This is a preprint of an article accepted for publication in Journal of the American Society for Information Science and Technology. Huang, H., Jörgensen, C. (in press, 2012). Characterizing user tagging and co-occurring metadata in general and specialized metadata collections. *Journal of the American Society for Information Science and Technology*,

Characterizing user tagging and co-occurring metadata in general and specialized metadata collections

Hong Huang

School of Information, University of South Florida, Tampa, Florida, 33620-7800.
Telephone: (813) 974-6361; Fax: (813) 974-6840; E-mail: honghuang@usf.edu

Corinne Jörgensen

School of Library and Information Studies, Florida State University, Tallahassee, Florida, 32306-2100.

Telephone: (850) 644-5775; Fax: (850) 644-6253; E-mail: cjorgensen@fsu.edu

ABSTRACT

This study aims to identify the categorical characteristics and usage patterns of the most popular image tags in Flickr. The "metadata usage ratio" is introduced as a means of assessing the usage of a popular tag as metadata. The authors also compare how popular tags are used as image tags or metadata in the Flickr general collection and the Library of Congress' photostream (LCP), also in Flickr. The Flickr popular tags in the list overall are categorically stable, and the changes that do appear reflect Flickr users' evolving technology-driven cultural experience. The popular tags in Flickr had a high number of generic objects and specific locations related tags and were rarely at the abstract level. Conversely, the popular tags in the LCP describe more in the specific objects and time categories. Flickr users copied the Library of Congress-supplied metadata that related to specific objects or events and standard bibliographic information (e.g., author, format, time references) as popular tags in the LCP. Those popular tags related to generic objects and events showed a high metadata usage ratio, while those related to specific locations and objects showed a low image metadata usage ratio. Popular tags in Flickr taggers appeared less frequently as image metadata when describing specific objects than specific times and locations for historical images in Flickr LCP collections. Understanding how people contribute image tags or image metadata in Flickr helps determine what users need to describe and query images and could help improve image browsing and retrieval.

KEYWORDS

Social tagging, image tagging, image retrieval, power law, collaborative tagging, image metadata, image indexing

Introduction

With the availability of shared digital collections through the Internet, social tagging systems have grown rapidly in recent years, and tags contributed by users can be used to describe many different materials in a variety of formats. A tag is a term or phrase associated with an item and contributed personally in an open social setting by a tagger (who is not necessarily the creator). Tagging has become a popular activity and is producing a large amount of additional descriptive terminology for items on the web. Users utilize collaborative tagging for a variety of interactive online activities (Macgregor & McCulloch, 2006). These Popular Tags are sometimes regarded as "power tags," that reflect the "collective intelligence" about the content in the annotated resources; therefore, they are viewed as potentially enhancing information retrieval (Peters & Stock, 2010). These Popular Tags are often presented to users as Tag Clouds within web collections and have come to serve as a representational and navigational tool for finding people or resources within groups sharing similar interests, as well as finding items. Research has suggested that tags being used in a collaborative environment might form a "nascent consensus" among the participants with a predictable and stable tagging pattern reflecting their socio-cultural engagement (Golder and Huberman, 2006). While users of collaborative tagging systems (e.g., Flickr) exhibit much variety overall in the sets of tags they employ, the frequencies of the top ranking tags ("Popular Tags") used as tags or metadata may indicate trends in the current aggregated tagging behaviors of individual users. Prior research has suggested the general patterns of occurrence of all-time popular tags, and these popular tags are frequently displayed in Tag Clouds as valid representations of content. Our concern here is that these Tag Clouds are based on simple specific word occurrence frequencies rather than more sophisticated categorical analyses or statistical measures. This study takes a step towards understanding the nature of these

popular terms and how they might fit into an overall model of document representation through the additional information provided by categorical analysis as well as users' changing interests and social tagging trends.

Tagging as a phenomenon is no longer unusual and has engaged a larger community of contributors than previously possible in providing additional descriptions for items in collections. It can aid in the discovery of more information about an item by empowering users to contribute what they may know about an item, yet can also introduce spurious or erroneous information about that same item. Many unanswered questions concerning tagging remain, questions surrounding the types of tags contributed and the utility and authority of user contributed tags, and tagging activities and systems have become an area of high interest to researchers recently. The research reported herein contributes to a growing body of knowledge and understanding of these systems by pursuing research questions in a large resource which tends to attract active tagging, the Flickr database, a popular open tagging photo-sharing system.

The Flickr General Collection (hereafter referred to as the FGC) refers to all the photos uploaded collectively in the Flickr image repository. The FGC includes various broad genres of images: people visiting places, their daily life activities, their social activities and their interests and hobbies. The FGC is a very large collection, having accumulated over 6 billion images since it was launched in 2004.

In 2008, the Library of Congress Prints & Photographs division began a collaboration with Flickr by placing two history and heritage related images collections in Flickr, the Library of Congress' photostream (LCP). The goal of this collaboration is to gain additional knowledge about and metadata for the photographs by allowing Flickr's users to offer comments, provide new tags, make corrections, and add precise information related to actual locations, individuals,

and events contained in the pictures (Springer et al., 2010). The Flickr LCP collection contained 16,272 historical images provided by the Library of Congress in 2010. The Library of Congress also provided metadata and two machine recognizable tags for each uploaded Library of Congress photo in Flickr. By analyzing the tags associated with an LCP image, researchers can discover if end users tag differently for the historical, heritage, or archival images as opposed to the tagging of more general shared photos posted to Flickr, since the available metadata in museums or cultural heritage collections can be too specialized for the general public (Srinivasan et al., 2009). Additional questions that can be explored are whether the addition of more specialized metadata to an image influences tagging behavior of other users and what kinds of terms users add when they are not already present as metadata, which may suggest users perceive these as useful terms which are lacking. Much of the prior research has analyzed the general patterns of the all-time popular tags, but fewer discuss whether or how the terms may change or adapt over time. By comparing the Flickr all time popular tags in two time periods, from the first launch of the Flickr database (Feb, 2004) to 10/01/2006 (providing a baseline up until the LCP images were added), and from Feb, 2004 forward to 10/01/2010, several research questions are explored.

The first question addresses the nature of the overall FGC:

1) *Are the most popular tags used in Flickr changing over time, and if so, how?*

This question is investigated by comparing the popular tag frequency rankings and tag categories in the entire FGC in the years 2006 and 2010 respectively. The second and third questions investigate whether there are differences in tagging in a general collection and a subset of that collection which is governed by a specific collection program.

2) *Are there differences in end-user tagging behavior between users of a non-domain specific collection (the FGC) and users of a subset of images comprising a specific collection (the LCP)?*

This question is investigated by comparing the popular tag frequency rankings and tag categories between the FGC and the LCP collection at one point in time (which may serve as a future baseline).

3) *Does the presence of image metadata influence tagging behavior in a subset of the FGC, specifically the LCP historical images collection?*

This question is investigated by comparing the frequency of popular tags and co-occurring metadata between the Flickr general collection and the LCP collection, and suggesting a measure, the "image metadata usage ratio" to measure this influence.

The popular tags lists (all time most popular tags) in Flickr (http://www.flickr.com/photos/tags/) and the Library of Congress' photostream (LCP) in Flickr (http://www.flickr.com/photos/library_of_congress/tags/) can be harvested to analyze the frequencies and overlap of terms used as image tags or metadata. Answers to these questions may help researchers better understand end-user needs and preferences for metadata, and enhance the quality of existing controlled vocabularies or generate new ones.

Definitions

The following are the working definitions adopted for use in this specific research.

Tag*:* any term or phrase in a Flickr record that appears in the "Tag" category.

Popular Tag*:* the most popular tags used as tags in Flickr. These are the all-time most popular tags listed as tag clouds in Flickr (http://www.flickr.com/photos/tags/) and the Library of Congress' photostream (LCP) (http://www.flickr.com/photos/library_of_congress/tags/).

Metadata*:* data about a resource that has been structured or displayed for some kind of use (this is more fully operationalized in the methods section).

Tag Category*:* a tag–category matrix used to categorize tags at different levels of meaning (this is more fully operationalized in the methods section).

