, $\mathbb{U}_x \triangleq \{u_p \mid p \in \mathbb{P}_x\}$ and $\mathbb{U}_{S,x} \triangleq \mathbb{U}_S \cap \mathbb{U}_x$ to denote users that exist in the set of photos $\mathbb{P}_S$, users that have used the tag x, and users that have used the tag x in the set of photos $\mathbb{P}_S$, respectively.

Using such data we wish to find tags that are most "representative" for each given geographical area. It is important to note that these representative tags are often not the most commonly used tags within the area under consideration. Instead, we wish to surface tags that uniquely define sub-areas within the area in question. For example, if the user is examining a portion of the city of San Francisco then there is very little to be gained by showing the user the `San Francisco`[4] or `Bay Area` tags, even if these tags are the most frequent, since the tags apply to the entire area under consideration. Instead, we would ideally show tags such as `Golden Gate Bridge`, `Alcatraz` and `Pier 39` which uniquely represent specific locations, landmarks and attractions within the city.

The first step in determining the "representativeness" of a tag is to have an intuition of what the term implies. While there are no formal models to define how representative a tag is for an area, we follow some simple heuristics that guide us in devising the algorithms. The heuristics attempt to capture the human attention and behavior as represented in the photos and tag dataset, and include the notions that:

- The number of photographs taken in a location are an indication of the relative importance of that location.
- The importance of a location increases with the number of individual photographers that have taken photos there.
- Users are likely to use a common set of tags to identify the objects/events/locations that occur in photographs of a specific location.
- Tags that occur in a concentrated area (and do not occur often outside that area) are more representative than tags that occur diffusely over a large region.
- The more users that used a tag in an area, the more representative the tag is for that area.

Based on the data, the requirements from our analysis are therefore:

- Identify important regions for every map region and zoom level.
- Select representative tags for the identified regions.

Finally, the World Explorer system is designed to provide a visualization of tags that represent any given area. Since the users would be able to browse any area of the world, the system must be able show the selected tags for any arbitrary geographic area and at any level of detail. The backend must be able to support this visualization with near real time responses to provide a seamless user experience.

4. DATA ANALYSIS

This section explains the algorithms and the framework underlying our work, including the pre-processing steps, the

---

[4]We use this format to represent tags in the text

data storage, and run-time retrieval. We begin by limiting our problem definition to generating the most representative tags for a pre-defined area. We then extend the description to show how we pre-compute tags for the entire world, and serve "representative tag" results for arbitrary geographical bounding boxes. Finally, we provide the details of our storage and retrieval scheme for the pre-computed tag data.

4.1 Computing Tags for a Geographic Area

As mentioned above, we start by assuming that the system considers a single given geographic area G, and the photos that were taken in this area, $\mathbb{P}_G$. The system attempts to extract the representative tags for area G. This computation is done in two main steps: in the first step, we cluster the set of photos $\mathbb{P}_G$ using the photos' geographic locations. In the second step, we score the tags in each cluster.

In the first step, the system geographically clusters the set of photographs $\mathbb{P}_G$. We use the k-Means clustering algorithm, based on the photos' latitude and longitude. Geographical distance is used as the distance metric, and the stopping condition for the k-Means algorithm is when each cluster's centroid movement drops below 50 meters. The number of seed points used for the k-Means algorithm was based on $|\mathbb{P}_G|$, the number of photographs in the area under question. Based on empirical observation of the results, we set the seed value to range from three for sparse areas (under 100 photographs) to 15 for denser areas (greater than 4000 photographs). If the clustering results with centroids that are close to each other (less than 20% of the width of G) we merge the clusters – this merging is done due to the a-priori nature of seed selection for the k-Means algorithm. Future implementations could easily use different clustering algorithms which do not require an a-priori decision about the number of clusters in the area, or deploy other criteria, such as the Bayesian Information Criterion (BIC) [19] to search for the value of k (in earlier work [8], we have used the Hungarian Clustering method for a similar application).

Once the clusters have been determined, the system scores the cluster's tags to see if it is possible to extract representative tags for each cluster. In other words, we consider each cluster C, and the set of tags $\mathbb{X}_C$ that appear with photos from the cluster. We rank each tag x in the set so that the top tags are, at least according to our heuristics, the most representative tags.

One of the factors we use for scoring is a TF-IDF approach that assigns a higher score to tags that have a larger frequency within a cluster compared to the rest of the area under consideration. Again, the assumption is that the more unique a tag is for a specific cluster, the more representative the tag is for that cluster. Of course, we do not wish to use tags that only appear a few times in the cluster; therefore, the term frequency element prefers popular tags.

