ZoneTag's Collaborative Tag Suggestions: What is This Person Doing in My Phone?

Mor Naaman and Rahul Nair
Yahoo!

We describe ZoneTag, a camera phone application that allows users to capture, annotate, and share photos directly from their phone.

Thousands of people visit the Golden Gate Bridge every day. Many (if not most) of these visitors take photographs of the bridge from the same picture spots. Despite being eight years into the 21st century, the overwhelming majority of these photographers will return home and load the photos onto some folder on their computer's hard drive, assigning each photo a highly nondescript name such as DSC02211.jpg. Surely, we can do better than that.

Collaborative tagging, combined with the growing availability of context-aware and network-connected capture devices, can address some aspects of this unfortunate situation. ZoneTag is a prototype mobile application that uploads camera phone photos to the photo-sharing site Flickr. ZoneTag can assist the users with context-based tag suggestions that are derived from multiple sources. A key source of suggestions is the collaborative tagging activity on Flickr. That photo of the Golden Gate Bridge, within seconds and at a user effort of no more than a couple of clicks, can be saved to Flickr, annotated with the exact geographic coordinates as well as the tag *Golden Gate Bridge*.

In this article, we describe the ZoneTag application and its tag suggestion feature. We describe the simple mechanism for deriving tag suggestions and the ensuing interaction design for presenting these suggestions to the user. We also discuss quantitative and qualitative results from an 18-months deployment of ZoneTag, emphasizing the way people use and understand tag suggestions. In addition, we highlight several emerging issues that could play an important role in the collaborative tagging for multimedia as well as other resources. While the quantitative study on the use of the suggested tags feature implies clear benefits for tag suggestions, a set of qualitative studies imply that while tag suggestions are helpful, there are multiple issues that arise and require careful consideration.

Collaborative tagging and context

We do not need to provide here an in-depth review of the benefits of collaborative tagging. While tagging is a lightweight way to organize and annotate information, collaborative tagging systems can enhance and advance the benefits of tagging. Examples of collaborative tagging's benefits include enhancing the span and impact of tags for any specific resource, enhancing the number of tagged (and therefore, organized) resources in the system, and distributing the workload for organization among multiple contributors.[1]

Assisting the tagging process could therefore improve the information in the system for the individual as well as for the community. One way to assist tagging is by supplying users with tag suggestions, aiding tagging by providing an easy input mechanism (typically just a selection, rather than typing) and by lowering the mental effort of thinking of new tags to add. Several existing systems (for example, del.icio.us) implement tag suggestions. In addition, several research efforts discuss tag-suggestion systems.[2,3]

However, there is a dark side to tag suggestions. Suggested tags might lead to a quick convergence of tags, and to consolidation that isn't necessarily beneficial. At worst, tag suggestions can potentially reduce the breadth and diversity of information for each particular resource, or for the system as a whole.[4] Such reduction in diversity of descriptors can damage

1070-986X/08/$25.00 © 2008 IEEE

Published by the IEEE Computer Society

a user's ability to find tagged resources in the system.[5]

Two observations about tagging systems will be helpful to frame the discussion here: the source of suggested tags and the structure of tagging systems. Generally, suggested tags can be generated from other tags assigned to the same resource (usually by the other users, as with del.icio.us[3]) or from tags assigned to similar resources. In ZoneTag, tag suggestions for a resource (a newly-captured image) are based on tags assigned to similar resources. The system judges the similarity using contextual metadata in the social, spatial, and temporal dimensions. Secondly, Flickr (and therefore, ZoneTag) is a set-based, self-tagging system according to the Marlow et al. model of tagging systems.[1] In other words, the images on Flickr are largely tagged by the users that uploaded each image. Those tags are shown as a simple list, and a tag cannot be added to the same image more than once.

ZoneTag's tag suggestions

To understand ZoneTag and its collaborative tagging features, we first briefly discuss the interaction flow for ZoneTag image capture, tagging and upload. We then describe the way ZoneTag generates tag suggestions.

