This is a preprint of an article accepted for publication in Journal of the American Society for Information Science and Technology. Jörgensen, C., Stvilia, B., & Wu, S. (in press, 2013). Assessing the relationships among tag syntax, semantics and perceived usefulness. *Journal of the American Society for Information Science and Technology*.

Assessing the Relationships among Tag Syntax, Semantics and Perceived Usefulness

Corinne Jörgensen[1]

College of Communication and Information, School of Library and Information Studies, Florida State University, Tallahassee, FL 32306-2100, United States. E-MAIL: cjorgensen@fsu.edu

Besiki Stvilia

College of Communication and Information, School of Library and Information Studies, Florida State University, Tallahassee, FL 32306-2100, United States. E-MAIL: bstvilia@fsu.edu

Shuheng Wu

College of Communication and Information, School of Library and Information Studies, Florida State University, Tallahassee, FL 32306-2100, United States. E-MAIL: sw09f@my.fsu.edu

[1] Corresponding author

This is a preprint of an article accepted for publication in Journal of the American Society for Information Science and Technology. Jörgensen, C., Stvilia, B., & Wu, S. (in press, 2013). Assessing the relationships among tag syntax, semantics and perceived usefulness. *Journal of the American Society for Information Science and Technology*.

Abstract

With the recent interest in socially created metadata as a potentially complementary resource for image description in relation to established tools such as thesauri and other forms of controlled vocabulary, questions remain about the quality and reuse value of this metadata. This study describes and examines a set of tags using quantitative and qualitative methods and assesses relationships among categories of image tags, tag assignment order, and users' perceptions of usefulness of index terms and user-contributed tags. The study found that tags provide much descriptive information about an image but that users also value and trust controlled vocabulary terms. The study found no significant correlation between tag length and assignment order, and tag length and its perceived usefulness. The findings of this study can inform the design of controlled vocabularies, indexing processes, and retrieval systems for images. In particular, the findings of the study can advance the understanding of image tagging practices, tag facet/category distributions, relative usefulness and importance of these categories to the user, and potential mechanisms for identifying useful terms.

Keywords: image tagging, image indexing, social indexing, social metadata, image syntax, image semantics, image interpretation

Introduction

With the implementation of image sharing sites such as Flickr (which has grown rapidly to several billion images, many of which are made available for viewing by all

Flickr users), the issues associated with describing and finding images have moved from a research area only of interest to a smaller group of scholars and professional content providers to being of interest to a much wider audience, including both those who seek images and those who post and describe them (Springer et al., 2008). Content creation, including description of that content, has resulted in an increasing number of web sites that allow the addition of "tags" (user contributed descriptive terms) and user comments, both of which can be mined for descriptive purposes or ontology creation and/or searched by end-users. This phenomenon, sometimes referred to as "democratic" (Hidderly & Rafferty, 1995; Albrechtsen, H. 1998) or distributed indexing in Web based social content creation and/or sharing systems (Munk & Mork, 2007; Rafferty & Hidderley, 2007) has many implications for more traditional methods of indexing, which usually has requirements and guidelines aimed at producing a constrained, consistent, and authoritative description which provides a level of confidence in that description.

While it is still unclear how the nature of the relationship between controlled vocabularies and user-contributed tags will ultimately resolve, there has been much interest in this socially created metadata as a potentially complementary resource to established tools such as thesauri and other controlled vocabularies used for document description (Rolla, 2009; Jörgensen 2007) and a source for automatic generation of ontologies (Zhitomirsky-Geffet, Bar-Ilan, Miller & Shoham, 2010). The growing public participation in social content creation communities (e.g., Flickr, LibraryThing, and Wikipedia) makes it possible to accomplish the task of describing items at a relatively low cost, including describing millions of photographs and generating new knowledge

organization systems (KOS, e.g., DBpedia). Because of the quantity of data generated through tagging, there is also interest in identifying potentially useful tags automatically (e.g. Sigurbjörnsson and Zwol, 2008).

There has also been an increase in efforts to harvest additional or missing metadata for existing image or photograph collections by deploying these collections in social content creation communities. One of the most notable of these is the Library of Congress's incorporation of a set of 7,192 historical images from the Prints and Photographs Division (2008) within the photo sharing site Flickr. A potential benefit of this is that user-contributed metadata could compensate for some of the difficulties associated with controlled vocabularies used to describe these collections, such as the need for constant maintenance and revision as changes in domain knowledge, culture, activity systems, technology, terminology, and user expectations occur.

However, there are also a number of factors that could motivate against the use of socially created metadata for additional description. Socially created metadata exhibit all of the problems that controlled vocabularies attempt to limit. For instance, different types of users create social metadata in different contexts and for different purposes (Cunningham & Masoodian, 2006; Stvilia & Jörgensen, 2009). The terms users select to search for images may differ from those they use to describe images (Chung & Yoon, 2008). Furthermore, while quality is being recognized as contextual and dynamic (Jörgensen, 1995; Strong, Lee, & Wang, 1997; Stvilia, Gasser, Twidale, & Smith, 2007), adding new metadata may not necessarily lead to value increase or cost reduction for the maintenance and upkeep of KOS (Stvilia & Gasser, 2008). Thus, questions remain about

the quality and reuse value of such metadata as it exhibits many of the inconsistencies of natural language. There are also questions about the extent to which users would trust the added tags. Mai (2007) suggests that the issue of trust in tags is related to the transparency of a system and argues for a social constructivist approach in which multiple interpretations are allowed. Before integrating socially created metadata with traditional image KOS, it is necessary to evaluate this newer type of metadata in general and, for the purposes of this research, for image indexing in particular.

The use of knowledge organization and representation tools is both well-established and ubiquitous in libraries, museums, and other information-intensive settings. Earlier studies evaluated the added value of social metadata by comparing it to traditional controlled vocabularies for images (Jörgensen, 1994; Rorissa, 2010), but the majority of this research has used text documents, rather than images, as test cases. The current study builds on previous research results showing that "social" terms (terms added by end-user tagging) can add value, both in subjective (participant ratings of terms) and objective (degree of added coverage provided by the terms) analyses (Stvilia & Jorgensen, 2010; Stvilia, Jorgensen, & Wu, 2012). Previous work by the authors also explored relationships between the participants' perception of the value of both tags and controlled vocabulary and participant demographic characteristics and indexing or tagging experience. The research reported here examines quality measures of image indexing terms from the user's perspective by asking what terms are useful to a user to describe a particular image and what criteria make the terms useful to the user. It also examines whether there are syntactic and semantic aspects of terms that could facilitate selection of

these inexpensively—that is automatically. The semantic aspects were addressed through coding of the image tags into their semantic attribute classes developed in previous research.

