PhotoSense: Emergent Semantics Based Approach To Image Annotation

Rohit Ashok Khot
Center for Security, Theory, and Algorithmic Research (C-STAR), International Institute of Information Technology, Hyderabad,
Hyderabad, INDIA
rohit_a@research.iiit.ac.in

Kannan Srinathan
Center for Security, Theory, and Algorithmic Research (C-STAR), International Institute of Information Technology, Hyderabad,
Hyderabad, INDIA
srinathan@iiit.ac.in

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
Tagging of images using descriptive keywords (tags), contributed by ordinary users, is a powerful way of organizing them. However, due to the richness of the image content, it is often difficult to choose tags that best describe the content of the image to the viewing audience and ensure access to the image. In this paper, we present a novel tagging framework based on the theory of emergent semantics to assist the user in the tag selection process. Our idea is to enrich the current "looking at" experience of tagging with the "looking for" experience of searching. We describe the design of our approach along with a preliminary user study conducted with a prototype Flickr application.

Keywords
Emergent Semantics, Tagging, Image Annotation

ACM Classification Keywords
H.3. [Information storage and retrieval]: Content Analysis and Indexing.

General Terms
Experimentation, Human Factors, Theory.

Introduction
Recent advances in digital technology have simplified the art of photography. With handheld digital cameras and cell phone cameras, we can easily capture

Figure 1: If indeed, an image can be worth a thousand words then which words will best serve the intent of the collection and ensure access to it?

photographs and share them through a web based archive. Photo making has become almost a routine part of life and visual information (images and videos) is now widely available on diverse topics and from multiple sources. This is an advantage and at the same time challenge since user cannot view such a large quantity of information at once. It is imperative, therefore to devise efficient ways to browse through the image collection and to perform queries based on the specific need.

In the recent times, *Tagging* has emerged as a simple yet effective way of organizing images using descriptive keywords (also known as tags). Unlike the traditional methods which involve hiring people such as librarian and indexers to review each image in the collection and add keywords to it, tagging enlists ordinary users in the image annotation process, effectively giving us keywords (tags) at a very low cost [6]. Tagging is open ended and non hierarchical in nature, which allows users to assign tags freely to an image based on the cognitive connection they share with the image. Tagging is quite popular among users [1] since it helps in managing personal information and provides a much desired social platform to express themselves, share ideas, and have fun. Notable examples of tagging based image annotation systems include a public photo sharing website, Flickr [2] and online game ESP [7].

However, the task of selecting tags that best describe and explain the content of the image to the viewing audience is a difficult one. The inherent richness of an image can create a dissonance between the assigned tags and the possible search queries for the image, thereby hindering its accessibility. Suppose, "*Teddy bear*" and "*Key chain*" are the assigned tags for the image in Figure 1. However, if any user searches for the same image with different tags such as "Teddy bear" and "Green coat" then she may not find it, due to the mismatch in the interpretation of the image with the tagger.

In this paper, we introduce a tagging framework based on theory of emergent semantics to assist users in the tag selection process. Unlike the previously proposed approaches [5, 8], which mostly concentrate on tag recommendations, we advocate a novel approach: We recommend images that share the same tag and provide feedback about the tag rank (Similar to a tag based search result). Tag rank denotes the number of images within the collection that share the same tag. The idea here is that "*knowing about*" the image collection and current tagging trend will enhance the experience of "*looking at*" the image and will prompt (encourage) user to improve the quality of the tags. In this paper, we describe the design of our approach along with a preliminary user study conducted with a prototype Flickr application.

A Perspective on Tagging

The motivation for this work comes from apparent differences that are observed while tagging an image and while searching for an image. Let us first look at the tagging scenario.

Tagging: A "Looking at" Experience

Most systems present the image to be tagged to the user and ask her to tag it as shown in Figure 2. This is a typical "*looking at*" experience, where the user tries to answer the question "*What I am seeing?*" with the entered tags.

