Emotion-based tags in photographic documents: The interplay of text, image, and social influence / Les étiquettes basées sur des émotions dans les documents photographiques: l'interaction entre le texte, l'image et l'influence sociale

Diane M. Neal

# Emotion-based tags in photographic documents: The interplay of text, image, and social influence

# Les étiquettes basées sur des émotions dans les documents photographiques : l'interaction entre le texte, l'image et l'influence sociale

Diane M. Neal
Faculty of Information and Media Studies
University of Western Ontario
London, Ontario N6A 5B7
dneal2@uwo.ca

Résumé : Cette étude examine le rôle communicationnel joué par le texte, l'image et l'interaction sociale dans des documents photographiques de Flickr portant une étiquette basée sur des émotions et classés selon leur pertinence forte ou faible. À l'aide de l'analyse du discours, il a été possible d'identifier des thèmes textuels et visuels susceptibles de transmettre une signification émotionnelle. Des mesures non-paramétriques ont révélé des différences statistiquement significatives entre les images de forte et de faible pertinence, entre le nombre de vues et le nombre de favoris pour chaque émotion, et entre les images portant des étiquettes émotionnelles différentes.

Mots-clés : photographies, extraction d'images, moteurs de recherche, étiquetage, folksonomie, signification, communication, émotion, documents photographiques

Abstract: This study investigated the communicative roles played by the text, image, and social interaction in high- and low-relevance ranked Flickr photographic documents with an emotion-based tag. Using discourse analysis, textual and visual themes regarding the conveyance of emotional meaning were identified. Non-parametric measures found statistically significant differences between most relevant and least relevant pictures, between the number of views and number of favourites within each emotion, and between pictures with different emotion-based tags.

Keywords: photographs, image retrieval, search engines, tagging, folksonomy, meaning, communication, emotion, photographic documents

## Introduction and literature review

Regardless of whether one believes that the collective intelligence of the many is more reliable than the knowledge of a select few (Surowiecki 2004), the presence of user-generated content on the current instantiation of the World Wide Web is undeniable. User-generated content, and the corresponding utilization of collective intelligence, are major tenets of "Web 2.0" (O'Reilly 2005). Web 2.0 principles also allow users to interactively access and manipulate data on the website, communicate with one another, and, in many cases, contribute to the site's continued development. Current popular Web 2.0 destinations include the social networking websites Facebook, Twitter, and MySpace, the video-sharing site YouTube, the online encyclopedia Wikipedia, the social bookmarking site Delicious, and the photograph-sharing site Flickr. *Time* rated Flickr the number one website on its "50 Best Websites 2009" list, providing the following support:

> Computers don't handle visual imagery with the same native ease with which they parse text or crunch numbers. Flickr was the first site to solve this problem with something called collaborative tagging. The idea is that if everyone is allowed to tag everyone else's uploaded photos, then a rough-and-ready categorization will naturally emerge from the wisdom of the crowd. It works because it has to—there aren't enough librarians in the world to look after Flickr's archive of 3 billion photos, much less file them away for future reference. But it also works because the many really is smarter than the sum of its parts. The Library of Congress has even started to poll the Flickr hive mind when cataloging its own photos. (Fisher 2009)

On many Web 2.0 sites, tags, or free-text keywords that users assign to online documents (such as Flickr photographs), collectively result in a folksonomy (Neal 2007), a term that is a combination of the words *folk* and *taxonomy* (Vander Wal 2007). The subjectivity, inter-indexer inconsistency, and other related issues inherent in describing and indexing *any* document (Wilson 1968)—images in particular (Jörgensen 1998; Krause 1988; Markey 1984; Shatford 1986)—would logically lead one to understand the potential utility of folksonomy as a method of image description and retrieval. In a folksonomy-based setting, people can contribute terms that are meaningful to their personal interactions with pictures. Research has shown that people desire control over the representation of their own photographs, which folksonomies allow them to do (Neal 2006). On the basis of the popularity of tagging, it could be assumed that many people enjoy tagging, although the debate continues

over whether users tag primarily for their own later retrieval (Golder and Huberman 2006), or for the benefit of all users (Angus, Thelwall, and Stuart 2008). Misspellings and other inconsistencies in users' tagging practices have led some to suggest that a combination of tags and authoritative assistance would improve access to tagged documents (Neal 2008; Rafferty and Hidderly 2007). This philosophy is starting to manifest itself in the Web 2.0 development community, with Wikipedia's recent decision to try adding oversight to user contributions, and with Flickr co-founder Caterina Fake backing its decision: "Without rules like those the site is testing, the encyclopedia would devolve 'into chaos'" (Sutter 2009).