Literature Review

The collaborative tagging environment in Flickr allows users to organize their digital content and to share their collections with others. Users can contribute a set of terms as tags and/or provide a title, descriptions, notes, or comments as unstructured metadata for photos to describe both the concept and the content of the photos that they share (Golder & Huberman, 2006). Analyzing the tagging patterns in a specific community could identify the terminologies and conceptual structures that users select for images (Furnas et al., 2006; Stvilia & Jörgensen, 2009). Compared to traditional indexing and classification schemes created by library professionals, non-expert generated tags in social tagging systems (e.g., Flickr) have unclear hierarchies or classification schemes and are heterogeneous in semantic meanings; these are often referred to as "folksonomies" (Golder & Huberman, 2006). User-supplied tags in the collaborative environment can represent rich image content shared by diverse user groups that can eventually be tested as indexing terms, providing added value to the image (Jörgensen, 1998; Stvilia & Jörgensen, 2010).

The distributions of image tags in different categories can also reveal the types of terms users choose for tagging activities. An early theory of structured descriptions of images was proposed by Panofsky (1955), who used three classes of meaning for visual art images: "pre-iconographical description," "iconographical analysis," and the "iconological interpretation." Shatford (1986) further regarded the "pre-iconographical description level" as the "generic" level of description; "iconographical" as "specific," and "iconological" as "abstract." She also added four additional facets to each level: who, what, where, and when, creating a three by four category matrix for image description. Shatford's classification has been applied to Flickr tag analysis and to users' queries (Choi & Rasmussen, 2002; Yoon, 2008; Yoon, 2009), as well as to image archiving and indexing (Armitage & Enser, 1996; Conduit & Rafferty, 2007). Yoon (2008) also classified image tags using three additional attributes: color, denotation, and connotation. Rorissa (2010) compared Flickr terms with image index terms traditionally employed by libraries, using an empirically established ten-level conceptual framework (Jaimes & Chang, 2000) for layering the syntactic and semantic image attributes. The ten-level conceptual framework helps users make inferences to understand local or global scenic content of the image (Jörgensen, 1998, Jörgensen et al. 2001). Flickr terms were also used to test and develop a new semantic description model for images (Lee & Neal, 2010).

Image metadata can be classified according to its different purposes: descriptive, administrative, and structural (Caplan, 2003; Zeng & Qian, 2008). Unlike administrative metadata which focuses on managing resources or structured metadata useful for comparing resources, descriptive metadata helps users store, organize, access, and retrieve items (Stvilia & Jörgensen, 2010). McCay-Peet & Toms (2009) indicated that both descriptive and conceptual image attributes that could be used as metadata were widely used to locate an image. Cunningham &

Masoodian (2007) suggested that descriptive image metadata were helpful to improve browsing or searching for specific information needs. Jeong (2009) compared tags and other metadata (e.g., title and description) on YouTube, and found that more than 50% of terms were shared among tags and metadata fields. Exploring how users apply Popular Tags as tags (Edmunson-Morton, 2010; Vaughan, 2010) or metadata to different image objects (e.g., historical or non-historical) could help develop terminology for image description for both non-historical and historical image archives.

Methods

To obtain our test collection of Popular Tags, the researchers first obtained the lists of the all-time most popular tags (about 150 tags) in Flickr (http://www.flickr.com/photos/tags/) on two dates (10/01/06 and 10/01/10), and in the Library of Congress' photostream (http://www.flickr.com/photos/library_of_congress/tags/) on 10/01/10. The frequencies of these popular tags used as tags or metadata were retrieved by searching as "Tags only" or "Full text" from the Flickr "Advanced Search" page. In addition to tags, image metadata in this research refers to the title, descriptions, and additional notes associated with the image, whether contributed by the LC (structured metadata) or by end users (unstructured metadata). When searching a term by "Tags only," Flickr searches only the photos with that term denoted as a tag by users, and returns the number of hits for the specific term used as a tag. So, the frequency of usage of a term as metadata (e.g.: in the title and/or the description for the image page, but not included in the tag list for that photo) can be computed by using the search results of "Full text" minus "Tags only."

A web crawler program (a .NET and Visual Basic application) connected to the Flickr API was used to harvest the usage frequencies of the most popular tags as "Tags only" and "Full

text" searches by the two time periods from the first launch of the Flickr database (Feb, 2004) to 10/01/06 and then up to and including 10/01/10 in the Flickr database. The research results of the most popular tags in the Library of Congress' photostream were also searched by the date before 10/01/10 (the LC photostream was added to Flickr as a pilot project in 2008). In order to compare category distributions of the popular terms used as image tags, the 150 most frequently occurring tags ("Popular Terms") were further categorized by the researchers using the groups in the Shatford model (Shatford, 1986; Choi & Rasmussen, 2003; Chung & Yoon, 2009) that make up the three category (Generic, Specific, and Abstract) by four facet (Who, What, Where, and When) matrix. The researchers added the category "LC-supplied" to identify LCP terms that were provided by the Library of Congress and adopted by users as tags. An additional category, "Image-related" tags (I), denotes tags describing the production and form of the image. In order to describe the tendency of a particular term to be used as metadata, we defined a ratio to indicate the metadata usage as follows:

For a given tag,

$$Metadata\_usage\_ratio = \frac{\#of\_times\_used\_as\_image\_metadata}{\#of\_times\_used\_as\_image\_tag}$$

The ratio indicating a term used as metadata versus the same term used as a tag is computed by using the frequencies of a particular term as metadata (searching results of "Full text" minus "Tags only") divided by its frequencies as image tags (searching result of "Tags only") in Flickr. The smaller the value, the less likelihood the tag will be used as metadata. Finally, descriptive statistics, Fisher's Exact Test, chi-square tests, Mann-Whitney tests, and the Kruskal-Wallis test were used to analyze the data using SPSS software (Chicago, IL, USA).

Findings

RQ1: Are the most popular tags used in Flickr changing over time, and if so, how?

To answer this question, a comparison of Popular Tags in 2006 and in 2010 indicates that the most Popular Tags are stable, with approximately 87% remaining constant in the top 150 popular tags list. An analysis of the classification of Flickr Popular Tags using the Shatford model (Table 1) shows that tags relating to a specific location (Loc-S) were most frequent with a slight drop from 32% in 2006 to 27% in 2010. This is followed by generic objects (Obj-G), which increased from 21% in 2006 to 26% in 2010, and generic event/activity (Ev-G) related tags, which remained at 12%. Generic place (Loc-G) tags decreased from 9% in 2006 to 7% in 2010; however, there was an increase from 6% in 2006 to 8% in 2010 in image-related (I) tags. The contingency 2x2 tables were created and tested using Fisher's exact tests in table 1. None of these changes were statistically significant, except for the specific time ($p < 0.05$) tags (T-S) which decreased from 6% to 1%.

An examination of the terms that dropped off or were added to the popular tags list between 2006 and 2010 (Table 2) suggests that users may be adding slightly more emotive or Abstract-level terms (e.g., "fashion," "love," "old", "vintage") in the organization of their shared photos. Other additions were tags describing new technologies and devices for producing images (e.g., "iphone"). Flickr's growth in the international community is shown through the addition of non-English prefixes to the Popular Tag words (e.g., "la", "de"). In addition to sharing images of popular outdoor activities (e.g., "camping," "roadtrip," both popular in 2006), users document a wider variety of activities or social events (e.g., "band," "model," "dance" in 2010).

In Flickr, some format-related terms were widely used as image tags: “macro,” “blackandwhite,” “portrait,” “cameraphone,” and “Nikon” (Table 1). Users choose terms that describe their or others activities. For example, “graffiti” remained a Popular Tag used to record the street art Flickr users encounter in daily life or on longer trips. In addition, the popularity of “me,” “myself,” “portrait,” and “new” as tags shows that tagging in Flickr is still used largely for personal identification and self-expression. Location and time terms are the major facets used in Flickr (56% of the Popular Tags in 2006 and 44% in 2010). As opposed to visual content tags (e.g., color) describing the objects in the images, location and time tags provided a nonvisual (e.g. bibliographic) component of the image, such as where or when it was created. For instance, the term “geotagged” was used to label all the images taken from a specific location within an interactive geographical map. As cameras themselves are adding more metadata to the image, such as time and location, that may account for the overall decrease in the time and location terms as tags. Furthermore, as one may expect, there are multiple meanings for a single word, but when words are associated with images in Flickr they may take on a more consistent meaning. For example, in most cases the tag “live” signifies something that is broadcast as it is happening rather than the verb form of the word.

RQ2: Are there differences in end-user tagging behavior between users of a non-domain specific collection (the FGC) and users of a subset of images comprising a specific collection (the LCP)?