The TF-IDF is computed with slight deviation from its regular use in Information Retrieval. The term frequency for a given tag x within a cluster C is the count of the number of times x was used within the cluster, or in our terms, $\mathrm{tf}(C, x) \triangleq |\mathbb{P}_{C,x}|$. The inverse document frequency for a tag x usually reflects the ratio of documents (i.e., clusters) that contain photos with this tag in the entire area G under consideration. In our case, we modify the measure to consider the overall ratio of the tag x amongst all photos in the region G under consideration: $\mathrm{idf}(x) \triangleq |\mathbb{P}_G|/|\mathbb{P}_G, x|$. This deviation from the standard is done due to the small num-

ber of clusters that we derive for each area. If we were to use the presence of a tag within a cluster then we could face large changes in the TF-IDF weights if even a single photograph in the cluster contains a tag. By using this approach we are able to 'smooth' the results. Also note that we only consider a limited set of photos ($\mathbb{P}_G$) for the IDF computation, instead of using the statistics of the entire dataset (as in the visualization efforts from [5, 23]). This restriction to the current area, G, allows us to identify local trends for individual tags, regardless of their global patterns.

Finally, to get the TF-IDF score, we multiply the two terms, $\mathrm{tfidf}(x, C) = \mathrm{tf}(x, C) \cdot \mathrm{idf}(x)$. The higher the tf-idf score the more distinctive the tag is within a cluster.

While the tag weight is a valuable measure of the popularity of the tag, it can often be affected by a single photographer who takes a large number of photographs using the same tag. To guard against this scenario, we include a user element in our scoring that also reflects the heuristic mentioned above that a tag is more valuable if a number of different photographers use it (as a further guard, we assign a score of 0 to any tags that was used by less than 3 photographers in a given cluster). In particular, we factor in the percentage of photographers in the cluster C that use the tag x: $\mathrm{uf} \triangleq \mathbb{U}_{C,x}/\mathbb{U}_C$.

The final score for tag x in cluster C is computed by $\mathrm{Score}(C, x) = \mathrm{tf} \cdot \mathrm{idf} \cdot \mathrm{uf}$. For each cluster, we retain only the tags that score above a certain threshold. The threshold is needed to ensure that the selected tags are meaningful and valuable for the aggregate representation. We use an absolute threshold for all computed clusters to ensure that tags that are picked are representative of the cluster.

4.2 Indexing and Storage

In this section we discuss the indexing and pre-computation methods that allow us to store the representative tags for the entire world. These methods are mostly based on standard geographic data-structures. The pre-computation and indexing are required to support the high volume of requests at near real-time rates required by the visualization.

Our algorithm first divides the world into equal sized non-overlapping "tiles" such that neighboring tiles are roughly equal in size. Since the photographs are referenced using the geographic coordinates system, we use the coordinates system to divide the space into tiles with equal horizontal and vertical length (in terms of coordinates). Though the tiles created in this way are not strictly equal in physical area, they are close enough in size to their neighboring tiles to make no practical difference. We then pre-compute the tags for each of the tiles in the manner described above (Section 4.1). As mentioned above, we retain all the tags that score above a threshold value (empirically determined to be 1.0). The stored information includes the tag itself along with associated metadata such as the computed score, id of the originating tile and the coordinates of the originating cluster.

Since one of our requirements was to be able to serve data for any area of the world, at any zoom level, we performed the pre-computation using several different tile sizes. This way we could provide a correct score for tags for different area sizes. For example, we expect the tag `Alcatraz` to score high when considering the San Francisco area, but the tag `San Francisco` should only score high when, say, the area of consideration is the state California. Clusters and

scores computed for the entire United States are, of course, different than the collection of clusters computed for every individual part of the country (notice that we use the areas known as "San Francisco", "California" and "United States" as illustration here; again, the tiles were created arbitrarily without consideration to physical boundaries).

We use a quad tree [6] index for our geographic tiles, where the tile sizes are determined by the level in the tree. The root node of the tree covers the entire globe and is split into two children that represent the eastern and western hemispheres of the earth. From the hemisphere level downwards, each node is split into 4 equal, square, non-overlapping children that each cover one quarter of the area (again, using the coordinate system and not exact physical area) of the parent. The subdivision of nodes continues until either the tree reaches a maximum depth, or there are no more photographs available in the area covered by the current node. Once the division is complete, each node is considered to be a single tile, and the tags for each tile are computed and stored as described in Section 4.1.