There is a growing body of Flickr-related research in the library and information science literature. Beaudoin (2007) created a category list of Flickr tags, concluding that place names, compound tags, things in the picture, people in the picture, and events in the picture comprised the five most popular categories. Lee and Neal (in press) found similar results by examining the most popular Flickr tags of all time and subsequently testing a hierarchy of tag categories via elicited descriptions. Stvilia and Jörgensen (2009) examined tags at Flickr's photoset (collection) level and found that users more commonly tag at the photoset level than at the individual image level. Ding et al. (2009) concluded that relatively few users tag on Flickr and that strong user interest lies in the processes of photograph commenting and sharing. Cox's (2008) remarkably insightful critique of Flickr from a social and leisure perspective defines the site's larger context as "the existence of an alternative social world governed by a different, more artistic ideology" (499).

In almost all cases, digital photographs are searched and retrieved online via text-based keywords, whether through the use of tags or controlled vocabularies such as the *Art and Architecture Thesaurus* (AAT) or Library of Congress Subject Headings (LCSH). Controlled vocabularies can guide users in selecting search terms but have many drawbacks, such as the use of narrow, expert-oriented vocabularies and inflexible, pre-coordinated terms (Jörgensen 2003). Additionally, as Svenonius (1994) stated, "There are instances where a message expressed in one medium cannot adequately be transposed to another" (600). Similarly, O'Connor posited that semantic meaning gets lost in the translation between a picture and its textual representation (Greisdorf and O'Connor 2008; O'Connor and Wyatt 2004). It is difficult to argue against their assertions. For exam-

ple, Barthes's *punctum*, or something that particularly grabs a person's interest in certain photographs, cannot be communicated in words (Barthes 1981).

Unfortunately, however, attempts at describing pictures using other pictures as descriptive surrogates have been mostly unsuccessful, or perhaps unpopular at best (Enser 1991; Rorvig 1986; Rorvig, Turner, and Moncada 1999); this area of research remains open for exploration. At the same time, the textual description that contextualizes a photograph (for example, it might explain where, when, and why it was taken) provides richness to the viewer's experience of a photograph. Additionally, Flickr's prominent social aspect supplies another dimension to the site's user experience. Visitors to a picture's Flickr page can add tags and comments; they can "favourite" it, they can invite the photographer to add it to a collection of specialized or outstanding photographs, and so on. Photographers can join groups of interest, elicit feedback on their pictures from visitors, and monitor statistics on the number of times their pictures have been viewed, and many other things. The combination of visual, textual, and social elements present on a page for a Flickr photograph will be referred to as a "photographic document" in this article. The use of this combination stems from Barthes (1964), in an essay in which he discussed the denotational (concrete) and connotational (abstract) linguistic messages of texts linked with images. He described the elucidating effect that words have on their associated image: "From the moment of the appearance of the book, the linking of text and image is frequent" (38).

The role of affect in image retrieval can be explored through theories of image description. Individual affect could be considered an iconological concern within Panofsky's (1955) three hierarchical levels of meaning in visual works. *Pre-iconographical description* is the first level, which addresses items physically present and easily identifiable in a visual document, such as a person. *Iconographical analysis* is the second level of meaning; this requires knowledge acquired through literary education; an iconographical description might take the "person" analysis a step further to identify the subject of a visual document as Jesus. Finally, the *iconological interpretation* of a visual document refers to the intrinsic or symbolic meaning of a document. As Shatford (1986) noted, the iconological level of description is the most difficult to index, for it can vary among individuals; for example, a painting of Jesus might symbolize

freedom to a Christian but symbolize oppression to a Muslim. However, as Shatford rightly stated, iconology should not be ignored because of these challenges. Interpreting emotion on the basis of facial expressions is a universal ability and is therefore a simple pre-iconographical task (Panofsky 1955). However, it could certainly be argued that some aspects of emotional description fall within the realm of iconology if the emotion attached to it is symbolic in association.

A challenge, however, is the highly differentiated response to pictures: "We cannot state that all people feel the same when they tag an image with, e.g., *disgust*" (Schmidt and Stock 2009). Similarly, perhaps not everyone will feel disgusted upon viewing a prototypically "disgusting" photograph. In Greisdorf and O'Connor (2002), participants described how they thought a supplied list of terms applied to a supplied list of images, and also provided their own free-text terms that they felt described the images. From this, the authors found that emotion-based queries may be important to facilitate image retrieval, despite inherent challenges related to individual subjectivity. Ornager (1997) also concluded that pictures must be indexed with objective and subjective terms for satisfactory search and retrieval.