Within Flickr, users are free to add up to 75 tags for each photo in the LCP collection. The pattern of these tags demonstrates context-dependent tagging behaviors in a social tagging environment, where tags are duplicated from LCP-supplied metadata; other user-generated tags

could provide metadata category recommendations for the Library of Congress's image management and archiving processes. Historical images may also stimulate taggers to specify aspects such as person/thing, location, and historical periods. Of the LCP Popular Tags, only 23% co-occur in Flickr general collection. The top three categories of the Popular Tags in the LCP were Obj-G (34%), Loc-S (15%), and Image related (I, at 8%). There were no Obj-S related popular tags in the FGC (Table 1). Fisher's exact tests showed that LCP had statistically significant differences in Obj-S ($p < 0.01$), T-S ($p < 0.01$), and Loc-S ($p < 0.05$) in comparison to the popular tags in Flickr general collection, and LCP had no T-G related tags. In addition, users used the metadata that library professionals provided (LC supplied: 9%) to tag the historical images (Table 1). Most of the LC-supplied Popular Tags describe specific objects or events (Table 3). LC-supplied tags were used to describe the sources of the image (e.g., "bainnewsservice") or individual names of the image creators (e.g., "howardhollem"). LCP popular tags describe people and social status (e.g., "soldier, " "royalty"), events such as wars and associated objects (e.g., "aircraft," "battleship," "war," "worldwar2,"), sports (e.g., "baseball," "al" for American League, "pologrounds"), and specific events (e.g., "iwd" for International Women's Day), outfits (e.g., "cap", "hats", "suit", "uniform"), fashion (e.g., "moustache"), picture formats (e.g., "4X5," "LF: large format," "glassnegative"), and techniques (e.g., "colorized," "photochrom"). In comparison to the Flickr general collection, the LCP collections showed a higher proportion of *Who*-related tags ($\chi^2 = 9.93$, $df = 1$, $p < 0.05$) but a lower proportion of *Where* tags ($\chi^2 = 10.08$, $df = 1$, $p < 0.05$) (Figure 2). This indicates that LCP popular tags mostly describe people and things, rather than the locations that were seen in the Flickr 2006 and 2010 collections. Location information is one of the often missing pieces of information that the LC is definitely interested in adding to many of its historical photos.

RQ3: Does the presence of image metadata influence tagging behavior in a subset of the FGC, specifically the LCP historical images collection?

An image in Flickr contains non-textual information (the picture itself), a tag or tags associated with the concept(s) it represents, and whatever textual description for the picture may have been provided. Investigating the distribution of the terms used as image tags and as metadata (defined for this research as all fields of description included in Advanced Search except "Tag") identifies the metadata use patterns for image description in the collaborative tagging environment. In Figure 3, the frequencies of popular Flickr terms used as image tags are ranked, and they show the predicted long tailed distribution (Tonkin, 2006; Huang, 2006). However, the frequencies for some of the terms (e.g., "me", "day," "park") used as tags and also as image metadata are exceptionally high (Figure 3). The image tags' metadata usage ratio of Flickr terms ranged between 0 and 18.16 while the ratio of LCP collections ranged between 0 and 23.18. These results seem to indicate that users apply terms in the Popular Tags list differently when used as an image tag versus when used as image metadata. The metadata usage ratios for some specific related terms such as "Christmas" (Flickr2010:0.17, LCP:0.08) , "newyorkgiants" (Flickr 2010: 0, LCP: 0), and "newyorkcity" (Flickr 2010: 0, LCP: 0) were small and close to zero, indicating that these terms appear mostly as image tags. However, the ratios for some generic related terms like "city" (Flickr 2010: 2.09, LCP: 4.81) and "house" (Flickr 2010: 3.13, LCP: 3.13) are much greater than zero which indicates that these popular terms are widely used both as tags and as image metadata. The rankings of the metadata usage ratios show that Obj-G terms have higher rankings while Loc-S terms have low rankings in both Flickr 2010 (Table 4) and LCP collections (Table 5). The rankings for the ratios in Obj-S terms

are also low in LCP collections (Table 5). These results suggest that users overall prefer terms at the generic, or "basic object" level (Rosch and Mervis, 1975), contributing these terms as tags themselves or drawing them from supplied metadata.

The Kruskal-Wallis test of dependence of the *Generic*, *Specific,* and *Abstract* term categories for the frequency ratios of image metadata usages in LCP and FGC showed that the *Generic* term ratios difference between FGC and LCP was statistically significant ($\chi^2 = 8.25$, $p = 0.004$). *Specific* and *Abstract* terms, however, were not significantly different. In addition, the Kruskal-Wallis test of dependence of the *Who*, *What*, *Where*, and *When* terms for the frequency ratios in LCP and Flickr showed *Who* ($\chi^2 = 8.64$, $p = 0.003$) term ratios were statistically different (Table 6) while *Where, What* and *When* term ratios were not significantly different.

Figures 4 and 5 show the arithmetic average of the number of times Popular Tags were used both as image tags and metadata in both the 2010 FGC and LCP collections based on the classification in Table 1. Only those categories that were greater than 5% of total Popular Tags are shown in both Figures 4 and 5. All metadata usage ratios in LCP are close to or below one. However, T-G, Obj-G, and I-related metadata usage ratios in the FGC are close to two. Interestingly, the Ev-G related metadata usage ratio is low (<1) in the FGC but high (>1) in LCP. Both the FGC and LCP show a low metadata usage ratio for specific location related (Loc-S) tags. The same situation was observed for the LC-provided, specific object (Obj-S), and specific time (T-S) related tags in the LCP collection. Other generic tags (Obj-G, Ev-G in LCP; Obj-G, Loc-G, T-G in Flickr 2010) as well as Image related tags (I) showed a high image metadata usage ratio.

A Mann-Whitney *U* test indicated a significant difference in image metadata usage ratios between *Generic* and *Specific* terms (n = 107, z = 6.73, $p < 0.01$) in Flickr 2010. The ratios in

Generic terms are higher (*M* rank = 71.57, n = 66) than *Specific* ones (*M* rank = 29.48, n = 41). Similar patterns were also observed in LCP 2010 collection, the image metadata usage ratios between *Generic* and *Specific* terms is significantly different (n = 107, z = 4.40, $p < 0.01$), and *Generic* terms show a higher usage ratio as metadata (*M* rank = 68.88, n = 60) than *Specific* ones (*M* rank = 41.49, n = 47).

Mann-Whitney U tests also found that the Obj-S category was significantly different in image metadata usage ratios than in the Loc-S (n = 32, z = 2.62, $p < 0.01$) and T-S (n = 22, z = 2.89, $p < 0.01$) categories in LCP 2010. Obj-S related terms have a lower metadata usage ratio (*M* rank = 10.59, n = 11) than Loc-S ones (*M* rank = 19.60, n =21); and also a lower metadata usage ratio (*M* rank = 7.77, n =11) than T-S ones (*M* rank = 15.88, n =11).

Discussion

The first research question asked if the usage of Popular Tags had changed over time, and if so, how. The results indicate that the most popular tags did not vary much during the time period observed. Users typically begin with their own tags but eventually these user-supplied tags converge on a consensus set of popular tags. These stable tags could be used to improve image indexes or access points for efficient information retrieval by subdividing large collections (Jörgensen 2007; Ransom & Rafferty, 2011). Alternatively, these very popular tags may only have meaning at the level of the individual contributing them. The fact that the tag “new” is popular (Table 1) suggests the Flickr database is a growing, evolving, and dynamic image tagging system. As other studies have indicated, the majority of the Popular Tags in Flickr are related to casual activities, travels, and social events. Some of the change in popular tags represents a technology-driven consumer market evolution, as with the change from

"cameraphone" to "iphone." These temporal influences for some tags indicate that tagging can be an inexpensive way of enhancing terminology in a period of rapid change, especially with technology or consumer products, allowing systems to adapt quickly to vocabulary changes.

The analysis of Popular Tags in the FGC indicates that generic and specific terms are more frequently used than abstract ones (Figure 1), which supports similar findings from other studies (Hollink et al., 2004; Ransom & Rafferty, 2011). Some researchers have used Wordnet to cluster and map tags to provide richer semantics for them in an automated way (Ungrangsi and Anutariya, 2011). Lower usage of Popular Tags to describe abstract objects and scenes indicates that users describing images tend to be more concerned with describing specific image objects and scenes rather than their abstract meaning (Jörgensen, 1998), but other types of tasks such as sorting or searching bring out other abstract or emotive terms. Grammatical analysis in the FGC Popular Tags reveals that image descriptions are primarily nouns or adjectives with fewer verbs as tags describing an action or task (Huang, 2006). As Flickr tags are largely related to the extant visual content of the images, the grammatical property of a tag in relation to the perceptual image content must be carefully evaluated to disambiguate meaning.