ZoneTag is a mobile application available as a public prototype for Nokia and Motorola phones, available at http://zonetag.research.yahoo.com; our description here applies to the Nokia version of ZoneTag. With ZoneTag, users can upload a newly-captured image from their camera phone to Flickr in as few as two clicks. After a photo is captured, ZoneTag displays an upload dialog over the image (see Figure 1a). The user can review the photo and decide whether he or she wishes to upload it to Flickr. If the user decides to upload the photo, ZoneTag shows more options, including the photo's privacy settings, title, and location, which the user can update or hide. Clicking the same button again will upload the image, keeping all the settings and tags entered for the previous photo.

Most importantly, in the pre-upload dialog, the user has the opportunity to type in or select tags to apply to the photo when it's displayed on Flickr. The tag-entry dialog is optional and only appears when the user selects it. This dialog includes the context-based tag suggestions that are prefetched from the ZoneTag server and sorted by the likelihood that the user will select a particular tag. This ranking is especially important on a mobile phone's small screen because the user can typically see only seven tags on one screen without scrolling. The tags are grouped into categories for ease of lookup; users can tab right or left to see the tags in each category. Within each tab, a user can search through the available tags by scrolling or by entering the first few letters of a tag in the search box. The categories roughly correspond to the different sources of suggested tags. Figure 1b shows a sample screen of the tagging interface.

(a) (b)

Figure 1. (a) ZoneTag's upload dialog and (b) the optional tag suggestions and entry dialog.

ZoneTag captures the cell-tower or GPS (if available) information for location data from the phone.[6] This location data is used both for automatic annotation of the image with location information, as well as for the tag suggestion mechanism. There are three types of location data that are available through the ZoneTag client. The cell tower information is always available; the real-world location of the cell tower is sometimes available from our cell-tower-to-location database;[6] and GPS is available on certain devices. When possible, when a photo is uploaded, the system converts the phone's location information into human-readable location labels (that is, city, state, country, and zip or postal code) that are automatically added as tags to the photo's page on Flickr together with the set of user-provided tags.

ZoneTag uses the location information to retrieve tag suggestions. While ZoneTag is

running on the phone, the client periodically sends the location data to the server. The location data is used to provide context and help the server generate the tag suggestions, which are then sent back to the ZoneTag client. This prefetching of tag suggestions is done in an interval of fewer than 10 minutes, so that whenever the user takes a new photo, the list of tag suggestions for that photo is already stored on the phone (reducing the latency for displaying the tag suggestions). At the same time, the tag list is relatively fresh, usually updated minutes before the photo was taken.

Deriving tag suggestions

The tags that ZoneTag suggests are derived from several different sources. In the following sections we describe the various tag sources, and how the user's context determines the specific tags that ZoneTag suggests. In other words, we discuss how we take the context data and produce a prioritized list of tags to be used in the ZoneTag interface. Thanks to the interface autocomplete feature, the tag list can be quite long (often including hundreds of tags), as the user does not necessarily have to scroll down to find the tag of choice. However, the most likely tags would ideally appear on the top of the list for easy selection (as shown in Figure 1b).

The main contextual driver of ZoneTag's tag suggestions is the available location information, in the form of a cell tower, a zip (postal) code, or exact GPS coordinates. In the following discussion, we assume some location context G where G could be any one of the following: cell tower, zip code, radius, bounding box, and so on. We describe how we produce suggested tags for a user in that location. When a cell tower ID with no real world location represents G, we refer to the location as a *virtual location*, and some of the methods described cannot be used.

Tag sources and weighting

The first source for tag suggestions is the user's own tagging history. These tags, prioritized to the current context, are often a probable match for the user's new photo. In other domains, research had demonstrated that most users have a relatively small tag vocabulary, and the growth rate of the vocabulary diminishes over time—two facts that suggest that the user's own past tags are indeed useful suggestions.[1,4] In ZoneTag, the tag suggestion algorithm suggests tags the user had applied in the same context (that is, in the same location G). This way, for example, if the user is at work, the tag *work* is likely to be suggested if the user applied that tag to prior photos from that location. Beyond that, the list of tag suggestions includes the user's 400 most frequently used tags, and all tags the user has ever entered on the phone, as these are likely to be reused.

The second source of tag suggestions is the tagging activity in location G by other users in the Flickr community. We differentiate between tags used by the user's friends and family (as defined by the user on Flickr's social network), and tags used by the general Flickr community at that location. For example, if the user is at home, the tag *home* might be suggested, if their spouse used that tag in that location before. In another example, the tag *Golden Gate Bridge* might be suggested in a location where many Flickr users had used it, say one of the popular viewpoints of the bridge.