4.3 Retrieval

As mentioned earlier, our algorithm indexes the world according to geographical coordinates and thus the indices may not match up with areas of user interest. Though the users can only request rectangular areas (defined using the coordinates of the bottom left and top right corners), it will be quite rare for any of the user requests to perfectly align with any of the existing nodes. Instead we must fit the requested area into existing nodes such that we are able to use the pre-computed tags from the nodes to calculate the appropriate tags. This section describes the retrieval algorithm that allows us to use our pre-computed tag index to return representative tags for any arbitrary geographical area.

Our first step is to find equal sized nodes that cover the requested area where size is calculated based on the geographical coordinate system. The equal size requirement is due to the TF-IDF algorithm which determines the "representative" tags based on size the overall area being queried. However our quad tree index has a limited levels of nodes while the requests can be of virtually any size. Instead of increasing the number of levels in the index, we "fit" the requested area into an existing node level. Since both the request and the nodes are described in geographical coordinates, it is a simple comparison to see which node level (either larger or smaller) is closest in size to the requested area. Once the closest node level is determined, we search for nodes at that level which completely cover the area requested by the user. The search will result in between 1 and 4 nodes that completely cover the requested geographical area. We then choose the top scoring tags from those nodes that fit within the requested area and return them along with their score and coordinates. If the same tag occurs multiple times we use the highest scoring instance of the tag. If the highest scoring tag of a cluster is being used elsewhere we use the next highest scoring tag and so on.

5. VISUALIZATION

The World Explorer visualization was designed as an interactive prototype that would encourage people to freely explore any place in the world.[5] The specific goals were:

Figure 2: Primary tags for central California

- Engage users in serendipitous discovery of interesting locations and photographs.
- Allow users to examine and understand a particular area of interest.
- Assist in exploration of areas of interest previously unknown to the users.

In this section we describe the main elements of the prototype and the interaction we developed to meet these goals.

The main navigational component of the visualization is a map. The user can pan the map by dragging the map area using the mouse; zoom in and out of the viewed area using a slider; or directly enter a location in the search box that appears next to the map.

The principal element of the visualization is the "primary tags". The primary tags, as shown in Figure 1 and Figure 2, are overlaid by the application on top of the viewed map area. The primary tags are meant to give the user an idea regarding the landmarks, points of interest and other items available in the viewed area. Whenever the user changes the focus of the map, the application retrieves the relevant primary tags for the current area and the current zoom level from the server. In reality, the application always retrieves tags for an area slightly larger than the currently viewed region, so that the user can perform some panning without waiting for server response. These primary tags are the top-scoring tags that represent the viewed area according to our analysis. In other words, for each area and zoom level, as described in Section 4, the system retrieves the appropriate clusters and returns the top-scoring tag for each cluster, the tag's score, and the tag's position (the cluster's centroid). The font size in which the tag is displayed is relative to the tag's computed score. If the same tag is the top-scoring tag in two different clusters, we select to display the second highest scoring tag for the second cluster. It is interesting to contrast the tags selected and displayed in Figure 1 with the tags in the higher zoom level of Figure 2: in different zoom level, different concepts are selected for representation.

The principal interaction method in the application is clicking, or simply hovering the mouse over one of the tags (both actions produce the same result; for convenience, we shall only refer to "clicking" below). For example, clicking
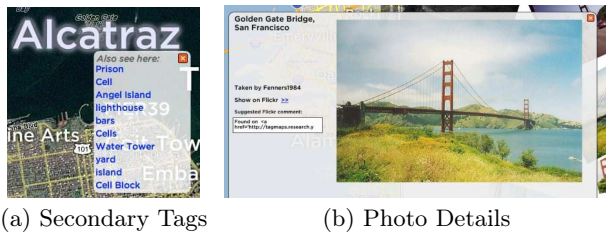
(a) Secondary Tags   (b) Photo Details

Figure 3: Secondary tags and expanded photo view in World Explorer.

a primary tag will show "secondary tags" as well as relevant photos.

The "secondary tags" can be thought of as an extension of the primary tags that are readily exposed on the screen for every map region. The secondary tags are displayed as a drop-down menu when the user clicks one of the primary tags, thus avoiding clutter while still allowing for more details to be exposed. As shown in Figure 3(a), these tags are usually other prominent tags that are relevant or representative of the map area around each primary tag. In Figure 3(a), the secondary tags are shown for the primary tag `Alcatraz`, representing the famous San Francisco Alcatraz prison, suggesting that these tags serve as an "also see here" feature. Note that the secondary tags in Figure 3(a) include both sub-features (`cell`, `lighthouse`), other descriptions for Alcatraz (`prison`), as well as objects that can be photographed from that location (`Angel Island`). The secondary tags are simply derived from the computation described above: for each area and zoom level, we retrieve the top non-primary tags from all nearby clusters (according to the centroid of the primary tag, and a radius based on the zoom level). In most cases, these tags are simply the other top-scoring tags from the primary tag's cluster. The derivation of the secondary tags is independent of the primary tag.