Research questions

The purpose of this study is to further evaluate socially created metadata for images in relation to its semantic and syntactic features and to assess the quality and applicability of such metadata to image indexing from the users' perspectives. The study also explores whether there are inexpensive mechanisms for identifying useful terms for image description in a set of socially created metadata. Specifically, the study posed the following seven research questions in relation to a specific set of Flickr images and their tags. The first four investigate aspects of the semantics of tags:

1. What kinds of terms do users assign to images as tags?
2. What kinds of terms would users select as tags in image description, as indicated by their ratings of a set of pre-assigned terms from most to least useful?
3. What are the criteria that make an index term useful to the user in image description?
4. What kinds of terms would users choose to create a query for an image, and do these differ from terms chosen to describe the images?

The last three research questions investigate syntactic aspects of tags and whether there is an easily discerned relationship to semantics:

5. Is there a relationship between types of user assigned image tags and the tag assignment order?
6. Is there a relationship between tag length and tag assignment order?
7. Is there a relationship between tag length and a user's perception of its usefulness?

This research presents additional fine-grained analysis of the set of data from an earlier study reported in Stvilia et al. (2012) to further evaluate syntactic aspects of the data and develop the category semantics associated with the data through coding, as well as testing these results against a current sample of Flickr tag data.

Related work

Traditionally, image access has relied primarily on human indexers and the use of image knowledge organization and representation systems to translate sensory data into concepts and entities that are understandable and searchable by humans (Heidorn, 1999). These systems generally use some sort of controlled vocabulary as a tool to insure consistency in indexing and thus in retrieval, and a users' search terms must be translated to those terms chosen as access points within this vocabulary. A parallel body of research within computer science focuses on automated methods of parsing visual content (called content based image retrieval or CBIR) and utilizing these results to search for syntactically similar images (Smeulders, Worring, Santini, Gupta, & Jain, 2000). There has also been work done to create models capable of translating data produced by these methods into higher level concepts allowing object recognition (thus bridging a "semantic gap" between visual percepts and semantic concepts) (Tang, Yan, Hong, Qi, & Chua, 2009; Chua et al. 2009), but these methods are generally more successful within

constrained domains and produce less satisfactory results in a more generalized collection of images.

With the availability of systems that facilitate wider participation in providing descriptions of image content, such as tags, researchers have begun analyzing the nature of these "uncontrolled" vocabularies and evaluating their potential for contributing to increased –and potentially less expensive – access to visual materials. Art museums have been particularly active in investigating the utility of user tagging (Smith, 2011), and leading museums, such as the Metropolitan Museum of Art and the Guggenheim Museum, have developed prototype systems. A study by Trant (2009) found that museum professionals find the user assigned tags generally useful. Tags have also been investigated in relation to their potential added value in library catalogs (Kakali & Papatheodorou, 2011; Redden, 2010). There is now a growing body of research suggesting that socially created metadata (e.g., user-generated tags) could be complementary to traditional KOS (e.g., Jörgensen, Stvilia, & Jörgensen, 2008; Matusiak, 2006; Rolla, 2009; Stvilia, Jörgensen, & Wu, 2012; Wetterstrom, 2008, Petek, 2012). However, concerns still remain about the quality and consistency of tags, the nature and value of structured and unstructured description, and what part they can play in providing different levels of access to images, and opinions range from implementing authority control tags through implementing various kinds of structured formats (Spiteri, 2010; Smith, 2011) to simply "letting them be."

One of the main challenges of indexing in information retrieval is to determine which terms and phrases are most informative surrogates for the document's semantic content.

Similarly, in concept based image indexing and retrieval, it is important to determine which terms and phrases are most informative about the image's semantic content. In traditional manual subject indexing this is accomplished by conducting careful subject analysis of the document's content, identifying significant topics of the document and then selecting appropriate subject headings through the use of elaborate cataloging rules and controlled vocabularies (http://www.loc.gov/cds/products/product.php?productID=79). In automated indexing, a term or phrase frequency based measure often attenuated by the term's occurrence in other documents of the collection is used for assessing the significance of the term as a representation of the document's overall semantic content (Manning, Raghavan, & Schutze, 2008). In social tagging environments, however, one cannot assume the user to be familiar with the rules and codes of subject cataloging. Furthermore, in social content sharing and tagging systems such as Flickr where a particular tag can be assigned only once to an image, the set of tags of the image is not a 'bag of words' representation of the image's semantic content. Therefore, a tag frequency based measure may not be useful for the automated assessment of the tag's significance to the photo's semantic content. To address this challenge, researchers have suggested the use of heuristics such as tag assignment order to identify useful index terms automatically (Golbeck, Koepfler, & Emmerling, 2011), but the utility of these approaches has not been demonstrated.

4. Methods

Data collection

The data analyzed here were generated in a larger study reported in Stvilia, Jörgensen, and Wu (2012) where the methods are explained in depth. To briefly recap, the study used a mixed methodology and adapted the controlled experimental design used by Jörgensen (1998) and Chen et al. (1995) to collect data from participants to answer the research questions. The experiment involved a sample of 35 participants, including students (both graduate and undergraduate) and faculty and staff members recruited from the College of Communication and Information at Florida State University. Participants first performed a free-tagging task (Description Task) and then rated a set of pre-assigned terms on their perceived utility (Evaluation Task). Participants later wrote queries to retrieve these same images in an approximation of a known-item search (Query Task). Participants performed the tasks on a copy of Steve tagger software (http://sourceforge.net/projects/steve-museum/) modified for the purposes of this research (Figs. 1 & 2). Pre-experiment questions indicated that very few participants were familiar with the concept of controlled vocabularies (somewhat surprising given the broad discipline most of the sample was drawn from), and while more were familiar with tags, the majority had not done tagging before. For a comparison set of results, a set of tags was sampled from a download of all tags on the Flickr Library of Congress image collection; these were coded using the same methods as were used for tags in the Description and Query Task.

In the Description Task, participants were presented with ten historical photographs selected from the Library of Congress Flickr photostream which on the download date (on September 13, 2009) contained 7,192 photographs. Criteria for selection included the variety of topics and reasonable clarity of the image. A pretest determined the optimal number of images which would allow participants to complete the task without undue fatigue. Participants were asked to describe the images by tagging with terms they felt described the image. In contrast to how these same images are viewed on Flickr, the images were presented in isolation, with none of the text and notes provided by the Library of Congress that displays on Flickr. People viewing the LC images on Flickr can also see user-contributed tags; these were not displayed either in the experimental interface. Thus, during tagging the participants had only the image (which in some cases had written notes or captions on the photographs) and a text box within which to enter their tags (Fig. 1). The entry screen for tagging displayed all ten images at once (Fig. 2).

Insert Fig. 1 here

FIG. 1. The Steve tagging interface.

