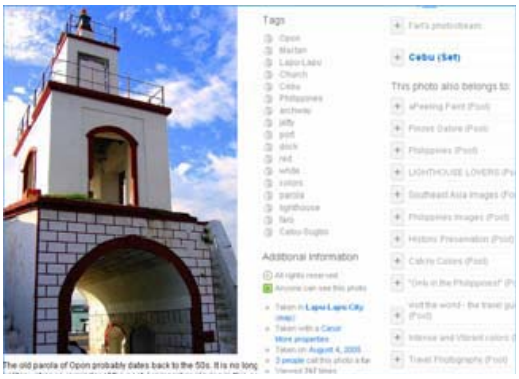
Harvesting Geospatial Knowledge from Online Social Networks

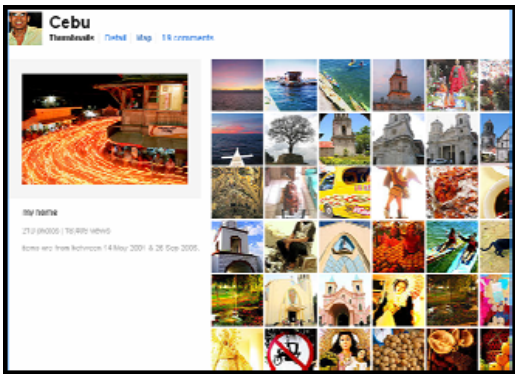
KRISTINA LERMAN
Information Sciences Institute
University of Southern California
Phone: (310) 448-8714
Email: lerman@isi.edu

Social Web has moved knowledge production from the hands of the experts and professionals to the masses. Today online social networking sites, such as Twitter, Facebook, YouTube, and Flickr, allow ordinary people not only to create massive quantities of new data, but also organize it, use it, and share it with others. Unlike earlier information technologies, the Social Web exposes social activity, allowing each person to observe and be influenced by the actions of others in real time. How will such real-time, many-to-many communication change how we discover, use, and manage information? And how will it transform society and how we solve problems? My research addresses these questions by developing methods to harvest social knowledge.

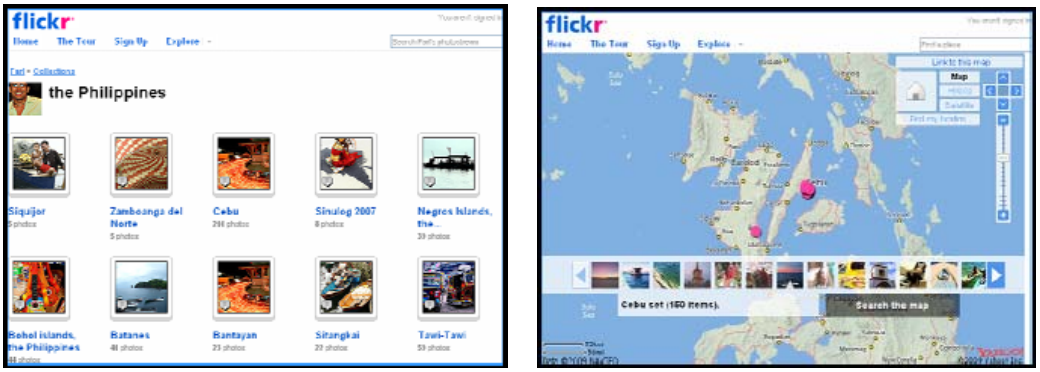
Consider a gazetteer, for example, Geonames.org, which compiles geospatial knowledge within a directory of places and place names, often organizing it hierarchically within taxonomy of geospatial concepts. Such gazetteers have been useful for creating geo-aware applications and integrating geospatial knowledge. However, since gazetteers are manually and painstakingly created by an expert or a small group of experts, they are rarely complete or comprehensive, do not reflect the variety of views, and fail to keep up with our changing ideas about places.



(a)



(b)



(b) (d)

***Figure 1**: Examples of data and metadata created by a Flickr user. (a) Tags assigned to an individual image (geotags are not shown), (b) images in the set "Cebu", (c) sets in a collection called "the Philippines" created by the user, (d) geotagged images in the "Cebu" set displayed on the map.*

Social Web has given ordinary people the ability to organize information, including geospatial information. A person can annotate content they create, whether a photograph, a blog post, or a tweet, with descriptive labels known as tags, attach geographic coordinates to it, or organize it hierarchically within personal directories. These annotations help people browse content or retrieve specific items at a later date. While an individual's annotation expresses her particular worldview, collectively social annotation provides valuable evidence for harvesting social knowledge. My group is developing machine learning methods [Plangprasopchok et al., 2010a; 2010b] to combine annotations created by many individuals into a common hierarchy, a *folksonomy*, that reflects how a community organizes knowledge [Plangprasopchok & Lerman, 2009].