Tagging:
No Assistance

Searching:
Assistance with
Related images

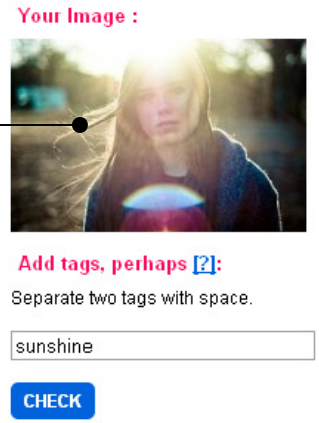

Figure 2: A common tagging scenario: "look at" the image and tag

Although the intent of the tags is to make the image accessible to others who might be searching for the same image, it is difficult to predict the tags they will be searching with, accurately. Therefore, while tagging, user often follows her own interpretations of the image and tags accordingly. However it is likely that in doing so, she may fail to notice other possible interpretations of the image. In effect, the tagged image remains accessible only to those who interpret the image similarly or it is lost in the crowd of similarly described images.

Searching: A "Looking for" Experience

Unlike the tagging scenario discussed earlier, searching follows a "*looking for*" experience. However, most of the times, user is not sure about what she is looking for. Therefore, the system assists user with a set of images (found for the given tag from the image collection) to help her decide what she actually wants as shown in Figure 3. She can then either browse through the presented set to find the desired image or refine her tags to get a better set of images.

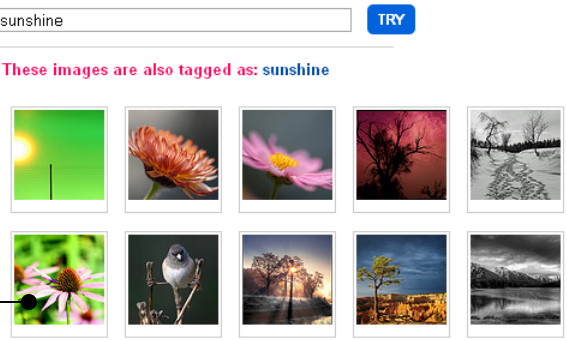

Figure 3: Typical searching scenario: "look for" the images that closely match what you desire, else refine tags.

Notice that there is a difference in the manner a user searches for an image and the manner in which she tags the same image. To illustrate, tagging involves listing down possible image interpretations normally in one go (which can be problematic as discussed earlier). Searching on the other hand is more iterative, where user adds some basic tags to begin with and finally after '*n*' iterations settles with the desired tags for the image. Each iteration helps user with set of related images (images that share the same tag) in making a judgment on the next set of tags for the image. Most of the times, the final tags that a user settles with are the tags that best describe the content of the image. We therefore, ask following question:

"*If we merge the tagging and searching scenarios together, will it be a better tagging experience?*"

To answer the above question, let us follow a small tagging experiment and try to describe the image in Figure 4.

Figure 4: How will you describe (tag) this image?

We are quick to reply as "*It is an Apple*". Now, place the same image, within a group of other images as shown in Figure 5 and now again try describing the first it. How will you describe the same image now?

Figure 5: Is 'apple' still a good tag for the first object (image)?

Our discriminating mind now notices that even if all images are of *Apple*, the first image is different from the rest of the images. That is, first image is of *Apple fruit*, while remaining images are in fact *Apple logos*. We therefore, update our description for the image, and say "*It is an apple fruit.*" If we progress in the same way and assemble the same image with other images of *Apple fruit*, then our description becomes even more precise and we say "*It is a green apple fruit, partially eaten from left.*"

This small tagging exercise shows our tendency of oversimplifying things; that is "*we do not always say what we see"* [3]. We all knew and saw the features of the image from the beginning, but we never felt the need to express them completely. Rather we described the image with a generic description - '*an Apple*'. However, when the image was shown along with other related images, we were compelled to describe the image with more precision. In fact, this observation is well supported by the theory of Emergent Semantics [4]. According to this theory, image in general does not have meaning, but the meaning emerges from the interaction with the user and by placing the image in the context of other images. In other words, meaning of the image can be manifested either by differentiation between an image which possesses that meaning and image which does not or by association between different images that share the same meaning. The tagging exercise in the prequel was the proof of later. Thus, we strongly believe, showing related pictures (similar to a search result) can improve the tagging experience and will yield better tags for an image.