Insert Fig. 2 here

FIG. 2. Steve tagging entry screen

Next, in the Evaluation Task, for each of the same ten photographs participants were shown a set of pre-assigned terms developed by the researchers representative of those in a controlled vocabulary. The pre-assigned terms were generated through several

processing methods from the following six sources: (a) the LC Thesaurus for Graphic Materials (LCTGM), (b) the Library of Congress Subject Headings (LCSH), (c) tags assigned to the photographs by Flickr members, (d) a folksonomy generated from the Library of Congress Photostream on Flickr, (e) the folksonomy from the complete Flickr database, and (f) the English Wikipedia. Further detailed description of how each dataset of pre-assigned terms were obtained can be found in Stvilia, Jörgensen, and Wu (2012). As this process produced an extensive list of terms, the researchers discussed the merged lists of terms, eliminated word variations and term overlap between the LCTGM and the LCSH (the LCTGM terms are derived from the LCSH), and chose a subset of the most representative (N=348) to make the task manageable for the participants during the one hour timeframe that the pretest indicated would be needed for the tasks.

While six very different resources were used to generate terms, the majority of the terms came from the LCTGM, the tags, and the Wikipedia folksonomy. The Flickr database also provided a large number, while the folksonomy from the LC Photostream folksonomy added only a few terms (the number chosen does not suggest a value judgment, but rather depends on the goals each set of terms fulfills for its system). The participants rated these preselected terms on a five-point Likert scale (i.e., ‘strongly disagree’, ‘disagree’, ‘neutral’, ‘agree’, ‘strongly agree’) by their usefulness for describing the content of the photographs. The number of terms evaluated by subjects varied among the images and ranged from 9 to 98. At the end of the Evaluation Task, participants completed a semi-structured interview with regard to their decision-making process in rating the usefulness of the pre-assigned terms. Questions included the

difficulty of the task (prompts as needed: too long, too many terms to evaluate, software difficult to use, terms difficult to understand); if the participants found the terms were useful, they were asked to describe in their own words what made these terms useful for them (prompts as needed: relevant, informative, easy to use/connect to the image). Likewise, if terms were not useful, participants were asked to describe why.

The final task, the Query Task, took place two weeks later; 28 of the previous participants performed the task and viewed the same ten images again. For each image, the participants were asked to formulate a query that, in their opinion, would allow them to locate the image with a hypothetical search engine with the least effort (i.e., with the least amount of browsing and query revision).

4.2 Data analysis

To identify the kinds of self-generated terms users assign to images as tags (Description Task) and the kinds of terms generated during a Query Task, two of the researchers coded both the tags (which could be single terms or phrases) assigned by participants in the Description Task and the terms generated by participants during the Query Task described above according to Jörgensen's (1995, 1998) scheme of general classes composed of a number of specific types of attributes. The data were divided among two of the researchers for coding, then these data sets were exchanged and coded by the second researcher. Differences that occurred when the two sets of coding were compared were easily resolved (most were the result of one of the researchers being less familiar with the attribute definitions) and produced excellent agreement across the codes. Individual tags and multi-tag phrases were coded, and as many codes as necessary were

added to capture the types of attributes making up the larger classes. Phrases were broken apart and the individual terms were coded if necessary for meaning (e.g. "woman carrying small dog" was coded as People, Activity, Object, and Descriptor). The third researcher analyzed the syntactic aspects of attributes using a set of algorithms and statistical measures.

During the coding process the researchers decided to add a new category to the original set, that of "Group." This refers to organizations, institutions, or social groups with a particular purpose or goal and which exist over time, as opposed to a transient or spontaneous gathering of people (Conduit and Rafferty, 1997, also used "group" in their indexing template to indicate quantity of people, a slightly different definition than that used for the current research). While terms with this meaning had been assigned to another category in previous research, their frequency in this particular set of historical data, especially where additional text information about the image was visible to Flickr viewers through the notes and tags provided, suggested the utility of adding Group as a separate category. Rather than trying to determine whether a tag is at the specific or generic level, which can be problematic, for tags that were proper nouns a simple tag (PN) denoting this was added.

The detailed list of specific attributes was used to assist in coding and data analysis by providing more detail about the types of attributes that make up the larger classes. To be consistent with prior research, the 1993 terms or phrases generated by the participants resulted in 2829 coded attributes, which were then grouped into ten classes of attributes (two of the original classes concerning user reactions were not applicable to this research).

While the classes used in this research are not the same as the categories used by the Library of Congress in their 2008 analysis, there appears to be reasonable correspondence between these two coding schemes, in both coverage and occurrence (Table 1). Further direct comparison is not possible however because of issues of granularity and interpretation.

Insert TABLE 1 here.

For the final three questions examining the relationships among tag categories, tag length, and tag assignment order, for each combination of tag, image, and participant, tags were assigned numbers indicating their length and the order in which the participant assigned the tag to the image. As the Wilks-Shapiro test determined that neither tag assignment order nor term ratings were normally distributed, the researchers used a nonparametric Kruskal-Wallis test to examine relationships among tag categories, tag assignment order, tag length, and term rating.

Results

The first research question asked what kinds of terms users generated during the task of assigning tags to images (Description Task). The coding of these user-generated tags indicated that, similar to prior research findings using the same schema (Jörgensen, 1998; 1999; Petek 2012), Object terms remained the most frequently used category (Table 2). For this set of photographs, the next most frequent category was composed of attributes belonging to the "Story" of the image. These two categories, along with People

and People-related attributes (16%), accounted for approximately 70% of the terms, as found in prior research. There were no tags representing Visual Element attributes, very few Color and Location attributes, and very few tags relating to the format for this set of photographs. As there was no additional text but notes that were written directly on the images, there were (not surprisingly) very few proper nouns (6.5% of the total participant tags, referring most frequently to a person's name and to the perceived locale of the image, such as Europe). There were also few dates assigned (58, or 2.9% of the total). Most of the images (8 out of 10) contained some text directly on the photograph, either short handwritten notes (or in one case text on signs depicted within the image itself). In cases where the text was clearly legible and indicated location, person, or date (two images), not surprisingly these were added as tags. This finding is in line with the analysis done by the Library of Congress (2008), which found that for LC images in Flickr, the more text an image record contained the more these LC descriptions were copied as tags, with the majority being for creator, place, and time period. For images with no text, tags indicating time or location were largely incorrect (e.g., participants tagged one LC image in the Flickr Commons with several different guesses of time period or date: 1800s, 1900s, 1930s, and twentieth century).

Insert TABLE 2 here.

To evaluate whether the results of the coding were consistent with the larger set of tags which occur on the LC Flickr site, in 2012 one of the researchers downloaded the full set of unique tags assigned to the LC images in Flickr (N=36,681) and a similar size sample (.05%, N=1821) of these was randomly selected for coding. Overall, there was good

correspondence in the results of the coding, with one important difference between these two sets of data: taggers on Flickr had the additional text information in the LC record and prior assigned tags available to them for the LC images. When taggers in general can view this additional text, it impacts the tagging process, with taggers not only copying visible text (names of people) into the tag field, as was noted in the LC's 2008 report (Springer et al. 2008), but copying the same information in multiple forms (e.g. Glenna Tinnin; Tinnen; Mrs. Glenna; Glenna Smith Tinnin). Taggers on Flickr seem to be far more likely to list people's names when these appear as text with the image and somewhat less likely to name objects. However, in the sample of 10 images selected for the research reported here, one picture appeared to be an outlier, as it was a complex image with many objects depicted, whereas the other images were more general scenes; consequently, many more objects were noted for this one image than for the other images (this one image contributed 49.4% of the total Object terms for the 10 images). We also see a large percentage of Story terms; similarly, in the LC tag sample the information about where the photo was taken appeared frequently in the text, and this one attribute, "Setting," accounted for over 50% of the Story tags in both samples.