Analysis of Popular Tags in Flickr shows a rich set of terms for generic visual properties. Some are related to global image properties (Jörgensen, 1998). For example, color is often referred to with basic terms such as "red" or "green;" texture is indicated by terms like "rock;" shape is described with terms like "flower;" and spatial information is given using terms like "sky" or "ocean." Color, texture, shape, and spatial information are important features in development of Content-Based Image Retrieval System (CBIRS). The traditional text-based approach, which is a high-level, user-centered approach, begins with a conceptual interpretation of the perceptual content of the image/object and moves toward a more generic terminology

(Jörgensen, 1998). The CBIRS digitizes the image/object and employs basic color processing algorithms to enable identification of image syntax and object geometry (Smeulders et al., 2000) and to use the results to identify similar patterns in other images. Thus Popular Tags in collaborative tagging systems that describe visual image properties as those in CBIRS processing enable leveraging of machine-based learning through user decriptions in image retrieval to the increase the speed and accuracy of these methods (Datta, Joshi, Li, & Wang, 2008). Flickr users use conceptually-based textual terms referring to non-textual digital content as tags to improve their browsing and searching capability, demonstrating the integration of human perceptual experience of pre-semantic awareness, visual configuration, and structural processing (Enser, 2008).

The second research question investigated whether there are differences in end-user tagging behavior between users of a non-domain specific collection (the FGC) and users of a subset of images comprising a specific collection (the LCP). The users of the FGC use it both for storing and/or sharing their own photographs and also for group activities surrounding image subsets that have a specific focus or a specific subject (Stvilia and Jörgensen, 2010). Those who tag photographs in the LCP may be assumed to share some similar motivations toward benefitting a larger community with their tags. Flickr users thus provide new tags that potentially enrich image descriptions for the Library of Congress collection.

Popular Tags in the LCP are very different from those in the general collection. For instance, Flickr LCP users show specific interests in fashions (e.g., mustache) and activities (e.g., "boxing," "baseball") that popularly occurred in the past. The current research demonstrates that a higher proportion of Popular Tags were used to describe historical people or subjects and that more specific and detailed image tags were required to describe historical heritage collections.

The emerging multifaceted, socially-generated tagging system in Flickr shares similar social concepts and describes objects and features with a similar level of complexity as is available in library cataloging systems or metadata management schemes (e.g.: Museum management systems), suggesting that a community-based system can to some extent engage in similar image description approaches as those of professionally-developed general cataloging systems. Most of the Popular Tags used in LCP are user-generated, providing a good resource for image tagging at the generic level in other available historical image archive collections.

However, image searchers also have needs for or create requests describing specific objects or events (Armitage & Enser, 1997; Markkula & Sormunen, 2000; Chen, 2001). This research demonstrated that Flickr taggers also copy Library of Congress supplied metadata and use these terms as tags to describe specific historical objects or events. As has been found in other research (Stvilia, Jörgensen, & Wu, 2012) this most likely occurs because users do not have sufficient knowledge to identify the image sources, authors, or specific background information for historical images. Over 90% of the Popular Tags are user provided, with the potential for greatly enriching the tagging strategies for Museum collections (Trant, 2006). In addition, the current research found that untrained taggers (e.g., Flickr users) use Popular Tags differently when annotating FGC images than when annotating cultural and heritage collections. Understanding the tagging behaviors by analyzing the usage patterns for popular tags in the LCP collection in Flickr is beneficial for the library and information science field in understanding how public visitors access collections, in designing better systems for browsing and querying, and in eventually helping to create personal digital collections (Marty, 2011).

The third research question asked if the presence of image metadata influences tagging behavior in a subset of the FGC, specifically the LCP historical images collection. The metadata

usage ratio is proposed as a metric to characterize the relationship of a tag's semantic category and corresponding likelihood of metadata adoption. The main finding is that Popular Tags (e.g. nouns and adjectives with a generic linguistic meaning) are used more frequently as Flickr metadata than tags with specific meanings (e.g. specific objects, locations, time, or events). This finding has implications for facilitating browsing in these systems, as using image indexes to capture the semantic content of an image is difficult and browsing becomes an important functionality in image retrieval (Enser, 2008). Browsing terms normally carry a general meaning and allow users to peruse a broad range of images while adapting their information seeking process to current needs and narrowing their search terms, as effective querying terms are more specific and return smaller sets with more accurate results. The image metadata usage ratio between the frequencies of image terms used as tags and image terms used as metadata could assist image annotators in considering more generic terms for browsing and then choosing more specific tags and metadata for searching (Enser, 2008). The differences in image metadata usages among just the top 150 Popular Tags could inform the creation of a thesaurus for browsing in addition to one for querying, enabling improved retrieval systems with better precision or recall.

Secondly, Jörgensen (1998) indicates that major image indexing systems (e.g. the *Library of Congress Thesaurus for Graphic Materials* [LCGTM] and the *Dewey Decimal Classification* (DDC)) do not contain any terms for specification of an element's location in the internal structure of an image (e.g., "behind," "next to," "on" or "against"). Interestingly, in Flickr, the terms used for specific location of objects in the internal structure appear mostly in image descriptions (image metadata) as opposed to image tags, likely because these descriptions can be longer; the image metadata usage ratios for these terms (e.g., "behind": 72.63, "on": 538.60) are exceptionally higher than other terms. In other words, users normally tag the image using terms

representing global location and use specific location-related terms as metadata to describe the internal structure of the image. These words are usually considered stop words in searching and are not normally included in indexing (or in tools such as Wordnet), yet they can reveal important differences in composition or clues to object identification in CBIRS.

Lastly, image tag metadata usages are different when comparing metadata for the FGC and the LCP historical images. When describing historical photos in image archives, users choose terms as tags or use in them in metadata when describing persons and things (*Who*), or generic objects, events, actions, and times (*Generic*) differently in comparison to their annotations of other kinds of images. For example, users might show more interests in using image tags to describe the specific persons and things (*Who*) in LCP historical images than in the FGC's, and users could select more generic event (Ev-G) related terms as image metadata to describe the event/activities reflected in LCP historical images than in the FGC's.

For historical images in the LCP collections, Flickr taggers used Popular Tags less as image metadata when describing specific objects than specific times and locations which could appear both as image tags and other types of image metadata. Users might also prefer using more specific objects/persons related metadata (e.g., bibliographical information supplied with the image: title, creator) to tag the historical images. These specific objects/person related tags were partially Library of Congress provided or may have appeared directly on the image in the margins, thus were not part of the non-textual content of the image. When these terms are added as tags they sometimes provide specific bibliographical information (e.g., creator, rights) missing in tagging. Or in specialized subsets of the FGC or other specialized collections, there may not be enough good quality alternative tags available to the users to semantically describe the content

of a historical image, as users may lack sufficient background knowledge to provide their own tags to accurately describe the historical contexts for image objects.

Limitations

This study is not without several limitations. The Popular Tags subset represents the most frequent but also a very limited portion of all the tags in Flickr. Further research needs to be conducted to explore other popular tag clouds and larger samples in Flickr (e.g., users' groups for holiday photos) to compare their category types and metadata usage ratios. The Flickr database has several other confounding factors: compound terms ("de," "la") were included that did not have conceptual meaning, as were synonyms ("photo," "photography") for the same concept, and the system concatenates multiword phrases, making categorization difficult. The LCP collection also has a limited number of photo sets available to Flickr users, therefore these research results cannot be generalized beyond this specific collection. However, it is useful to discover differences in a specialized image subset such as this as these differences may enlighten future indexing, tagging, and retrieval of these subsets. With the expansion of the number of photos in LCP depicting rich, diverse cultural and historical objects and events, the collection is worthy of revisiting as it grows to further analyze tagging patterns and metadata usage ratios for improving image archiving or digital image collection management in libraries. In LCP, some of the Popular Tags were not user supplied (e.g., bainnewsservice) but rather were copied from the Library of Congress metadata. Further investigation is needed to explore the connections in Popular Tag usage patterns between user-generated tags and expert-provided tags in LCP and in more generalized collections.