We balance all the tag sources to generate a prioritized suggested tag list. As mentioned previously, most suggested tags are derived for a specific location G; to generate a ranked list of suggestions, we use several heuristics that take into account the tags' social and temporal context, and other measures that weigh the tag frequency to create a final score. These heuristics include the following:

- *Spatial*. We consider all tags used in location G regardless of the exact location. Alternative schemes could assign a higher score to tags that appeared in photos taken in close proximity to the current photograph.

- *Social*. Tags the users themselves applied in a given context are more likely to apply to their current photo than tags used by others. Tags used by the user's social contacts on Flickr are more likely to match the user's photo than tags applied by others in the Flickr community.

- *Temporal*. We apply temporal heuristics for the tag computation: recency, uniformity, and cyclicality. Tags are more likely to

apply to a photo if they have been used recently (for example, in the last few hours by other users, or in the last 24 hours by the same user). Tags used across a long time period in a location are more likely to apply than tags that just appeared on a single occasion. Finally, tags used cyclically in the current context are given a higher likelihood. For example, if the tag *mobile seminar* is used in location G every Friday, it would be given a higher weight on Fridays.

ZoneTag uses each of these heuristics to compute a probability score for each tag. The score from the individual heuristics are then fused together to compute a final score for each tag, indicating the likelihood of the tag's use in the user's current context. Finally, the tags are ranked in decreasing score order so that the most likely tags appear at the top of the list presented to the user.

Additional tag sources

In ZoneTag, we use additional sources for tag suggestions beyond the user and community activity on Flickr. In general, tag suggestions can be generated from other sources on the basis of an item's content (for example, image or text) and context, whether virtual (links), physical (location and time), or other. In our case, ZoneTag uses context information to find contextually-relevant tag suggestions using spatial and event databases. The Yahoo! Local database is used to find a list of restaurants and cafes and display those as tag suggestions, as Flickr users often apply this type of tags to images.[7] Yahoo! Upcoming provides a user-generated listing of events and venues that serve as tags: for example, a user at a U2 concert might take a photo and then find the phrase *U2 live* and the venue name at the top of the tag suggestion list. Though the user sees only the names of the restaurants, venues, and events as text tags, ZoneTag maintains an internal representation connecting the text tag to a semantically identified object.

Finally, we allow users to extend the functionality of the ZoneTag suggested tags by accepting any RSS 2.0 feed—, for example, a feed of entries from the user's online calendar—as a tag source. However, users rarely discovered this feature.

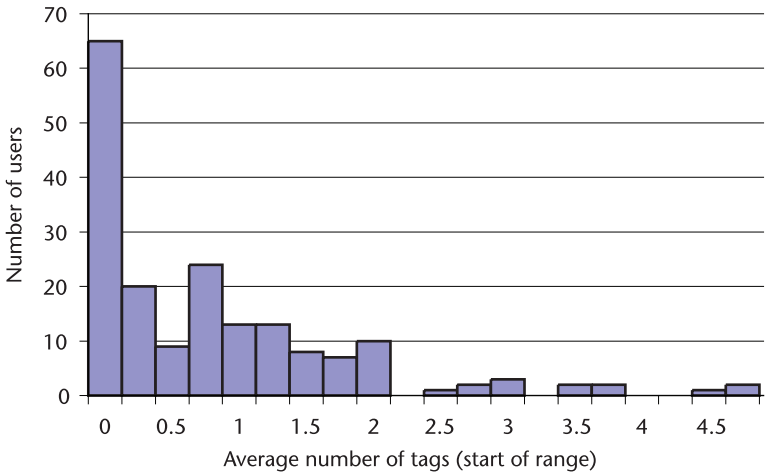

Figure 2. Histogram of average number of tags per user.