PhotoSense: Framework and Prototype Design

We present a novel tagging framework based on the theory of emergent semantics to assist user in the tag selection process. It is an iterative process of adding tags to an image by repeatedly placing the image in the context of other images that share the same tags. At the beginning, there can be many images those correspond to the entered tags, but gradually over time, the process helps users to come up with better tags for the image so that less number of images corresponds to the added tags, improving accessibility of an image. Figure 6 shows a working example of tagging an image using the proposed framework. We explain the working with a simple online tagging application called PhotoSense which we designed for Flickr users (Refer Figure 7).

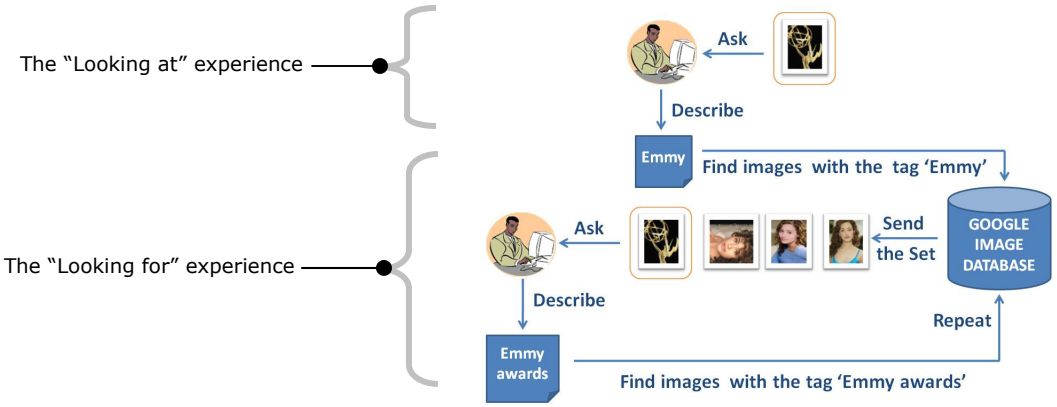

Figure 6: A tagging example using our proposed emergent semantics based framework

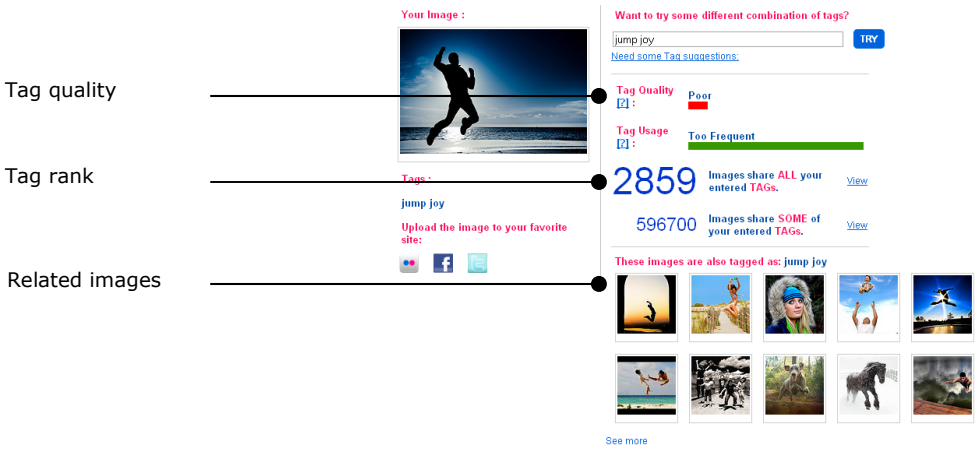

Figure 7: Screenshot of the PhotoSense application.