Conclusion

This research provides an overview of tagging behaviors in relation to the most Popular Tags in a collaborative environment. While there are limitations, the results provide much opportunity for future study. For instance, the LCP can be studied over a period of time to see if "curated" collections undergo similar changes noted here for the FGC. Popular Tags reflect the culture of the tagging population and also serve as evidence of the social impact of collaborative social tagging activities. It is also well-known that the discriminatory power of terms in a language is not necessarily found in the most popular, or the least frequently occurring, words, but rather in a middle range of the Zipf distribution (Losee, 2001; Jörgensen, 2007). Future study will include studying a sample of these "mid-range" terms and their relationships to tags and to metadata.

Understanding social tagging patterns in a community-based system defines the common vocabularies of the community and can enhance the representational predictability of image description and retrieval in these systems. Finding popular tag distributions in Flickr can potentially reveal unique social tagging distributions while the culturally accepted understandings of the corresponding tags may enable the building of more efficient image retrieval systems integrated with content-based image retrieval techniques in Flickr.

The proposed image metadata usage ratio metric characterizes a term used as a tag or metadata in user-generated metadata collections, and while the mechanisms and rationale behind why end users copy metadata as tags for LCP images is still unknown, this metric could guide formal metadata creation, or distinguish among different subsets within a collection (e.g, the FGC or the Flickr LCP collection) for browsing. It may also assist in characterizing tags semantically based on potentially differing metadata usage ratio scores. Further analysis using the proposed metadata usage ratio between tags in the FGC and the LCP may help identify

general and specific needs for browsing and searching historical image collections and thus begin to bridge the semantic gap between the language used by the end user and the library/archives professional; librarians could provide additional types of image metadata based on the metadata usage ratios of categories of terms.

Collaborative tagging could also help develop interactive, efficient cross language museum collection curation. Future work could improve the tag browsing experience by applying similar methods in constructing tag clouds, e.g., by presenting tags while considering term popularity and user interests of specific groups of people. Social tagging and folksonomy have a strong potential for connecting cultural institutions with the individuals and subcultures of those who live in a particular society, and enabling a two way process of communication and understanding among groups of individuals and institutions for knowledge creation.

References

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Augmenting navigation for collaborative tagging with emergent semantics. In *Proceedings of the 5th International Semantic Web Conference: ISWC2006* (pp. 58-71). doi:10.1007/11926078_5

Armitage, L., & Enser, P. (1997). Analysis of user need in image archives. *Journal of Information Science*, 23(4), 287–299.

Bearman, D., & Trant, J. (2005). Social terminology enhancement through vernacular engagement: Exploring collaborative annotation to encourage interaction with museum collections. *D-Lib Magazine*, 11(9). Retrieved from http://www.dlib.org/dlib/september05/bearman/09bearman.html

Caplan, P. (2003). *Metadata fundamentals for all librarians*. Chicago, IL: American Library Association.

Chen, H. (2001). An analysis of image queries in the field of art history. *Journal of the American Society for Information Science and Technology,* 52(3), 260–273.

Choi, Y., & Rasmussen, E. (2003). Searching for images: The analysis of users' queries for image retrieval in American history. *Journal of the American Society for Information Science and Technology*, 54, 6, 498–511. doi:10.1002/asi.10237

Chung, E., & Yoon, J. (2009). Categorical and specificity differences between user-supplied tags and search query terms for images: An analysis of Flickr tags and Web image search queries. *Information Research,* 14(3). Retrieved from http://informationr.net/ir/14-3/paper408.html

Conduit, N., & Rafferty, P. (2007). Constructing an image indexing template for The Children's Society: Users' queries and archivists' practice. *Journal of Documentation*, 63(6), 898–919. doi:10.1108/00220410710836411

Cunningham, S., & Masoodian, M. (2006). Looking for a picture: An analysis of everyday image information searching. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 198-199). doi:10.1145/1141753.1141797

Datta, R., Joshi, D., Li, J., & Wang, J.Z. (2008) Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys, 40(2), 1-60, 2008.

Ding, Y., Jacob, E.K., Fried, M., Toma, I., Yan, E., Foo, S., & Milojevic, S. (2010). Upper tag ontology for integrating social tagging data. *Journal of the American Society for Information Science and Technology,* 61(3), 505–521.

Edmunson-Morton, T. (2010). Talking and tagging: Using CONTENTdm and Flickr in the Oregon State University Archives. *The Interactive Archivist: Case Studies in Utilizing Web 2.0 to Improve the Archival Experience*. Retrieved from http://www.lib.byu.edu/sites/interactivearchivist/case-studies/flickr-at-osu/

Egghe, L., & Rousseau, R. (1997). Duality in information retrieval and the hypergeometric distribution. *Journal of Documentation*, 53(5), 488-496. doi:10.1108/EUM0000000007208

Enser, P. (2008). The evolution of visual information retrieval. *Journal of Information Science*, 34 (4), 531–546.

Golder, S., & Huberman, B. (2006). User patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. doi:10.1177/0165551506062337

Huang, H. (2006). Tag distribution analysis using the power law to evaluate social tagging systems: A case study in the Flickr database. Poster presented at the 17th ASIS&T SIG/CR Classification Research Workshop, Austin, TX. Retrieved from https://digital.lib.washington.edu/ojs/index.php/acro/article/viewFile/12497/10996

Hollink, L., Schreiber, A.Th., Wielinga, B.J., & Worring, M. (2004). Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61, 601–626.

Jacob, E.K. (2004). Classification and categorization: A difference that makes a difference. *Library Trends*, *52*(3), 515-540. Retrieved from http://hdl.handle.net/2142/1686

Jaimes, A., & Chang, S.F. (2000). A conceptual framework for indexing visual information at multiple levels. IS&T/SPIE Internet Imaging. San José, CA: IS&T/SPIE.

Jeong, W. (2009). Is tagging effective? – Overlapping ratios with other metadata fields. *International Conference on Dublin Core and Metadata Applications*, 31-39, Seoul, South Korea.Jörgensen, C. (1998). Attributes of images in describing tasks. *Information Processing & Management*, *34*(2-3), 161-174. doi:10.1016/S0306-4573(97)00077-0

Jörgensen, C. Jaimes, A., Benitez, A. B., & Chang, S.F. (2001). A conceptual framework and empirical research for classifying visual descriptors. *Journal of the American Society for Information Science & Technology*, 52(11), 938–947.

Jörgensen, C. (2007). Image access, the semantic gap, and social tagging as a paradigm shift. Retrieved from http://dlist.sir.arizona.edu/2064/01/Jorgensen_Update_SIG-CR_2007.doc

Jörgensen, C., & Jörgensen, P. (2005). Image querying by image professionals. *Journal of the American Society of Information Science and Technology*, *56*(12), 1346-1359. doi:10.1002/asi.20229

Lee, H., & Neal, D. (2010). A new model for semantic photograph description combining basic levels and user-assigned descriptors. *Journal of Information Science, 36*(5)*,* 547-565. doi:10.1177/0165551510374930

Losee, R.M. (2001). Term dependence: A basis for Luhn and Zipf models. *Journal of the American Society of Information Science and Technology*, *52*(12), 1019-1025.

Macgregor, G., & McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Online Library Review, 55*(5), 291-300. doi:10.1108/00242530610667558

Markkula, M., & Sormunen, E. (2000). End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval,* 1(4), 259–285.

Marty, P.L. (2011). My lost museum: User expectations and motivations for creating personal digital collections on museum websites. *Library & Information Science Research*, 33(3), 211-219. doi:10.1016/j.lisr.2010.11.003

McCay-Peet, L., & Toms, E. (2009). Image use within the work task model: Images as information and illustration. *Journal of the American Society for Information Science and Technology, 60*(12), 2416-2429. doi:10.1002/asi.21202

Panofsky, E. (1955). *Meaning in the Visual Arts: Papers in and on Art History*. Garden City, NY: Doubleday.

Peters, I., & Stock W.H. (2010). "Power tags" in information retrieval. *Library Hi Tech*, *28*(1), 81-93. doi:10.1108/07378831011026706

Ransom, N. & Rafferty, P. (2011). Facets of user-assigned tags and their effectiveness in image retrieval. *Journal of Documentation*, 67(6), 1038–1066.

Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology, 61*(11), 2230-2242. doi:10.1002/asi.21401

Rosch, E. & Mervis, C. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, 7(4),573–605.

Rosch, E., Mervis, C., Gray W., David M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382-439. doi:10.1016/0010-0285(76)90013-X

Sigurbjörnsson, B., & van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *WWW '08: Proceeding of the 17th international conference on world wide web* (pp. 327–336). New York, NY, USA: ACM Press.

Smeulders A., Worring M., Santini S., Gupta A., & Jain R. (2000) Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1349-1380. doi:10.1109/34.895972

Springer, M., Dulabahn, B., Michel, P., Natanson, B., Reser, D., Woodward, D., & Zinkham, H. (2008). *For the common good: The Library of Congress Flickr pilot project.* Retrieved from http://www.loc.gov/rr/print/flickr_report_final.pdf

Srinivasan, R., Boast, R, Furner, J., & Becvar, K.M. (2009). Digital museums and diverse cultural knowledges: Moving past the traditional catalog. *The Information Society*, *25*(4), 265-278. doi**:**10.1080/01972240903028714

Stvilia, B., & Jörgensen, C. (2009). User-generated collection level metadata in an online photosharing system. *Library & Information Science Research. 31*(1), 54–65. doi:10.1016/j.lisr.2008.06.006

Stvilia, B., & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society of Information Science and Technology*, 61(12), 2477-2489. doi:10.1002/asi.21432

Stvilia, B., Jörgensen, C., & Wu, S. (2012). Establishing the value of socially created metadata to image indexing. *Library & Information Science Research*, 34(2), 99-109.

Tonkin, E, (2006). Searching the long tail: Hidden structure in social tagging. In: Proceedings of the 17th ASIS SIG/CR Classification Research Workshop.

Trant, J. (2009). Studying social tagging and folksonomies: A review and framework. *Journal of Digital Information,* 10(1). Retrieved 17 August, 2009 from http://dlist.sir.arizona.edu/2595/

Trant, J. (2006). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia*, *12*(1), 83-105. doi:10.1080/13614560600802940

Ungrangsi, R., & Anutariya, C. (2011). A comparison study of Flickr's folksonomies and ontologies. *Proceeding of Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 257-262, Nakhon Pathom, Thailand.

Vaughan, J. (2010). Insights into the commons on Flickr. *Portal: Libraries and the Academy*, *10*(2), 185-214. doi:10.1353/pla.0.0095

Yoon, J. (2008). Searching for an image conveying connotative meanings: An exploratory cross cultural study. *Library and Information Science Research, 30,* 312-318.

Yoon, J. (2009). Towards a user-oriented thesaurus for non-domain-specific image collections. *Information Processing and Management*, *45*(4), 452–468. doi:10.1016/j.ipm.2009.03.004

Zeng, Marcia Lei & Jian Qin. 2008. *Metadata*. New York, NY: Neal-Schuman.

Figures:

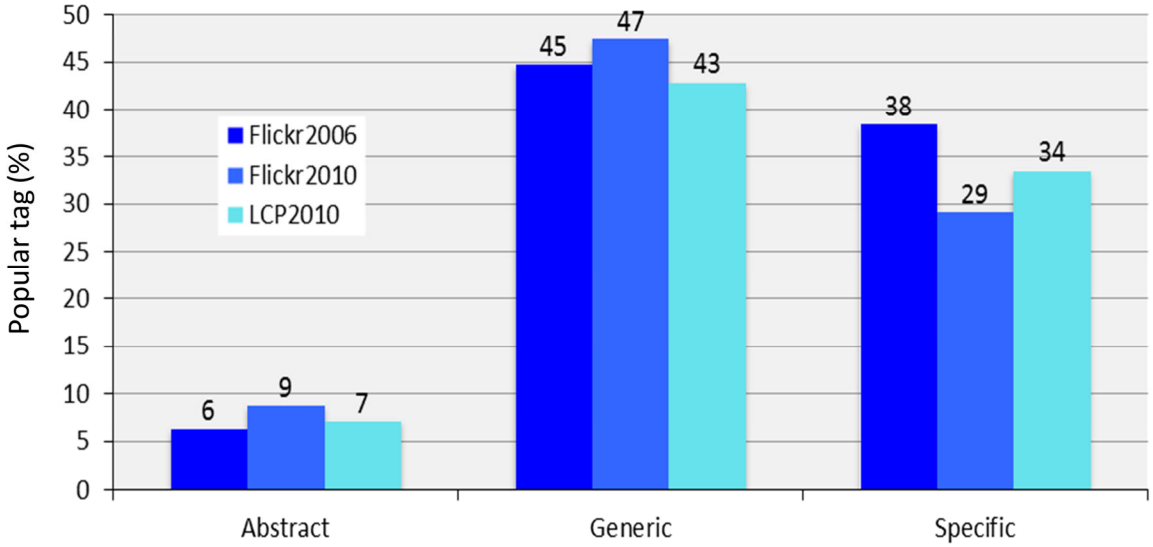

Figure 1: Distributions of "Abstract", "Generic" and "Specific" Popular Tags in Flickr 2006, 2010, and LCP 2010.

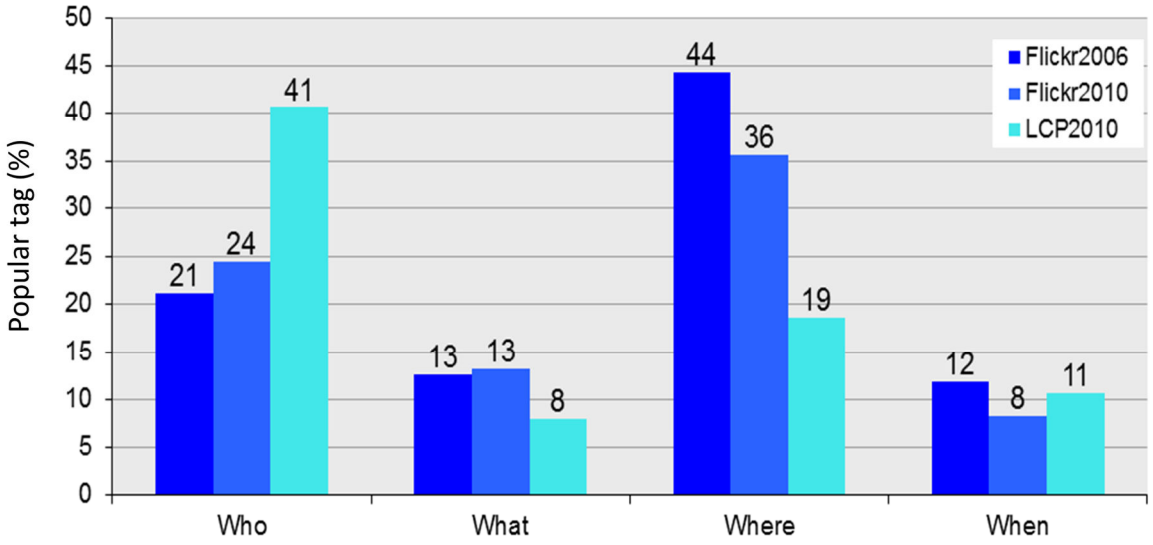

Figure 2: Distributions of "Who", "What" and "Where" popular tags in Flickr 2006, 2010, and LCP 2010.

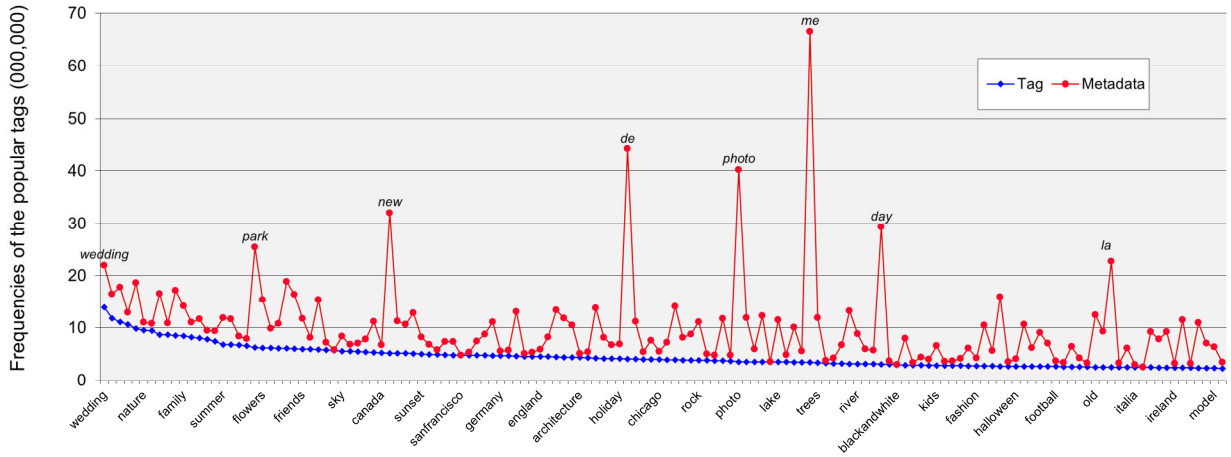

Figure 3: Usage frequencies of the Flickr 2010 popular tags as tags or metadata. The actual number in Y-axes is times one million.