Tagging by numbers

ZoneTag has been publicly deployed for over 21 months. In that period, more than 1,500 people have used the application, uploading more than 100 thousand photographs taken in 60 different countries. About 47.2 percent of the uploaded photographs had at least one human-entered tag associated with the photo. Furthermore, in only 18.4 percent of these photos with mobile tags did the user type in at least one new tag on the phone after taking a photo. On the other hand, 91 percent of the tagged photographs had at least one tag that was not keyed in (either suggested or kept from the previous photo). This latter finding suggests that the default behavior of keeping metadata from one photo to the next is effective, and indicates that tag suggestions are effective.

More users added tags to their photos from the mobile device rather than doing so while visiting Flickr's Web-based interface after the photo was uploaded. People used the Flickr Web interface to add tags to around 10 to 15 percent of the total number of photos uploaded using ZoneTag; roughly half of these photos already had at least one tag added from the mobile device (the exact numbers in this category are difficult to specify due to logging issues). We aren't able to tell how many users reviewed their photos on the Web after the photo upload from the mobile device.

Figure 2 shows the breakdown in tagging activity between the most active ZoneTag users (over 180 users that uploaded at least 40 photos each). The figure shows a histogram of the average number of tags applied by each user to their ZoneTag photos. For example, 65

users (36 percent) applied, on average, fewer than 0.25 tags per photo. The aggregate of all bars, excluding the leftmost four, indicates that 35 percent of ZoneTag users added, on average, at least one tag per every ZoneTag photo they uploaded. This count, and the other statistics we report in this section, don't include the automatic tags that ZoneTag adds to photos. Indeed, Figure 2 indicates a varied set of tagging habits among our users.

Does local context (and social proximity) help reduce the effort of tagging or encourage users to add tags? To investigate that question, we looked at the distribution of tags in two specific locations, both of which enjoy a high volume of ZoneTag use. First, many ZoneTag users work in the Yahoo! headquarters in Sunnyvale, California (see Figure 3a); some of them know each other and are connected via Flickr. Second, ZoneTag users were active in many locations in San Francisco (See Figure 3b); we show data for the Mission neighborhood, popular for its restaurants, entertainment, and other attractions. In total, the data represents 70 users in the Sunnyvale location and 42 in the Mission (San Francisco) location.

Figure 3a represents the tagging activity of the location- and socially-proximate group of users in Sunnyvale, and is an example for the significant tag overlap that can exist between users in the same location and social context. The figure shows the breakdown of all tags added to photos taken in the Sunnyvale zip code where the Yahoo! headquarters is located. The number of tag instances (the number of times each tag was applied to a photo) is binned by the number of users that employed each tag. For example, the top-right part of the pie (in black) represents the count of instances for tags used by just one person. These tags represent only 17 percent of all tag appearances in this location. Conversely, in 2,845 cases of the tag applications (a total of 39 percent of all tag instances, in the white area at the top left), the tag used was shared between at least 16 users (out of 70 users in that location). The most shared tags for this location, of course, were ones that generally described the place and its social context—tags such as *Yahoo!*, *Yahoo! Sunnyvale*, and *URL's Café*.

Figure 3b shows similar data for the San Francisco Mission neighborhood (all ZoneTag photos taken in zip code 94110). We expect less social connectivity between the users taking photos in this location. Indeed, there is a higher portion (36 percent) of tag instances where only one person used the tag. However, in a significant two-thirds of the cases where tags were used with photos in this area, there was an overlap with at least one other user. The most shared tags for this location were ones that refer to the neighborhood—for example, *The Mission*.

Figures 3a and 3b demonstrate that at least for these locations, it was more common than not for ZoneTag users to employ the same tags. This finding provides verification to the approach of collaborative tagging based on location. The differences between the two locations hint at some effect of the social context of tagging. In any case, these are just preliminary results and should be investigated in depth in a future study.

Considerations for tag suggestions

What was the effect of the tag suggestions on the users' tagging behavior and activity? In a series of interviews, we qualitatively investigated the tagging activity in ZoneTag and Flickr. In these interviews, we discussed each user's tagging activity. Elsewhere, we report on the set of tagging motivations that emerged in our interviews.[7] Here, we expand on the main findings regarding acceptance and use of tag suggestions, as reflected in that earlier work and subsequent interviews.

In all of our studies, we had some participants who were not initially familiar with tagging or Flickr. These participants often felt that tagging wasn't worth the effort; one oft-cited reason was that most photos would end up having the same tags (for example, the names of the user's kids). Some participants often entered titles to their photos instead, feeling that titles would more likely attract the attention of novice Flickr users.