The second research question analyzed the kinds of pre-assigned terms users would select as tags in image description, as determined by their ratings of the usefulness of these terms. The Kruskal-Wallis test showed that both specific attribute and general category codes were statistically significantly different on participants' usefulness rating of pre-assigned terms ($\chi2 = 549.8$, $df = 24$, $p < 0.001$; $\chi2 = 251.2$, $df = 11$, $p < 0.001$). The results are interesting in that they demonstrate that frequency of a tag or term does not

necessarily correlate with a perceived usefulness rating that participants assign (Table 3). For example, while Abstract category terms accounted for less than 1% of the tags, by percentage, they were rated third highest in usefulness. Similarly, Description terms, which can include adjectives describing qualities of objects that are not routinely indexed ("wooden," "broken"), are also highly rated. These two types of terms are widely considered too subjective and inconsistent to assess and generally are not found in controlled vocabularies. Story terms are also included in the top six Classes of terms, but aside from terms relating to the setting of the image these are also usually not indexed for the same reasons. In contrast to these, while Objects are consistently among the most described attributes, they were ranked lower than the top six categories. It should be noted that People and Group also rank highly; the high ranking for Group occurs because in many instances the group that is depicted is composed of individual people in the picture (for instance, "army" would refer to a group of people in soldiers' uniforms). The categories which participants rated the least useful in a known item search include terms describing technique and other aesthetic considerations (in Art Historical) and Visual Elements (texture, focal point) as well as (not unexpectedly) personal reactions ("wow").

Insert TABLE 3 here.

The third research question asked what criteria a participant uses to rate a term as useful in image description. After rating the pre-assigned terms on a 5-point scale, interviews were conducted in which participants were asked "Could you please give me some examples of the third party terms you found particularly useful, and could you explain why?" Interviews were recorded, and content analysis of the transcripts was conducted

using opening coding. The indicators participants named in evaluating the usefulness of image index terms were coded; in order to determine the range of indicators only the first occurrence of an indicator during the interview was coded. In most cases, the participants' own language was used to determine indicators and criteria. Adjectives were changed to nouns (e.g., "accurate" to "accuracy") and in some cases a longer explanation was summarized in a word or phrase (e.g., describing focus of the image). This produced a total of 35 indicators from which 15 broad criteria were developed. Two of the researchers created initial groupings by analyzing which phrases or terms were appropriate upper level criteria and which terms or phrases should be indicators of those criteria. The final criteria were determined by discussion and consensus among all the researchers.. The final criteria emerged in coding the transcripts and reflected the interview prompts as well as phrases that participants used to describe their decision processes in rating the terms.. For example, one participant stated that a term was useful if it "gives context, background information you may not know or tell from the image itself such as the date or the period of photo taken," while another commented on a tag which referred to an object in the picture: "I did not notice it, until I saw someone tagged with."

In order of frequency, the finalized criteria for usefulness across the participants were Informativeness (providing contextual information not available from viewing the picture (date, location, what was going on in the image); Relevance; Connecting to the Image (describing the process of understanding how a specific tag is relevant by examining details in the picture); Accuracy (correct spelling, unknown technical term that appears

"legitimate"); Specificity (more explicit terms, such as Tokyo rather than Japan), Descriptiveness, Level of Detail, Personalization (used when a tag was an exact match with what the participant would use), Ease of Use, Importance, Generality; Redundancy; Objectiveness; Moods; and Structure (syntax of phrase that suggests it comes from a controlled vocabulary). Two broad observations can be made from these criteria. The first three criteria suggest that these participants found terms that contribute to the viewer's understanding of the context of the image are helpful in first understanding the image, while the next four criteria seem to be more related to specific details about the image as well their accuracy. These seem to contribute not only to understanding the image but to the participants' confidence in trusting the terms. The last group seems to relate more to the individual experience of the image or to the knowledge level of a particular participant: "I would say easy to use and familiarity because some of those terms I did not even know what they were. So perhaps if I were a very specialized researcher, I would, you know, know to…Yes. Like for the Norwegian building like I knew it was a state church but I think the actual term was in Norwegian like "stavkirke" or something like that, and I do not think that most people probably know that."

The fourth research question asked what kinds of terms users would choose to create a query for a known image search, and if these differed from the kinds of terms chosen to tag or describe the images. For this Query Task, participants wrote queries while viewing the images, and they could use any format they chose to express the queries. The structure of the queries fell into two broad types and seemed to indicate the degree of familiarity a participant had with concepts of online searching. A few contained some

structure, either denoting terms that should appear together ("vintage images" "black and white" "steam ship" "nautical photography") or the relationship of items in the image (ship sailing with the smoke), but the vast majority of queries were an unstructured list of terms (boat, sea, wave, smoke). Participant's query terms were coded using the same coding method as was used for the tags. As text was visible on some of the images, these text terms were included in some of the queries across the ten images. The number of instances of copying text varied from one to 68 occurrences across the ten images and all 283 participant queries (29 participants started the task but one participant only wrote queries for three images).

Results indicate that query term class ranks (by frequency) do not mirror the ratings of attributes found in the Evaluation Task tag classes (Table 2, Table 3). Results from the coding reveal that, for the queries, the Objects and Story Classes account for the greatest percent of the query terms, with People and People-Related Attributes following close behind; the results for the Query Task are more similar to the Describing Task than the Evaluation Task. Although Art Historical terms also appear prevalent, this is accounted for by the way that many participants worded their queries, frequently beginning with "A picture of…" The other Classes appear far less frequently. One might expect that participants would have rated tags/terms somewhat equivalently in the Query Task and Evaluation Task, given that the latter was designed to elicit value judgments about the relative value of different types of tags. However, the difference here may lie more in the similarity between the Query Task and the Describing Task; in both cases participants were asked to generate terms. In contrast, in the Evaluation Task the participants did not

generate the terms themselves; they were given the terms and asked to evaluate them as to their potential usefulness, a fundamentally different task from the Query and Describing Tasks. While participants could evaluate the usefulness of terms that they may not have been very familiar with ("stave church"), generating these terms would probably have been almost impossible for most, if not all, of them.