Both the primary and secondary tags, then, supply the context for the user interacting with the system. The displayed tags provide an understanding of what is available, or the salient features in the dataset for each map region. The tags support both discovery and navigation. Indeed, in an early version of the visualization, tags were the only element shown [8]. However, the tags themselves are only a representation of the underlying data – photos, in our case. We therefore added the capability to show photos associated with each tag, via a dedicated photo pane.

The photo pane is a critical part of the visualization, in that it gives the user actual visual information for each area, based on the currently selected tag. Whenever a user clicks a primary or secondary tag, the application loads 20 photos that were annotated with that tag from the same geographic region; i.e., photos that visually explain and extend the tag information. In Figure 1, as mentioned above, the user clicked on the `Golden Gate Bridge` tag to see related photos. Notice that most photos shown in the photo pane are photos of the bridge, and a sharp eye could tell that most were taken by users standing on the bridge itself (around the area where the tag is shown). Also note that highlighting the same tag, if it appears in a different location in the appli-

cation, would retrieve images of the bridge from a different vantage point. Clicking a secondary tag issues a transformational query, getting photos that match the secondary tag and the location, but do not necessarily match the original primary tag.

Once the photos are displayed, any photo can be expanded to be examined in more detail by double-clicking on it, as shown in Figure 3(b). A higher-resolution version of the image is then shown, together with additional metadata about the photo. Most importantly, the data includes the title of the photo and the name of the user who took that photo. From this expanded photo display the user can click on a link to see the photo in full resolution and detail on Flickr.

6. EVALUATION

We performed a qualitative evaluation of our system, as we reported on a direct evaluation on parts of the World Explorer in previous work, and a completion-time-based evaluation of the application was not appropriate. The direct evaluation performed in [8] explored the system's performance with data limited to the San Francisco area, and without the visualization elements. As part of that direct evaluation, we verified that the algorithm selects meaningful locations, and assigns relevant tags to these locations. We do not repeat this evaluation here. Similarly, we felt that a completion-time-based evaluation is also not appropriate here; the experience of the World Explorer application is not geared toward user task completion or task speed. Furthermore, meaningful metrics that could allow us to compare World Explorer to other systems were not available.

Our evaluation, then, focused on collecting qualitative feedback regarding the visualization and the effectiveness of the extracted tag data. In particular, we wanted to examine whether recruited participants found the application useful and for which purpose. We wished to extract additional requirements and deficiencies of the World Explorer application and see if other issues arise. In addition, we were hoping to find out which application features were discovered and used and which were missed. We attempted to gather data about whether participants understood the tag and location metaphor, and were able to navigate and comprehend what each element of the visualization is used for. Finally, we had users interact with the Flickr integrated map system to see how users react to that system in similar conditions (i.e., similar tasks).

We first describe the dataset used to create the World Explorer visualization. We then describe the experimental setup and discuss the results.

6.1 Dataset

The dataset used in this research consisted of 6 million public geo-referenced ("geotagged") photographs on Flickr (data collected in October 2006). There are a number of possible sources for photographs' location information on Flickr, user entry being probably the most prevalent: much like LocationStamper [22], Flickr users can geotag their photographs using a map interface, where they can drag-and-drop photographs onto a world map at any zoom level. The zoom level on the map when the photograph was dragged is recorded, and the accuracy level of the photo's location information is assigned accordingly. Flickr has 16 levels of accuracy ranging from the most general (level 1) to the most accurate (street level, 16).

Map-based entry of location is not the only possible source of photo location. Various external automatic and manual tools (such as GPS synchronization, other map applications and more) can be used to update or input locations for photographs on Flickr. Location-aware applications, such as the ZoneTag application for cameraphones,[6] automatically geotag photos they upload to Flickr. We expect the rate of addition of geotagged photographs to Flickr to increase as more location-aware devices become available.

Of course, issues of errors and noise are possible in the Flickr data. Photos could be geotagged incorrectly, or inappropriately, or assigned a misleading accuracy level. Photos are also likely to be assigned text tags that are irrelevant to the location. Naturally, there is no guarantee that photos taken in San Francisco, for example, will all have the `San Francisco` tag, or any other tag relevant to the location. While the correctness of the time and location metadata for each individual photo is suspect [2], and the tags associated with the photos may vary, we show that trends and patterns can be reliably extracted and used.