Because some people never use tags (as also demonstrated by the data in Figure 2), the fact that the tagging interface was optional, and is skipped by default, proved to be a successful design choice. Even those who did use tags enjoyed being able to skip tag entry in certain situations that do not allow interruption in the moment of capture.

Another useful shortcut employed frequently by users was tag stickiness: the fact that tags remained from one photo to

another allowed users to set up tags once for an event and then upload photos with the same tags without having to interact with the tagging interface after each photo. In addition, the autocompletion feature for searching through the suggested tags list was welcome, especially in this mobile context. Several users commented on how suggested tags, via the autocomplete mechanism, reduced their tagging work and cut down on cumbersome text entry.

While the way that ZoneTag produced the suggested tags was not clear for all our users, occasions where the suggestions were helpful were appreciated. One participant reported how the suggestions were effective when she was attending a conference with other ZoneTag users who created many relevant tags. In other cases, users pointed out their delight when the suggested tags included the exact restaurant the user was in or the name of a friend who was captured in the photograph.

On the other hand, when ZoneTag did not get it right, tag suggestions were confusing, inexplicable, or even disturbing. Such occurrences included suggesting a tag with the wrong neighborhood name, or (worse) showing tags perceived by the user to be unrelated to their current context or their activities in general. In particular, because the system cannot distinguish different types (or semantics) of tags, unknown person names entered by another user near the same location sometimes appeared in the tag list, thus confusing some participants. One participant commented about an unfamiliar name appearing in her suggested tags: "This person was in my phone for a month! Who is she and what is she doing in my phone?"

Users selected tags from the suggested list even if they hadn't originally intended to add those tags. Some participants reported scrolling down the tag-suggestion list and adding any relevant tags. This phenomenon also has negative implications: some participants stated they added such tags from that list even if they weren't entirely relevant or accurate.

The list of suggested tags often served as inspiration for the users looking to tag their photo. The suggested tags therefore played a role even when not selected by the users. Seeing certain tags on the list inspired some participants to add their own tags and gave them direction as to the sorts of tags they should use. For example, on participant noted that seeing neighborhood names as suggested tags inspired them to add a tag for the neighborhood they were in, even though that name did not appear as a tag suggestion.

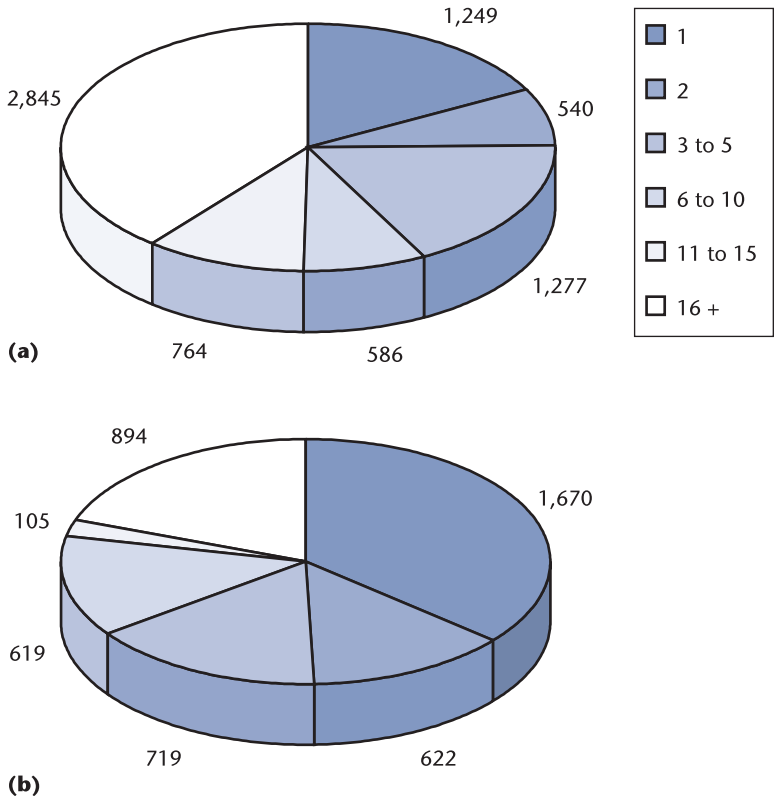

Figure 3. Breakdown of tag instances in one location: (a) Sunnyvale and (b) San Francisco, California. The categories are the number of ZoneTag users that tagged a photo in the location. The slices represent the number of instances of tags in each category.