The findings from the Description and Query Tasks, as well as the analysis of a sample of tags from the complete LC tag corpus, are in line with those of other recent research, although direct comparison is not possible as different coding schemes were used among the studies. As just one example, in one coding scheme used by several researchers both People and Objects occur in the same facet (Shatford, 1986). The most recent research done by Ransom and Rafferty (2011) indicates People and Objects are most frequently tagged, confirming several earlier research studies, while other researchers (Beaudoin, 2007) have found that Location ("Setting," part of the "Story" class in the current research) is most frequent. The research results reported here indicate the importance of all three (note: Setting, referred to in other research as Location, is part of the Story class in the current research).

The fifth, sixth, and seventh questions analyzed syntactic and semantic relationships among tag assignment order, tag categories, and users' perception of tag usefulness. The fifth research question examined the relationship between the categories of tags users assign to images and the tag assignment order. The Wilks-Shapiro test showed that neither tag assignment order nor ratings of pre-assigned terms were normally distributed. Hence, the researchers used a nonparametric Kruskal-Wallis test to examine relationships

among tag assignment order and tag categories. The Kruskal-Wallis test showed that both specific attribute and general tag categories were statistically significantly different on the order of tag assignment ($\chi2 = 132$, df = 28, $p < 0.001$; $\chi2 = 58.6$, df = 11, $p < 0.001$). Table 5 shows that, on average, Group was the first assigned general category of tags, followed by Color, Art Historical, Story, People, Human Attributes, and Abstract, with Objects assigned last.

Insert Table 5 here.

The sixth research question asked whether there was a relationship between tag length and tag assignment order. The study did not find a significant association between tag length and assignment order (Spearman's rho 0.06, $p<0.01$). The seventh research question analyzed the relationship between term length and the user's perception of term usefulness; as with the sixth research question there was not a significant association between pre-assigned term length and the usefulness ratings (Spearman's rho 0.07, $p<0.001$).

The preliminary findings of this study suggest that, for these last three questions, while in some instances users might assign general categories of terms (e.g., Group, People) that they perceive most useful first, order is not necessarily an indicator of perceived utility, as can be seen with Color, which although high in assignment order ranked low in perceived usefulness (Table 5). As Objects are named frequently it is also interesting that this category rates last in assignment order. Although tag order has been proposed as a guide that library and museum communities could use in identifying useful and important index terms for their image collections at a relatively low cost (Golbeck, Koepfler, &

Emmerling, 2011), these results suggest that at the broader level of semantic meaning represented by the broader classes of terms, the results are more variable when tags are evaluated in different tasks, such as search, query, and evaluation.

Discussion and Implications

This research project gathered data designed to shed light on semantic and syntactic aspects of tags in relation to controlled vocabulary. This research evaluated participant-generated tags on these aspects using both quantitative and qualitative measures and gathered data on user perceptions of quality of indexing terms generated from a variety of sources.

There is a significant body of literature that defines the indexing quality measures of textual items, including completeness, precision, exhaustiveness, specificity, coherence, and the structure of terms (Cleverdon, 1997; Rolling, 1981; Soergel, 1994). The findings of this study indicate that several quality criteria of image indexing are different from those of textual document indexing, such as context, moods, importance, and level of detail. The differences in quality criteria of image indexing could be explained by the fact that image content is not linguistic; the process of image indexing requires translating sensory input into socially defined and culturally justified linguistic labels and identifiers (Heidorn, 1999; Rasmussen, 1997). Yet it appears that providing access to other criteria dealing with emotive and affective aspects of images (Jörgensen; 1999; Schmidt and Stock, 2007; Liu, Dellandréa, Tellez, & Chen, 2011), as well as their context or "story," would broaden access to images to multiple communities that are searching for an image

that conveys more than a narrower interpretation of what composes "information" in more traditional bibliographic and image metadata.

The findings of the study suggest that while in some instances users might assign general categories of terms that they perceive most useful first, order is not necessarily an indicator of perceived utility. A number of factors enter in to image understanding, including complex (and little understood) interactions among perceptual features and cognitive categories culminating in the "Gestalt" of an image (Palmer, 1992), as well as task influence. Other researchers have suggested the importance of the Gestalt concept of figure and ground (or foreground and background) as a broader organizing principle for what is described when participants view images or scenes and have theorized there may be differences in how images are interpreted, with European cultures placing more emphasis on the main object or foreground, while Asian cultures interpret an image more holistically (Masuda & Nisbett, 2001; Boduroglu, Shah, & Nisbett, 2009; Chua, Boland, & Nisbett, 2005; Dong & Fu, 2010).

The study suggests that, while not all audiences perceive tagging to be a useful adjunct to controlled vocabulary within specific domains (e.g. MESH), tagging can provide access points which may be particularly applicable to documents with which users tend to engage on an interpretive, creative level, or emotive level (Schmitz, 2006; Namaan, Harada, Wang, Garcia-Molina, & Paepcke, 2004) and, in fact, can stimulate users to engage with images and participate in the construction of shared meaning or enliven the interaction with images by presenting differing ideas and viewpoints drawing upon differing levels of expertise or cultural or local knowledge. The financial aspects of using

images which are emotive, or through which users are stimulated to create their own meaning or connection, are well-understood by the producers of advertising, mass media, and online commercial image banks, but there are very few image indexing vocabularies or computer algorithms that could provide more than the most limited access to such materials. Images have shared the fate of other categories of documents that are less frequently indexed, such as fiction (Beghtol, 1991), through which a reader engages in a mediated and ongoing process of interpretation. As the Internet has "democratized" the production of texts, so too has it made available many more formats than were previously available and a variety of ways that users can interact with images, beyond simply posting them (e.g. Pinterest). There has even been a new term coined to describe these interactions: "produsage" (Peters & Stock, 2007; Bruns, 2012). Thus it is not surprising that users are also both creating and demanding new tools and representations for organizing and providing access to a wide variety of documents in a staggering array of subjects.

A recent semiotic analysis of Flickr's "all-time popular" (ATP) tags (Archer, 2010) adds evidence to the results of this research which indicate that tags represent a wide range of image attributes, some of which are not found in controlled vocabularies or are considered too subjective to address. Six connotative groupings of tags within the 144 ATP tags are described. The most visually dominant tags refer to events (e.g. Christmas, football, vacation, concert, party); these are part of the "Story" class in the current research schema. Another group refers to subject or topic (and includes Objects in the current research, which Archer (2010) notes as representing relational and hierarchical

ways of thinking. The most dominant by frequency includes location tags ("Setting" in the current research, part of the "Story" class). Two other groups are temporal tags (including seasons, again part of the "Story" class), and descriptive terms, including color. The final class is the production tags, referring to format and technology used to produce the image. Again, even though there is not a direct mapping between the connotative groups and the classes used in the current research, the semiotic analysis suggests a strong presence of "Story" and "Object" terms, as well as "Setting" and descriptive terms.