Ultimately, as described in Section 3, each photograph p in our dataset was associated with a location ℓ_p, geo-accuracy α_p, and the identity of the owner u_p. Almost 90% of the 6 million photos were associated with user-entered tags. While the heaviest concentration of geotagged photos was found in the United States and Western Europe, the dataset had very wide coverage, and included photos from almost every country in the world.

We excluded from our analysis photographs which had an accuracy lower than 10 (approximately city level). This exclusion had a marginal effect on the results: in the high-level tiles, enough accurate photos were available and the "fuzzy" ones were not needed; for the lower level tiles, it was best not to use photos with non-accurate location. We were left with 4.5 million photographs and 28 million tags.

The index was designed to follow the Flickr data and map interface characteristics. We used the index, as described in Section 4 with a maximum of depth of 16 levels (in order to mimic both the Flickr accuracy and map zoom levels in the visualization interface). At each node level, the tag computation was performed using all photographs that had an accuracy equal to or greater than the accuracy of the node level itself (but no less than accuracy level 10, as mentioned above). Additionally, in order to reduce noise, we excluded any nodes that had less than 30 photographs or less than three active users. In the case of nodes higher than level 7, the number of photographs in each node was often very large. In these cases, we ran the clustering and scoring algorithm on the tags computed from the lower level rather than the full dataset of photos. Finally, we automatically assigned a score of 0 to any tags that had been used by less than 3 users within a each cluster.

The entire pre-processing step was implemented using PHP scripts, and was not optimized for performance or parallelism. The execution on our dataset lasted a few days on a non-dedicated server. However, the run-time can easily be reduced with parallel execution on multiple machines, and using optimized code. In addition, it is easy to modify the code and our algorithm to lend itself to incremental changes such that additional data can be incorporated without going through the complete process.

_____

Conveniently, the number of node levels in our data structure is equivalent to the number of zoom levels of the map. This allows the system to directly translate the visualization state (or area and zoom level) to the level of data requested. In other words, if the user currently views the map visualization at level 5, the data served will be drawn from the 5th level of our data structure, regardless of the viewed area. The translation scheme, based on the zoom level and not just the viewed area, guarantees that for any screen size, from small screen to large devices, the number of objects (tags) delivered to the visualization will not be dependent on the view area, but on the appropriate level of detail.

6.2 Experimental Setup

We recruited a total of 10 participants to interact with our system. The recruitment was done via ads published on public forums and in an internal company mailing list. While recruiting participants, we excluded computer professionals from participating. The participants ranged in age between early 20s and over 55; with most of them in the 25–40 range. The participants background and technical experience varied from non-technical, yet computer-savvy employees of a high-tech company, to people with little experience with online photo sharing, let alone map-based photo browsing. Several participants required an explanation of the geotagged photo concept. All the participants were compensated for the study, which required about one hour of their time, by up to 20USD.

With each subject, we conducted a session that involved performing three different tasks. Each session started with a short explanation of the application and its features. We demonstrated the map navigation mechanism and the key interaction techniques of World Explorer. Before the experimental tasks began, we let the user explore the application independently for a few minutes (the application was set to view the Washington, DC area for this exploration). The different tasks attempted to cover the main tourism-oriented motivations for interacting with such online geotagged photo collections. Since the World Explorer data was available throughout the world, we could often tie the tasks to individual participants' experiences, as described below. Each task was followed by a short unstructured discussion to elicit feedback.

Task 1 was the "vacation recap" task. For this first task, we asked the participants for the location of their latest vacation. We then asked the participants to pretend the interviewer was a personal friend, and requested that they use World Explorer to tell the interviewer about their trip, using only the application and the public photos available through it.

Task 2 was the "San Francisco task". In this task, the interviewer asked the participant to tell them about San Francisco (all our participants live in the San Francisco Bay Area and were relatively familiar with the city and its features). Particularly, the participant was asked to pretend the interviewer was a friend visiting for the first time from out-of-town, and give the interviewer a quick photo-based overview tour of things to see/do in San Francisco.

Finally, Task 3 was the "unfamiliar area" task. We asked the users to explore an area that is unfamiliar to them. To make this task relevant and interesting, we asked the users about their future travel plans – which city are they going to visit next. The participants were then invited to point

the map to that area and explore it, possibly devising an itinerary of places that seemed to be worth visiting.

For the first and third tasks, we ensured that the places selected by the users had sufficient amount of data. For these two tasks, the participants were asked to choose the map area to interact with. We therefore had to make sure that the selected areas had sufficient amount of photos for a reasonable browsing experience. This requirement did not pose a problem; the first or second choice for each participant in each task fulfilled the requirement in all cases. To give an example, in Task 1, our participants told us about their vacations in Hawaii, Paris, Los Angeles and New York, among others. For the "unfamiliar area" exploration of Task 3, participants elected to explore Paris, Barcelona, Edinburgh, Chicago and more.