User begins the tagging process by adding some basic tags of her choice to the image. We then search for images that share the entered tags within Flickr image database and display them along with the tag rank as shown in Figure 7. Here, Tag rank is calculated based on number of images that corresponds to entered tags in Flickr database. This information is a vital feedback about the image collection and current tagging trend. If there are too many images that share the same tag, then there is a high probability that the image she wants to share may not be accessible with the entered tags, since searcher may not browse more than the first few pages of search results. User therefore updates (improves) her tags so that there are only few images that share the same tags. She can do it by trying different combinations of tags and by inspecting the returned search results. If we gather all tags that are tried by the user, then we will have a rich collection of tags, describing the content of the image, in different manner.

User study

We conducted a formal user study to evaluate the proposed tagging approach. In particular, we were interested in validating our claim that assisting users with related pictures would result in better and more number of tags for the images. We recruited a total of 35 users from the university campus by sending invitation mails. All participants were students with their age in the range of 19 to 28 with 13 female and 22 male. To avoid the cold start, monetary incentives were provided. The URL of the web based prototype of PhotoSense was mailed to all the participants along with the instructions about the task they need to perform. All the participants were asked to complete two tasks: *Task 1)* Annotate two images of their choice (choose from the set of presented image or upload one

of their own image.) *Task 2)* Annotate them again using the PhotoSense application.

Results

All participants successfully completed both the tasks. Results show that (Refer Figure 8), using PhotoSense, users were able to contribute more number of tags to the image (Mean tag count for task 1: 2.02, while Mean tag count for task 2 is 4.86).

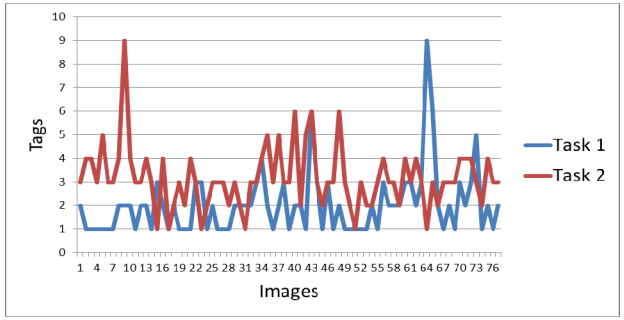

Figure 8: Number of tags per image in task 1 and task 2

We were also interested in knowing the quality of the tags, which we measured in terms of *Tag rank*, i.e. total number of images that correspond to the tags. We found that Mean tag rank for the tags in task 1 was 21058.32, while the mean tag rank for tags in task 2 was 153.35. This decrease in the tag rank provides good evidence about the improved tag quality since less number of images now shares the same set of tags. Another interesting thing to observe was the number of iterations users spent on each image in order to come up with good tags. Results show that, users required only 4 iterations (Mean 3.42) for tagging an image using PhotoSense application. When asked majority of the users (97%) said using PhotoSense was fun. We believe these evidences are good enough to show that with fewer attempts, users were able to contribute more number of quality tags to an image.

CONCLUSION

In this paper, we presented a novel tagging framework along with its prototype design. We believe our approach makes a significant contribution in the way it merges the useful characteristics of "looking for" and "looking at" experiences to improve tagging. Future work includes public release of the application and a comprehensive user study and analysis of PhotoSense.

Citations

[1] Ames, M. and Naaman, M. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of*. CHI '07. ACM, 971-980.

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

[3] O'Connor, B. C., and Griesdorf, H. F. 2007. *Structures of Image Collections: From Chauvet-Pont-d'Arc to Flickr,* Libraries Unlimited.

[4] Santini, S., Gupta, A., and Jain, R. Emergent semantics through interaction in image databases. *IEEE Trans. Knowl. Data Engg.* 13(2001), 337-251.

[5] Sigurbjörnsson, B. and van Zwol, R. Flickr tag recommendation based on collective knowledge. In *Proceeding of the* WWW '08. ACM, 327-336.

[6] Smith, G. 2008, Tagging: People powered Metadata for the social web, New Riders Press.

[7] Von Ahn, L., and Dabbish, L. Labeling images with a computer game. In Proceedings of CHI'04, ACM, 319-326.

[8] Wu, L., Yang, L., Yu, N., and Hua, X. Learning to tag. In *Proceedings of* WWW '09. ACM, 361-370.