Although not the direct products of this research, there are several observations relating to the methodologies widely used in tagging research and the methods of indexing/tagging and representation that can be made based on the experiences of the researchers in interacting with this body of data. The first is the difficulty, and even danger, of analyzing image tags without their context, in other words in isolation from the actual image, as is often done in transaction log analysis. In coding tags or terms, the researchers found that there can be considerable ambiguity and error in taking a tag at its face value. For instance, when the tag "elbow" appeared, the first thought is of a body part; upon viewing the image, the researcher saw that there were clearly two types of elbows pictured, that of a human, and a very prominent pipe elbow, making the tag ambiguous. The same type of ambiguity occurred with names ("Black" could refer to a person or a color). There were a number of instances where the correct interpretation of a tag could only be made by looking at the image, and doing so greatly reduced the number of unknown terms in this set of coding.

The same kind of observation can be made with foreign language terms, which are usually not coded. However, Flickr is being used worldwide, and for this research a variety of translators, as well as Wikipedia pages, were used to assist in interpretation (again aided by being able to view the specific image as well). For the 0.05% (N=1922) sample of all of the LC tags there were nineteen different languages represented (as well as non-western characters) including Esperanto, Welsh, Portuguese, and Hebrew. LC tags for the ten images used in current research added a twentieth, Japanese. As a result of this effort, for this research the percentage of "unknown" terms was much lower than reported in other studies (1.5%). Of the foreign language terms (not counting named people), the majority were name places, including local names and small villages. Objects followed next, with the rest scattered among events, activities, descriptors, and emotion. These terms could be a highly useful resource for cross-language retrieval; however, some tagging systems limit the number of tags (Flickr places a limit of 75), which could constrain this utility, especially in relation to object searches.

Limitations

One of the limitations of this study is that participants were self-selected from a single academic department. Replicating the experiment with a larger and more representative sample of participants and a larger variety of images (in addition to photographs) would be desirable to strengthen the findings of the study. Additionally, the set of data gathered and analyzed for this research is quite large. These results, while robust, are preliminary in that they do not report on interactions among the different types of data and type of task, a potential question for future research. Term consistency across

taggers was not evaluated, although "eyeballing" the data appears to confirm a high level of consistency. Further analysis will need to confirm this.

In addition to the coding scheme used for this research (Jörgensen, 1995) there are several other coding schemes for tags/descriptors that have been used in previous research (Shatford-Layne, 1994; Jorgensen et al., 2002), as well as combinations of several of these (Conduit and Rafferty, 2007). All of these contribute slightly different aspects to the process of making sense and organizing image tags/descriptors/attributes and in practice all have their limitations. The 1995 coding scheme used here was not developed to create a specific scheme for images, but rather to empirically establish the *range* of attributes necessary for full description of the content of an image. As such, it does not address information external to the image (such as provenance and other concerns of records and archives managers), and it does not specifically address the generic, specific, and abstract levels or the major facets of who, what, when, and where of Shatford-Layne (1992). More recent work attempting to integrate these different schemes for further research (Conduit and Rafferty, 2007) encountered difficulties. The 1995 scheme, composed of 10 descriptive classes and 39 attributes, was used in this research to continue to provide a comparable baseline of data gathered across a long time span, as it has been used a number of times with a variety of images, participants, and tasks and adequately covers the range of image content as described by participants. To be usefully applied as a descriptive template is an area of future research.

Conclusion

The current study has undertaken an intensive look at participant tagging behavior of a set of ten historical photographs in the Library of Congress's photostream on Flickr. Participants completed several tasks: a describing task, an evaluation task, and a query task with these images. In addition, as a baseline a random sample of image tags from the LC photostream and the complete set of existing tags for the ten images were analyzed for semantic meaning by using the same coding scheme used to code participant tags. The data were coded for their semantic categories, with all participant tags analyzed for both syntactic (term position, term length) and semantic relationships. Consistent with prior findings, Objects, People, and Setting (location) were most frequently described by tags. The evaluation task suggested that frequency alone cannot be taken as an indicator of importance, as other types of tags (Abstract, Group, the "Story") were seen by participants to have potential value as well.

From the research presented here and evidence presented by other researchers, one could also draw some tentative conclusions about the contributions that tags – and taggers – can make to image (and other document) access. As participants in this research demonstrated, for images they tag Objects frequently and often list Objects in queries, even though they did not rate Objects as highly as some other kinds of attributes in the evaluation task. Thus, taggers themselves can provide multiple "entry terms" (as they are called in controlled vocabularies) for a visual item (e.g. auto, automobile, car) enabling people to find such an image even if they do not know or think of the "correct" term. It also appears that for complex images taggers are willing to add a larger number of tags

(including Object terms) than is practical in a work environment. In the same vein, foreign language terms are also added as tags; these could be useful for cross-language retrieval (e.g. apple: Apfel, pomme, appel, manzana, and so forth).

As taggers engage with digital collections, they can contribute other types of information to these collections; taggers represent a variety of communities with varying levels of expertise – expertise that is needed to provide a full description of the image. As Fry (2007) notes with historical costume terminology, "torquetum" is a rare but essential term for a particular type of query, suggesting that even tags which may occur only once are important. In the LC Flickr collection there were several examples where an image was tagged with a location that was not correct; this stimulated discussion of changing historical boundaries among taggers and resulted in a correction to the image. Previously, Winget (2011) noted that tags contribute local geographic knowledge in providing details about place names and phenomena such as volcanoes and can add additional knowledge that cannot easily be found elsewhere. The LC report (2008) describes invaluable contributions made by taggers who are expert in local history or historical research methods. The research done to date suggests that tagging can bring added value to digital collections at low cost and can increase use of (and support for the existence of) digital collections by providing a wider variety of terminology, including tags that appeal to end users and tags reflecting detailed expertise and subject knowledge. While tagging may not be appropriate in specialized domains where a specific level of expertise and training is required (e.g., medical images), it has the benefits of engaging users in interaction with

collections, increasing awareness of collections, and adding a variety of types of information to collections, especially local knowledge and disciplinary expertise.

However, the current research also demonstrated that users value the information provided by more authoritative approaches such as classification and controlled vocabularies and that they place a level of trust in these, especially where they do not know the appropriate terminology (Jörgensen, Stvilia, and Wu, 2011). Rafferty (2011) points out that information loss occurs when specific historical context or vernacular language is not preserved in these within records. What we also know is that, as time moves on, culturally important information that resides within individuals and cultures is being irretrievably lost, information that can be captured if systems are opened to more user contributions (Jörgensen, 2004).

One area that can benefit from further research is the development of a simple descriptive metadata structure that would reduce the ambiguity found in tags. A large collection of annotated images to test computer methods for object and scene recognition has been a goal of the computer science community for well over a decade (Clough, Müller, & Sanderson 2010; Loy & Eklundh, 2005) and it now appears that image tagging is partially fulfilling this need (Escalante et al. 2010; Chua et al. 2009). Evidence suggests that if end-users are presented with a simplified structure in which to put index terms or tags of their own they can do so (Hollink, Schreiber, Wieling & Worring, 2004; Jorgensen, 1996). However, both the library and information science community and the computer science community also express a need to “tame” tags. Two complementary resources are needed to do this: 1) a simplified hierarchical vocabulary more relevant to

images versus the complexity and abstraction of WordNet, a widely used tool for ontology creation (e.g. Alves, & Santanchè, 2013), and 2) an easy to understand and use metadata structure. Ideally, these would be created in tandem, and they represent rich challenges for the research communities involved.