6.3 Discussion

We report below on the major positive and negative themes and considerations about World Explorer that have risen during the interview sessions. We supply quotes from participants for each team to illustrate key points and possible issues. Broadly speaking, the participants found the interaction fairly simple, and the visualization metaphor clear: tags represent photos taken by Flickr users.

6.3.1 Recall

One of the primary effects of the World Explorer interface was improving recall for participants while performing their tasks. The readily exposed tags in the interface often reminded the participants of locations they wanted to discuss. For Task 2 (San Francisco) in particular, many of the participants simply examined at the world explorer interface and explained the viewed tags rather than explicitly looking for specific tags and locations. As Participant 2 stated: "It brings out memories because you know exactly what to look for... the good thing is that a lot of the pictures it pulls up are familiar pictures, things that you could easily identify, point out".

For Task 1, telling about their vacation, participants were often reminded of locations that they did not otherwise remembered. The memories were triggered by both the tags and the photographs shown for tags. Participant 8 saw the `Diamond Head` tag and was reminded about visiting the Diamond Head beach in Hawaii. Participant 9 was examining Paris when she exclaimed "Oh my God! This place [points to `Rue Mouffetard`] has the best open air restaurants".

Of course, the danger of exposing tags is having the users concentrate on available tags and not dig further into the task at hand. However, this was usually not a problem, as participants often sought elements that wanted to discuss even when they were not immediately shown.

6.3.2 Discovery

A strong element of discovery was expressed in the interviews. The World Explorer visualization allowed participants to discover locations and elements that were previously unknown to them. While this aspect was most pronounced while they explored unknown places, it did also occur in places that the participants were familiar with. Participant 9, a music journalist, was surprised to find new music festivals while browsing Chicago (Task 3) – "I've never heard of this music festival [points to the tag `Pitchfork Music Festival`] or this one [points to `Intonation Music`

`Festival`)". The interface also prompted users to explore the known map areas and make serendipitous discoveries. For example, while looking for locations in New York city, Participant 1 was attracted by other tags from nearby locations ("I have heard lots of stories about Coney Island... now I'm already distracted.").

6.3.3 The Needle and the Haystack

A number of issues and deficiencies with World Explorer have been pointed out by the users and the tasks, and are discussed here. Most prominently, we address the "needle vs. haystack" question in information seeking behavior (also seen in [23]) as it was raised by our participants.

While the participants could usually perform their tasks well using World Explorer, most of them expressed the need for an explicit search interface to look for specific tags/landmarks. When the participants thought of something they did not see in the interface, the missed the ability to search for it. This need was most commonly expressed when participants were looking for specific locations or landmarks in a city that they were not familiar with (tasks 1 and 3). For example, while browsing through Barcelona for Task 3, Participant 6 said "I guess what I'm looking for are bull fighting pictures. Is there a way to actually [search]?". Participant 5, recapping her vacation in Los Angeles during Task 1, said "It was easy to navigate San Francisco because I know about it, but with a place like Los Angeles, that's just flat... its like – where was Culver City again?".

Another "needle" issue occurred when a participant knew exactly which map area to browse to find photos they wanted to show, yet World Explorer had no available detail in that area. In that case, the participants would often like to have seen the photos anyway, looking for photos from a specific street corner or point of interest that was not important enough to be represented in a tag. Participant 10 was disappointed not to be able to show a French-like street in San Francisco (Task 2): "they should have photographed the French part of town, I think people would be pleasantly surprised about it".

To contrast the haystack-style interaction of World Explorer with a "needle-oriented" service, we had conducted some additional interviews where (other) participants interacted with Flickr's map system. The interviews were done under similar conditions (i.e., similar tasks). As mentioned above, the Flickr system displays photos as markers on a map, for every map area the user navigates to. The markers could be enlarged to show the actual photos. The Flickr map system also supports a search feature, which allows the user to search for photos in the current map view by keyword. Indeed, most participants used the search by keyword feature as their primary method of finding photos of interest once they had navigated on the map interface to the location of interest. This search functionality proved to be an effective method for locations the user was familiar with, and therefore could recall relevant keywords for points of interest in that location. However, for geographic locations unfamiliar to the user, the search by keyword feature proved to be ineffective. Participant 11 said "If I went to a city where I knew what I was looking for but I didn't know the name of it, I don't think I would ever find it. This probably wouldn't be a system for you do that." Participant 12 said "It's more useful where you're already familiar with the area and know the names of specific locations. You need to

know what you are looking for." Participant 13 said "You have to have a sense of what the attractions are in order to do the search", and Participant 15 was stuck - "I've never been to Edinburgh, I wouldn't know where to start." To summarize, the support that World Explorer gave users in navigating unknown areas was substantially more effective than the regular map interface.