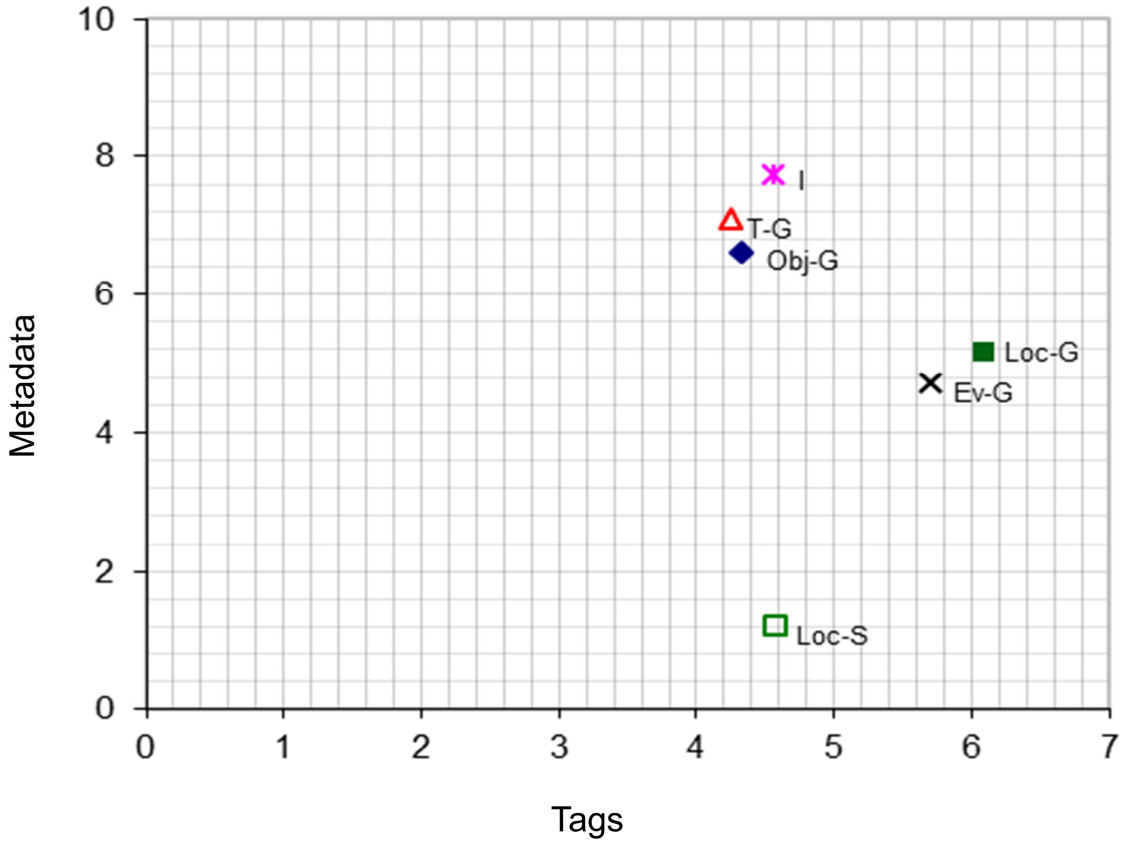

Figure 4: Scatter plot for the usage frequencies as tags or metadata for the Popular Tags classified by the Shatford schemes in Flickr 2010. The actual number in X, Y-axes is times one million.

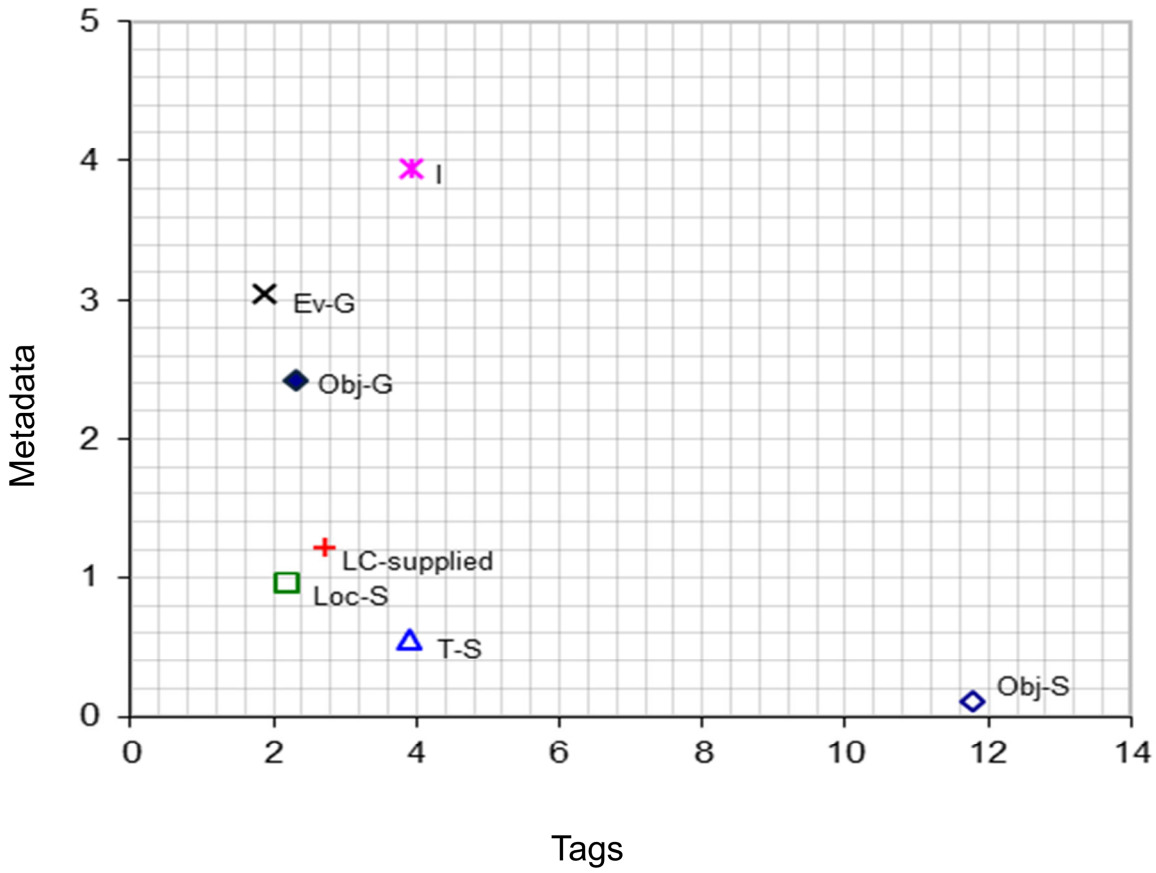

Figure 5: Scatter plot for the usage frequencies as tags or metadata for the Popular Tags classified by the Shatford schemes in LCP 2010. The actual number in X, Y-axes is times one hundred.

Table 1. Categories and facets of Popular Tags in Flickr 2006, 2010, and the LCP collection (2010).