At the institutional level, the challenge for tagging, in contrast to other widely-used formalized systems, is that there may be no one "right" way to implement it across all collections. As institutions implement tagging in their collections they need to share their successes and failures, challenges and benefits, and learning experiences with each other; no doubt there will be a variety of appropriate ways that tagging systems can work for institutions and their various audiences (Feinberg, 2006; Fry 2007). We may be at the point where we actually do understand enough about these two widely different approaches to describing materials to allow them to peacefully coexist and to do what each does best.

Acknowledgements

This research is partially supported by an OCLC/ALISE Research Grant, 2010. The article reflects the findings, and conclusions of the authors, and do not necessarily reflect the views of OCLC or ALISE.

References

Albrechtsen, H. (1998). The order of catalogues: Towards democratic classification and indexing in public libraries. Proceedings. IFLA Council and General Conference 63, Copenhagen, Danemark, 27(2), pp. 41-43.

Alves, H. & A. Santanchè (2013). Folksonomized ontology and the 3E steps technique to support ontology evolvement. Web Semantics: Science, Services and Agents on the World Wide Web 18(1), 19-30.

Angus, A., Thelwall, M., & Stuart, D. (2008). General patterns of tag usage among university groups in Flickr. Online Information Review 32(1), 89-101.

Archer, J. (2010). Reading clouds: An analysis of group tag clouds on Flickr. (Unpublished master's thesis). Wake Forest University, Winston-Salem, North Carolina.

Bailey, K. (1994). Methods of social research (4th ed.). New York, NY: The Free Press.

Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y., & Shachak, A. (2008). Structured versus unstructured tagging: a case study. Online Information Review, 32(5), 635-647.

Beghtol, C. (1991). The classification of fiction. In American Society for Information Science SIG/CR News (Special Interest Group/Classification Research News), 3-4.

Bischoff, K., Firan, C. S., Nejdl, W., & Paiu, R. (2008, October). Can all tags be used for search? In Proceedings of the 17th ACM conference on Information and knowledge management (pp. 193-202). New York: ACM Press.

Boduroglu, A., Shah, P., & Nisbett, R. E. (2009). Cultural differences in allocation of attention in visual information processing. Journal of Cross-Cultural Psychology, 40(3), 349-360.

Brown, P. & Hidderley, G.R. Capturing Iconology: a study in retrieval modelling and image indexing. *Proceedings of the 2nd International Elvira Conference*, De Montfort University, May 1995. ASLIB 1995 pp. 79-91.

Bruns, A. (2012). Reconciling community and commerce? Collaboration between produsage communities and commercial operators. *Information, Communication & Society,* 15(6).

Chen, H., Schatz, B., Yim, T., & Fye, D. (1995). Automatic thesaurus generation for an electronic community system. Journal of the American Society for Information Science, 46(3), 175-193.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences of the United States of America, 102(35), 12629-12633.

Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z., & Zheng, Y. (2009). NUS-WIDE: A real-world web image database from National University of Singapore. In Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR '09) (pp. 48:1-48:9). New York: ACM Press.

Chung, E., & Yoon, J. (2008). A categorical comparison between user-supplied tags and web search queries for images. Proceedings of the American Society for Information Science and Technology, 45(1), 1-3.

Cleverdon, C. (1997). The Cranfield Tests on index language devices. In K. Sparck Jones & P. Willet (Eds.), Readings in information retrieval. Morgan Kaufmann multimedia information and systems series (pp. 47-59). San Francisco, CA: Morgan Kaufmann.

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

Cunningham, S., & Masoodian, M. (2006). Looking for a picture: An analysis of everyday image information searching. In G. Marchionini & M. Nelson (Eds.), Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries (Vol. 3644, pp. 198-199). New York: ACM. doi:10.1145/1141753.1141797

Dong, W. & Fu, W.-T. (2010). Cultural difference in image tagging. In Proceedings of the ACM 28th International Conference on Computer-Human Interaction (CHI 2010), Atlanta, GA, USA, (pp. 981-984).

Escalante, H. J., Hernández, C. A., Gonzalez, J. A., López-López, A., Montes, M., Morales, E. F., ... & Grubinger, M. (2010). The segmented and annotated IAPR TC-12 benchmark. Computer Vision and Image Understanding, 114(4), 419-428.

Feinberg, M. (2011). An examination of authority in social classification systems. Advances in Classification Research Online, 17(1), 1-11.

Fry, E. (2007). Of torquetums, flute cases, and puff sleeves: A study in folksonomic and expert image tagging. Art Documentation (26) I, 21-27.

Geffet, M., Bar-Ilan, J., Miller, Y., & Shoham, S. (2010). A generic framework for collaborative multi-perspective ontology acquisition. Online Information Review (34) 1, 145-159.

Golbeck, J., Koepfler, J., & Emmerling, B. (2011). An experimental study of social tagging behavior and image content. Journal of the American Society for Information Science and Technology, 62(9), 1750-1760.

L. Hollink, Schreiber, A., Wielinga, B.J. &M. Worring (2004). Classification of user image descriptions. International Journal of Human-Computer Studies, 61(5), 601-626.

Heidorn, B. (1999). Image retrieval as linguistic and nonlinguistic visual model matching. Library Trends, 48(2), 303-325.

Jörgensen, C. (1994). The applicability of existing classification systems to image indexing: A selected review. In R. Green (Ed.), Advances in Knowledge Organization 5, (Proceedings of the Fourth International ISKO Conference), Frankfurt/Main: Indeks Verlag, 189-197.

Jörgensen, C. (1995). Image attributes: An investigation (Unpublished doctoral dissertation). Syracuse University, Syracuse, NY.

Jorgensen, C. (1996). Indexing images: Testing an image description template. In ASIS'96: Proceedings of the 59th ASIS Annual Meeting, 33, 209.

Jörgensen, C. (1998). Attributes of image images in describing tasks. Information Processing and Management, 34(2/3), 161-174. doi:10.1016/S0306-4573(97)00077-0

Jörgensen, C. (1999). Retrieving the unretrievable: Art, aesthetics, and emotion in image retrieval systems. In B. E. Rogowitz & T. N. Pappas (Eds.), Electronic Imaging ‘99 (pp. 348-355). Bellingham, WA: International Society for Optics and Photonics.

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

Jörgensen, C. (2004). Unlocking the museum–A manifesto (2004). Journal of the American Society for Information Science and Technology 55(5), 462-464.

Jörgensen, C. (2007). Image access, the semantic gap, and social tagging as a paradigm shift. In J. Lussky (Ed.), Advances in Classification Research Online.