6.3.4 Other Considerations

Several participants pointed out the possibility of using the World Explorer to share a location or experience with remote friends and family. Participant 8, for example, was interested in sharing the "vibe" of her neighborhood ("I have done Google image searches before to try and show my neighborhood that I live in here to friends... so I would use it for that"). Participant 2 wanted to share the browsing experience over the telephone ("The good thing is you can pick the phone and call someone in New York and tell them tap on Miami, Florida and you can see pictures of so and so... Its a great way to share information with other people").

The interaction with the maps and the exploration was received well by both users who are comfortable with maps, and users whose comfort level was not very high. Participant 1 ("I personally am a very map person") liked the integration for planning: "I can not only decide what I like and visit, I can even plot it out and first I do this and I don't need anyone's help...". On the other hand, despite claiming to have a "really horrible sense of direction", Participant 7 was able to not only find locations but to recall their names using the tags ("We went to record stores that I cant remember the name of either. Maybe this [points to `Easy Street Records`] was one of them").

Other specific issue raised was the lack of tags. Many of the users felt that there could be more tags in specific locations. This was most often expressed during tasks 1 and 3 and was very rarely mentioned while browsing San Francisco. We believe that this is due to relative sparseness of geotagged photographs in many locations and will be remedied as more data is added to the system.

Participants appeared to have difficulties when the visible tags did not match their mental model of a location. For instance, at certain zoom levels in the San Francisco area, the tag `Chinatown` was not visible, while `Japantown` was available; many of the users complained that Chinatown was more important and should be displayed before Japantown. This could be possibly remedied by giving the users a more personalized view of the tags where the tags are tailored to their interests. Participant 1 requested a filter to help plan a trip "If the tags were really interesting and could pick up what kind of person I am and what kind of trip I would do in New York which would be really different from someone like my dad then that filter would really help me visit New York in the way i would."

About half the participants also wanted to have more information about the tags especially in the unknown locations. Participant 3 summarized it as "now that I think about it just looking at the pictures doesn't really help, I need text to go with it. So for example 'opera' all I see are buildings that look like other buildings so it doesn't really give me a feel for what its like. If there is text behind it it will be more helpful because I don't know if I'm there to see an opera or a building...".

Many of the non-technical participants did not discover the zooming functionality of the map. While they were quite happy to drag and pan the map they did not seem to consider the possibility of zooming the map to get more information. It is possible that some better visual design on the application and the zoom slider could fix this issue, but this problem is out of the scope of this paper.

To conclude, some participants independently communicated the benefits of the World Explorer. Participant 9 would "definitely use this if I was going to a new city... It's useful because it is so detailed, a lot of the time when I'm sent to go work somewhere I'll go straight to the hotel... I would use this system to see what was in that general vicinity and use that rather than getting to there and talking to the hotel and find local information." Participant 2 liked that "if you are not familiar with that at all and have no idea what you are looking for... I mean its pretty much a 'no brainer' that the places that are highlighted are pretty popular and people do have a lot of pictures of, so it's probably a pretty important place to visit.". And finally, from Participant 6: "it would be super in trip planning".

7. CONCLUSIONS AND FUTURE WORK

We have presented a sample application that generates aggregate location-based knowledge from unstructured text associated with geographic coordinates. We have shown that salient location-based textual features can be extracted from the data and used in real-life settings. Using geotagged photos from Flickr, our World Explorer visualization allows users to explore and understand the data in a way that was previously not available.

The evaluation of World Explorer brought up issues we hope to address in future work. In particular, it was clear that user needs often required both the "needle" and the "haystack" view of the data, even while performing a single task. Clearly, the aggregate data had an important role in revealing initial trends and characteristics of the data, but for many individuals, the need to explore further for individual terms or items invited a "needle" view which our current implementation was lacking. Two easy enhancements that would help solve this issue are 1) allowing free-text search for tags associated with any photo in the currently-viewed map area, and 2) showing some indication of the existence of photos in the data even in places where we did not compute or display tags. Both these features exist in the current implementation of maps on the Flickr website. Of course, switching the between exploration, search and browse tasks is a tricky, and a heavily-studied research area (for a recent example and overview, see [20]).

In future work, we would also like to examine visualizing information based on geotagged textual data extracted from various resources such as geo-referenced blogs. Do the heuristics we employ here still apply? Can we apply the techniques described above in the same manner? Similar, yet non-location-based work in this domain [23] may indicate that the answer to these questions is positive.