Did it work?

Our quantitative and qualitative results suggest that ZoneTag's context-based tag suggestions are beneficial and useful for tag entry, and probably contribute positively to the number of tags per resource. A significant portion of our users participated in this collaborative-tagging activity. The tag overlap between users in the same location indicates that the system's suggestions have merit.

Mobile tagging features, such as the tag suggestions list selection and autocomplete from the suggestions list, were welcome and often used by those users who engaged in mobile tagging. These users also enjoyed the fact that the tags applied to one photo were automatically kept for subsequent images, showing a strong temporal aspect of tagging personal multimedia content. We saw little tagging activity on the Web, after photo upload, which indicates that the phone is the primary channel and almost the last chance for content annotation.

We also reestablished that tagging is not for everyone. Many of our users never engaged in tagging from the phone or the Web: some didn't understand or questioned the usefulness of tags; others didn't care to divert their attention after taking a photo, or didn't have time after the fact to add tags or curate their collection on Flickr. Other tagging options that don't require immediate attention after capture might somewhat improve this state of affairs.

Even for those users who do tag, the tag suggestions were often a mixed bag. When the tag suggestions were useful and appropriate, they were received positively. However, especially among novice users, suggested tags were often bewildering, and even bothersome when the tags didn't seem to match the user's mental model. More careful filtering of tags might be in order.

Finally, the issue of over-tagging, where users add any suggested tag that barely relate to the content, might eventually pose a system-wide problem of tag accuracy. There is no evidence that such practice is commonplace, but those working on tag-suggestion systems might want to consider this issue when developing applications to supply users with tag suggestions.

Overall, we think the positive elements in suggesting tags in a collaborative system outweighed the negative, as we allowed people to add context quickly and easily by tagging their personal photos. Future work might explore different tagging mechanisms, paths for encouraging people to tag, algorithms for generating better (more robust) tag suggestions, and an investigation of the information value of individual tags for each resource. **MM**

References

1. C. Marlow et al., ''HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read,'' *Proc. Hypertext 2006*, ACM Press, 2006, pp. 31-40.
2. S. Sood et al., ''TagAssist: Automatic Tag Suggestion for Blog Posts,'' *Proc. Int'l Conf. Weblogs and Social Media*, 2007; http://www.icwsm.org/papers/paper10.html.
3. B. Sigurbjörnsson and R. van Zwol, ''Flickr Tag Recommendation Based on Collective Knowledge,'' *Proc. World Wide Web Conf.*, ACM Press, 2008, pp. 327-336.
4. S. Golder and B.A. Huberman, ''Usage Patterns of Collaborative Tagging Systems,'' *J. Information Science*, vol. 32, no. 2, 2006, pp. 198-208.
5. G.W. Furnas et al., ''The Vocabulary Problem in Human-System Communication,'' *Comm. ACM*, vol. 30, no. 11, 1987, pp. 964-971.
6. S. Ahern et al., ''Reliable, User-Contributed GSM Cell-Tower Positioning Using Context-Aware Photos,'' 2006; http://infolab.stanford.edu/~mor/research/ahernUbi06celltext.pdf.
7. M. Ames and M. Naaman, ''Why We Tag: Motivations for Annotation in Mobile and Online Media,'' *Proc. Conf. Computer–Human Interaction* (CHI), ACM Press, 2007, pp. 971-980.

Mor Naaman is a research scientist at Yahoo! His research interests include social media, mobile computing, interactive multimedia systems, and location- and context-aware computing. Naaman has a PhD in computer science from Stanford University. Contact him at mor@cs.stanford.edu.

Rahul Nair is a researcher in the Early Stage Products group at Yahoo! Connected Life. His research interests include ubiquitous computing, intelligent interfaces, and context-aware computing. Nair has an MS from the human–computer interaction program at the Georgia Institute of Technology. Contact him at rnair@yahoo-inc.com.