Facet	Category	Flickr 2006 (N=141)	Flickr 2010 (N=139)	LCP (N=140)
Who (Obj)	Specific: Individually named person, group, thing (Obj-S)	0	0	11(8)
	Generic: Kind of person or thing (Obj-G)	30(21)	35(26)	46(33)
	Abstract: Mythical or fictitious being (Obj-A)	0	0	2(1)
What (Ev)	Specific: Individually named event, action (Ev-S)	2(1)	2(1)	4(3)
	Generic: Kind of event, action or condition (Ev-G)	14(10)	13(10)	7(5)
	Abstract: Emotion or abstraction (Ev-A)	2(1)	4(3)	1(1)
Where (Loc)	Specific: Individually name geographic location (Loc-S)	45(32)	38(27)	21(15)
	Generic: kind of place: geographic or architectural (Loc-G)	10(7)	10(7)	5(4)
	Abstract: Place symbolized (Loc-A)	6(4)	5(4)	2(1)
When (T)	Specific: Linear time: date or period (T-S)	8(6)	1(1)	11(8)
	Generic: Cyclical time: season or time of day(T-G)	8(6)	8(6)	0(0)
	Abstract: Emotion or abstraction by time (T-A)	1(1)	3(2)	5(4)
Expert-provided	LC-supplied: tags provided by Library Congress	0	0	13(9)
Others	Color: color, color value (C)	6(4)	6(4)	1(1)
	Image related: format, style (I)	8(6)	11(8)	11(8)
	Part of speech (P)	1(1)	3(2)	0

Note. Percentage values are shown in parentheses.

Table 2. Popular tags in Flickr 2006, 2010 by category and facet types.

	Year 2006		Year 2010
Obj-G	animals, art, baby, car, cat, clouds, dog, family, flower(s), food, friends, garden, girl, graffiti, house, kids, lake, light, me, music, people, portrait, river, rock, sky, snow, sun, tree, water, zoo	Obj-G	animals, art, baby, ***band***, ***bike***, ***birds***, ***bird***, car, cat, clouds, dog, family, flower(s), food, friends, garden, girl, graffiti, house, kids, lake, light, me, ***model***, music, people, portrait, river, rock, sky, snow, sun, tree(s), water, zoo
Ev-S	Christmas, halloween	Ev-S	Christmas, halloween
Ev-G	birthday, concert, festival, holiday, party, show, travel, trip, vacation, wedding, ***camping***, ***honeymoon***, ***roadtrip***, ***hiking***	Ev-G	birthday, concert, ***dance, football***, festival , holiday , party, show, ***tour***, travel, trip, vacation, wedding
Ev-A	Live, fun	Ev-A	Live, fun, ***fashion***, ***love***
Loc-S	***amsterdam***,australia, barcelona, berlin, ***boston***, california, canada, chicago, china, england, europe, florida, france, germany, hawaii, india, ireland, italy, japan, london, mexico, newyork, newyorkcity, ***newzealand***, nyc, paris, ***rome***,sanfrancisco, scotland, seattle, spain, ***sydney***, taiwan, texas, thailand, tokyo, toronto, uk, usa, washington, ***vancouver***,york, dc, africa, hongkong	Loc-S	***asia***, australia, barcelona, berlin, california, canada, chicago, china, england, europe, florida, france, germany, hawaii, india, ireland, ***italia***, italy, japan, london, mexico, newyork, newyorkcity, nyc, paris, sanfrancisco, scotland, seattle, spain, taiwan, texas, thailand, tokyo, toronto, uk, usa, Washington, york
Loc-G	beach, church, city, island, mountain(s), museum, ocean, park, sea, street	Loc-G	beach, church, city, island, mountain(s), museum, ocean, park, sea, street
Loc-A	Geotagged, nature, urban, Architecture, landscape, ***home***	Loc-A	Geotagged, nature, urban, Architecture, landscape
T-S	***06***, july ***, august***, ***june***, ***may***, september, ***october***, april,	T-S	july
T-G	autumn, day, fall, night, spring, summer, sunset, winter	T-G	autumn, day, fall, night, spring, summer, sunset, winter
T-A	new	T-A	new, ***old, vintage***

C	black, blue, green, red, white, yellow	C	black, blue, green, red, white, yellow
I	***cameraphone***, blackandwhite, canon, color, film, macro, nikon,bw	I	blackandwhite, bw, canon, color, film, macro, nikon, ***photography***, ***photo(s)***, ***raw***, ***iphone***
P	san	P	***de***, ***la,*** san

Note. Bold/Italics: Tags that dropped off (2006) or were added (2010) to the most popular tags in Flickr.

Table 3. Popular Tags in LCP by category and facet types.

Obj-S	al(American League), cymru, libraryofcongress, nl(National League), greatmustachesoftheloc, historicalphotographs, newyorkbaseballgiants, newyorkgiants,, unitedstatesnavy, usn(us navy), usnavy,
Obj-G	aircraft, airplane, army, athlete, aviation, baseball, battleship, beard, boat, boxer, building, cap, car, children, clouds, crowd, dress, factory, farm, flag(s), hat(s), horse(s), house, man, men, military, moustache, mustache, navy, plane, people, portrait, railroad, river, ship, sky, snow, soldiers, suit, train, tree(s), uniform(s), water, woman, women, worker(s)
Obj-A	fashion, royalty
Ev-S	Iwd (international women's day), worldwar2, worldwarii, ww2
Ev-G	boxing, football, parade, seated, sport(s), standing, war
Ev-A	suffrage,
Loc-S	america, california, chicago, illinois, ireland, manhattan, mexico, newmexico, newyork, newyorkcity, norway, ny, nyc, philadelphia, texas, us, usa, wales, washingtondc, york, pologrounds
Loc-G	city, field, mountains, stadium, street
Loc-A	architecture, landscape
T-S	1911, 1912, 1913, 1940, 1941, 1942, 1943, 1940s, forties, october1942, october
T-A	history, wartime, new, old, vintage
LC-supplied	alfredpalmer, alfredtpalmer, americanleague, bainnewsservice detroitpublishing, fsa(Farm Security Administration), georgegranthambaincollection, jackdelano, howardhollem, howardrhollem nationalleague, russelllee, WWII
C	White
I	4x5, blackandwhite, bw, color, colorized, glassnegative, largeformat, LF (large format), photochrom, slidefilm, transparencies

Table 4. Popular Tag category-facet distributions of metadata usage propensity ratios in Flickr 2010.

Flickr	Obj-S	Obj-G	Obj-A	Ev-S	Ev-G	Ev-A	Loc-S	Loc-G	Loc-A	T-S	T-G	T-A	LC-supplied	C	I	P
0~10		1				1					1	2			2	3
11~20		4						3		1				1	1	
21~30		5			2			2			1					
31~40		5			1			2						1	1	
41~50		5			1	1					2			1		
51~60		6			2			1						1		
61~70		3			1	1	1	1			1			1	1	
71~80		3			1		3				1			1	1	
81~90		2		1	1	1	1	1	1		1	1				
91~100		1			3		3		1						2	
101~110							8				1				1	
111~120					1		8								1	
121~130							9								1	
131~140				1			4		3							
Total		35		2	13	4	38	10	5	1	8	3		6	11	3

Note. The abbreviations for the corresponding categories can be found in Table 1. The cell value for each category refers to the number of popular tags within a particular distribution level in the "Flickr" column.

Table 5. Popular tag category-facet distributions for metadata usage propensity ratios in the Flickr LCP 2010.

LCP	Obj-S	Obj-G	Obj-A	Ev-S	Ev-G	Ev-A	Loc-S	Loc-G	Loc-A	T-S	T-G	T-A	LC-supplied	C	I	P
0~10		1			1		2			1		1	1	1	2	
11~20	1	4					1	1	2			1				
21~30	1	4			1			1		2		1				
31~40		5					2			1					2	
41~50		6			1		2	1								
51~60		6			1		1			2						
61~70		5			2		1	1		1						
71~80		7					2	1								
81~90	1	5	1			1	2									
91~100		2	1	1			4			1		1				
101~110	2	1		1			2			1		1			2	
111~120	3			1	1		1			1			1		2	
121~130	3						1			1			2		3	
131~142				1									9			
Total	11	46	2	4	7	1	21	5	2	11		5	13	1	11	

Note. The abbreviations for the corresponding categories can be found in Table 1. The cell value for each category refers to the number of popular tags within a particular distribution level in the Flickr LCP.

Table 6. Kruskal-Wallis test of dependence of the *Generic, Abstract, and Specific; or Who, What, Where, and When* related popular tags metadata propensities in Flickr 2010 and LCP 2010.

	Generic	Abstract	Specific	Who	What	Where	When
Chi-Square	8.25	0.00	0.564	8.64	2.22	0.53	1.35
df	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Asymp. Sig.	0.004*	1.000	0.453	0.003*	0.136	0.465	0.245

* $p<0.005$