Finally, we aim to mine the geographic metadata together with other types of metadata that is associated with the text like time to understand and extract knowledge beyond the visualization. For example, we have started to use location and time metadata to be able to automatically identify Flickr tags that relate to events or places [18].

8. DATA APIS

Both the Flickr geotagged data, and the processed World Explorer data described in this work are available via APIs. The Flickr API is at http://www.flickr.com/services/ while the World Explorer data and visualization is available at http://tagmaps.research.yahoo.com/. A 'TagMaps' module is also available as a customizable Flash component that can be embedded in any website, visualizing World Explorer data, or data from custom GeoRSS-formatted sources.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] A. Arampatzis, M. van Kreveld, I. Reinbacher, P. Clough, H. Joho, M. Sanderson, C. B. Jones, S. Vaid, M. Benkert, and A. Wolff. Web-based delineation of imprecise regions. In *Proceedings of the Workshop on Geographic Information Retrieval*, 2004.

[2] D. C. Bulterman. Is it time for a moratorium on metadata? *IEEE MultiMedia*, 11(4):10–17, October 2004.

[3] D. Cavens, S. Sheppard, and M. Meitner. Image database extension to arcview: How to find the photograph you want. In *Proceedings of ESRI Users Conference*, 2001.

[4] M. Davis, S. King, N. Good, and R. Sarvas. From context to content: leveraging context to infer media metadata. In *Proceedings of the 12th International Conference on Multimedia (MM2004)*, pages 188–195. ACM Press, 2004.

[5] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins. Visualizing tags over time. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 193–202, New York, NY, USA, 2006. ACM Press.

[6] R. A. Finkel and J. L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1), 1974.

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

[8] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 89–98, New York, NY, USA, 2006. ACM Press.

[9] V. Jung. Metaviz: Visual interaction with geospatial digital libraries. Technical Report TR-99-017, International Computer Science Insitute, 1999.

[10] R. R. Larson. Geographic information retrieval and spatial browsing. In L. Smith and M. Gluck, editors, *Geographic Information Systems and Libraries: Patrons, Maps, and Spatial Information*, pages 81–124. University of Illinois, 1996.

[11] Y. Leclerc, M. Reddy, L. Iverson, and M. Eriksen. The geoweb - a new paradigm for finding data on the web. In *International Cartographic Conference (ICC2001)*, 2001.

[12] E.-P. Lim, D. H.-L. Goh, Z. Liu, W.-K. Ng, C. S.-G. Khoo, and S. E. Higgins. G-portal: a map-based digital library for distributed geospatial and georeferenced resources. In *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 351–358, New York, NY, USA, 2002. ACM Press.

[13] P. Longley, M. Goodchild, D. Maguire, and D. Rhind. *Geographic Information Systems and Science*. John Wiley & Sons, 2001.

[14] C. Marlow, M. Naaman, D. Boyd, and M. Davis. HT06, tagging paper, taxonomy, flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA, 2006. ACM Press.

[15] S. Milgram and D. Jodelet. Psychological maps of paris. In W. I. H. Proshansky and L. Rivlin, editors, *Environmental psychology*, pages 104–124. Holt, Rinehart and Winston, 1976.

[16] M. Naaman, A. Paepcke, and H. Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. In *10th International Conference on Cooperative Information Systems (CoopIS)*, 2003.

[17] R. Purves, P. Clough, and H. Joho. Identifying imprecise regions for geographic information retrieval using the web. In *GISRUK '05*, 2005.

[18] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. ACM Press, 2007.

[19] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistic*, 6:461–464, 1978.

[20] R. Shen, N. S. Vemuri, W. Fan, R. da S. Torres, and E. A. Fox. Exploring digital libraries: integrating browsing, searching, and visualization. In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pages 1–10, New York, NY, USA, 2006. ACM Press.

[21] T. R. Smith. A digital library for geographically referenced materials. *Computer*, 29(5):54 – 60, MAY 1996.

[22] K. Toyama, R. Logan, and A. Roseway. Geographic location tags on digital images. In *Proceedings of the 11th International Conference on Multimedia (MM2003)*, pages 156–166. ACM Press, 2003.

[23] F. B. Viegas, S. Golder, and J. Donath. Visualizing email content: portraying relationships from conversational histories. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 979–988, New York, NY, USA, 2006. ACM Press.

[24] B. Zhu, M. Ramsey, H. Chen, H. Rosie V, T. D. Ng, and B. Schatz. Create a large-scale digital library for geo-referenced information. In *DL '99: Proceedings of the fourth ACM conference on Digital libraries*, pages 260–261, New York, NY, USA, 1999. ACM Press.