Error! Reference source not found.(a) shows an image on *Flickr*, along with metadata associated with it, which includes descriptive tags. *Flickr* allows users to group photos in folder-like *sets*, and group sets in *collections*. The image in Error! Reference source not found.(a) was grouped with other images taken around the Philippine province of Cebu in an eponymous set (Error! Reference source not found.(b)). This and sets devoted to other places around Philippines were grouped together in a collection called "the Philippines" (Error! Reference source not found.(c)). In addition to tags, users can attach geospatial metadata to photos. Error! Reference source not found.(d) shows geotagged images (purple dots) in the "Cebu" set on a map.

While geospatial knowledge expressed by a single user may be noisy, ambiguous, and incomplete (e.g., points in Fig. 1(d) only cover a small portion of Cebu), combining data from many different people will provide more accurate knowledge. We have developed a method that aggregates geotagged content created by thousands of users of the social photo-sharing site Flickr to learn geospatial concepts and relations between them [Intagorn et al., 2010]. Our method leverages geo-referenced data to represent and reason about places. We have evaluated the learned geospatial relations by comparing them to a reference ontology provided by GeoNames.org. Our approach achieves good performance

and also learns novel relations that do not appear in the reference gazetteer. Such folksonomies may eventually aid people in searching, browsing, visualizing, managing, and personalizing information. In addition to learning folksonomies, we have extended machine learning methods to learn better topic models of annotated documents [Plangprasopchok and Lerman, 2007; 2010]. We showed that these methods help us exploit social annotation to discover relevant new information sources more effectively than a search engine [Plangprasopchok and Lerman, 2007; Ambite et al., 2009].

What we learn from social annotation is often surprising. For example, some Flickr users annotate spider with the term "insect," as shown in Fig. 1, and Disneyland with "LA." Technically, they are wrong, since spiders, in possession of eight legs, are certainly not insects, and Disneyland is not even a part of Los Angeles County, much less the city of Los Angeles. Though it may be wrong, this information is still useful. If you want to find images of spiders, sometimes you will have to look under "insects." Likewise, if you are traveling to Los Angeles with children, you better know about Disneyland. This type of *folk knowledge* is often at odds with expert-generated knowledge expressed in formal taxonomies, but you need folk knowledge to make the most of social information.

References

[Ambite et al., 2009] Ambite, J.-L., Darbha, S., Goel, A., Knoblock, C. A., Lerman, K., Parundekar, R., and Russ, T. A. (2009). Automatically constructing semantic web services from online sources. In *Proceedings of 8th International Semantic Web Conference*. http://http//www.isi.edu/integration/papers/ambite09-iswc.pdf.

[Intagorn et al., 2010] Intagorn, S., Plangprasopchok, A., and Lerman, K. (2010). Harvesting geospatial knowledge from social metadata. In *Proceedings of 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM10)*. http://www.isi.edu/integration/papers/intagorn10-iscram.pdf.

[Plangprasopchok & Lerman, 2007] Plangprasopchok, A. and Lerman, K. (2007). Exploiting social annotation for automatic resource discovery. In *Proc. of AAAI workshop on Information Integration from the Web*. http://arxiv.org/pdf/0704.1675.

[Plangprasopchok & Lerman, 2009] Plangprasopchok, A. and Lerman, K. (2009). Constructing folksonomies from user-specified relations on flickr. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 781–790, New York, NY, USA. ACM. http://portal.acm.org/citation.cfm?id=1526814.

[Plangprasopchok & Lerman, 2010] Plangprasopchok, A. and Lerman, K. (2010). Modeling social annotation: A bayesian approach. To appear in *TKDD*. http://arxiv.org/pdf/0811.1319.

[Plangprasopchok et al., 2010a] Plangprasopchok, A., Lerman, K., and Getoor, L. (2010a). Growing a tree in the forest: Constructing folksonomies by integrating structured metadata. In *Proceedings of 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. http://www.isi.edu/~lerman/papers/sap10final.pdf.

[Plangprasopchok et al., 2010b] Plangprasopchok, A., Lerman, K., and Getoor, L. (2010b). Integrating structured metadata via relational affinity propagation. In *Proceedings of AAAI workshop on Statistical Relational AI*. http://www.isi.edu/~lerman/papers/star10final.pdf.