Jörgensen, C, Stvilia, B., & Jörgensen, P. (2008). Is there a role for controlled vocabulary in taming tags? In J. Lussky (Ed.), Advances in Classification Research Online.

Kakali, C., & Papatheodorou, C. (2010, February). Could social tags enrich the library subject index? In Proceedings of the International Conference Libraries in the Digital Age (LIDA 2010) (pp. 24-28).

Liu, N., Dellandréa, E., Tellez, B. & Chen, L. (2011). Associating textual features with visual ones to improve affective image classification. Affective Computing and Intelligent Interaction (ACII 2011). Lecture Notes in Computer Science (6974), 195-204.

Liu, D., Hua, X.-S., & Zhang, H.-J. (2011). Content-based tag processing for Internet social images. Multimedia Tools and Applications 51(2), 723-738.

Loy, G., & Eklundh, J. O. (2005). A review of benchmarking content based image retrieval. In Workshop on Image and Video Retrieval Evaluation, The Fourth International Conference on Image and Video Retrieval (CIVR 2005), Singapore. Lecture Notes in Computer Science (3568).

Mai, J.-E. (2007). Trusting tags, terms, and recommendations. Proceedings of the Seventh International Conference on Conceptions of Library and Information Science (Unity in diversity). Information Research 15(3). Available at http://informationr.net/ir/15-3/colis7/colis705.html

Manning, C., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Masuda, T. & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. Journal of Personality and Social Psychology (81) 5, 922-934.

Matusiak, K. (2006). Towards user-centered indexing in digital image collections. OCLC Systems & Services: International Digital Library Perspectives, 22, 283-298. doi:10.1108/10650750610706998

Munk, T., & Mork, K., (2007), Folksonomy, the power law & the significance of the least effort. Knowledge Organization, 34(1), pp.16-33.

Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H., & Paepcke, A. (2004, October). Context data in geo-referenced digital photo collections. In Proceedings of the twelfth ACM international conference on Multimedia (pp. 196-203). New York: ACM.

Palmer, S. E. (1992). Modern theories of gestalt perception. In G. W. Humphreys (Ed.), Understanding vision: An interdisciplinary perspective. Cambridge MA: Blackwell.

Petek, M. (2012). Comparing user-generated and librarian-generated metadata on digital images. OCLC Systems & Services: International Digital Library Perspectives 28(2), 101-111.

Peters, I. & Stock, W.G. (2007), Folksonomy and information retrieval. Proceedings of the American Society for Information Science and Technology, 44 (1), 1-28.

Rafiee, G., Dlay, S.S., & Woo, W.L. (2010, July). A review of content-based image retrieval. Proceedings of the 7th International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP) (pp. 775-779, 21-23). Los Alamitos, CA: IEEE.

Rafferty, P. (2011). Informative tagging of images: The importance of modality in interpretation. Knowledge Organization 38(4), 283-298.

Rafferty, P., & Hidderley, R. (2007, December). Flickr and democratic indexing: dialogic approaches to indexing. In Aslib Proceedings (59)4/5, pp. 397-410. Bingley, UK: Emerald Group Publishing Limited.

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

Rasmussen, E. M. (1997). Indexing images. In M.E. Williams (Ed.), Annual review of information science and technology (Vol. 32, pp. 169-196). Medford, NJ: Learned Information.

Redden, C. S. (2010). Social bookmarking in academic libraries: Trends and applications. The Journal of Academic Librarianship, 36(3), 219-227.

Rolla, P. J. (2009). User tags versus subject headings: Can user-supplied data improve subject access to library collections? Library Resources and Technical Services, 53(3), 174-184.

Rolling, L. (1981). Indexing consistency, quality and efficiency. Information Processing & Management, 17(2), 69-76.

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, 59, 1383-1392.

Schmidt, S. & Stock, W. (2009). Collective indexing of emotions in images: A study in emotional information retrieval. Journal of the American Society for Information Science and Technology, 60(5), 863–876.

Schmitz, P. (2006). Inducing ontology from Flickr tags. In Collaborative Web Tagging Workshop (WWW2006), Edinburgh, Scotland (Vol. 50).

Shatford, S., (1986), Analyzing the subject of a picture: A theoretical approach. Cataloging and Classification Quarterly, 6(3), pp. 39-61.

Shatford-Layne, S. (1994). Some issues in the indexing of images. Journal of the American Society for Information Science, 45(8), 583-588.

Sigurbjörnsson, B. & Zwol, R. (2008). Flickr Tag Recommendation based on Collective Knowledge. Proceedings, World Wide Web Conference 2008 (April, Beijing, China), 327-336.

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

Smith, M. K. (2011). Viewer tagging in art museums: Comparisons to concepts and vocabularies of art museum visitors. Advances in Classification Research Online, *17*(1), 1-19.

Soergel, D. (1994). Indexing and retrieval performance: The logical evidence. Journal of the American Society for Information Science, 45(8), 589-599.

Spiteri, L. F. (2010). Incorporating facets into social tagging applications: An analysis of current trends. Cataloging & Classification Quarterly, 48(1), 94-109.

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. Washington, DC: The Library of Congress. Available: http://www.loc.gov/rr/print/flickr_report_final.pdf

Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. Communications of the ACM, 40(5), 103-110.

Stvilia, B., Gasser, L., Twidale, M., & Smith, L. C. (2007). A framework for information quality assessment. Journal of the American Society for Information Science and Technology, 58(12), 1720-1733.

Stvilia, B., & Gasser, L. (2008). Value-based metadata quality assessment. Library and Information Science Research, 30(1), 67-74.

Stvilia, B., & Jörgensen, C. (2009). User-generated collection level metadata in an online photo-sharing system. Library and Information Science Research, 31(1), 54-65.

Stvilia, B. & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. Journal of the American Society for Information Science and Technology, 61(12), 2477–2489.

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

Sun, A., Bhowmick, S., Nguyen, K., & Bai, G. (2011). Tag-based social image retrieval: An empirical evaluation, (62)12, 2364-2381.

Tang, J., Yan, S., Hong, R., Qi, G.-J., & Chua, T.-S. (2009). Inferring semantic concepts from community-contributed images and noisy tags. In Proceedings of the 17th ACM International Conference on Multimedia (MM ’09) (pp. 223-232). New York: ACM Press.

Wetterstrom, M. (2008). The complementarity of tags and LCSH—A tagging experiment and investigation into added value in a New Zealand library context. The New Zealand Library and Information Management Journal (50), 296-310.

Winget, M. (2011). User-defined classification on the online photo sharing site Flickr… Or, how I learned to stop worrying and love the million typing monkeys. Advances in Classification Research Online, 17(1), 1-16.

This is a preprint of an article accepted for publication in Journal of the American Society for Information Science and Technology. Jörgensen, C., Stvilia, B., & Wu, S. (in press, 2013). Assessing the relationships among tag syntax, semantics and perceived usefulness. *Journal of the American Society for Information Science and Technology*.