Lee and Neal (2007) developed a framework and methodology for determining users' emotional responses to music. They used Power's (2006) model of basic emotions, which includes anger, fear, happiness, sadness, and disgust, for study participants to use in rating the level of each emotion that each piece of music caused them to feel. Lee and Neal discovered that because there is a wide range of responses, a Web 2.0–style system for an emotion-based music information retrieval system would be beneficial. This finding was validated in Neal et al. (2009). Schmidt and Stock (2009) followed Lee and Neal's methodology when investigating the use of tagging for emotion in images, and they found "prototypical images for given emotions" (863). Additionally, they found that many text-based descriptors could be categorized at the basic level of description (Rosch et al. 1976), or the level of categorical detail at which people tend to describe objects most often. In recent image-retrieval research, the "basic level" has been suggested as a potentially consistent phenomenon for preferred photograph description (Rorissa 2008; Rorissa and Iyer 2008; Yoon 2009). In a sense, the current study seeks to combine these two lines of research, using the basic level of pictures as a method of understanding their potential affective aspects.

The concept of relevance is important in this study, particularly because of the method of sample selection. Flickr's "most relevant" and "least relevant" pictures that were returned in tag searches for Power's five basic emotions are the subject of this study. The notion of relevance is still the subject of much debate and research in library and information science. From the perspective of an information retrieval system, a relevant document matches the user's query. In a systems-oriented, mathematical context of relevance, "precision" measures the number of retrieved documents that are deemed relevant, and "recall" defines the number of relevant documents that are retrieved based on all documents in the system (Korfhage 1997). However, many theorists maintain that relevance is human oriented, not systems oriented (Borlund 2003; Saracevic 2007; Schamber, Eisenberg, and Nilan 1990; Xu and Chen 2006). Within the user-centred view of relevance, many models for criteria and relative importance of individual relevance components can be found. For example, Wilson (1973) theorized that information is situationally relevant to a person when the information answers "questions of concern" (463) to that person. In more recent work, Greisdorf (2003) created a continuum of "relevance thresholds" based on participants' determinations of relevance. His model starts at users' notions of a document's topicality, progresses to pertinence, and ends at functionality. Xu and Chen's (2006) research concluded that topicality and novelty were more important relevance criteria than understandability, reliability, and scope.

Somewhat removed from these theoretical discussions, Flickr's relevance-ranking approach is mysterious. In 2007, a user posted a question to Flickr's Help Forum asking about group searching and Flickr's relevance definitions. Kevin, a Flickr staff member, responded in part, "We use traditional text search for keywords to determine relevance. Group keywords, group description and group title all figure into it. It's not working as well as we'd like, and improving the relevance is on the To Do list" (Flickr 2007). No newer official discussion of Flickr's relevance ranking could be located. Despite their need to improve the relevance ranking system, it is currently one of three prominent methods by which to sort search results. The other two sorting methods are "recent," which sorts by descending date, and "interesting." Flickr lists several factors in its "interestingness" ranking, including "where the clickthroughs are coming from; who comments on it and when; who marks it as a favorite; its tags and many more things which are constantly changing. Interest-

ingness changes over time, as more and more fantastic content and stories are added to Flickr" (Flickr 2009). While the nature of this measurement seems intriguing from a research perspective, extreme high and low relevance ranking was used to choose the purposeful sample of photographic documents used in this study, because relevance ranking is a widely known qualitative and quantitative measurement of the success or failure of an information retrieval system.

This study advances a line of research that seeks to improve emotion-based access to digital photographs. Prior research, as described above, demonstrates the indisputable popularity of user-generated and social photographic content, people's desire to describe and access photographs by subjective descriptions such as emotion, and the contentious nature of defining relevance for any one individual or document. Given these factors, it is important to first understand the interplay of factors that create and communicate people's notions of emotion within photographs, so eventual system elements can be designed accordingly. This summarizes the holistic goal of the research described in this article.

## Research questions

In order to examine the individual contributions and interactions among the textual, visual, and social elements of photographic documents on Flickr, as well as their relationships to relevance ranking, two research questions were formulated.

1.  Among high- and low-relevance ranked Flickr photographic documents with an emotion-based tag, what textual and visual elements communicate the emotion represented by that tag?
2.  Among high- and low-relevance ranked Flickr photographic documents with an emotion-based tag, what statistical relationships exist among the number of views, the number of favourites, relevance rankings, and the tagged emotion?

## Methodology

The sample of photographic documents used in the study was extracted using Flickr's flickr.photos.search Application Programming Interface (API) call. The API calls were configured to sort by relevance, to retrieve

photographs only, and to search by the specified tag. For consistency, adjective forms were selected for the tags on which to search (angry, sad, happy, afraid, and disgusting). The XML that was extracted from the API calls was imported into a text editor. The Extensible Markup Language (XML) provided essential identifying information about each photograph, such as the author ID, the photo ID, the picture's title, and other metadata. In order to gain a broader sample of various photographers' descriptive and visual styles, multiple entries from the same photographer were eliminated within each tag's result set. Thirty pictures in each of the following ten categories were identified, for a total of 300 photographs: angry (most relevant), angry (least relevant), sad (most relevant), sad (least relevant), happy (most relevant), happy (least relevant), afraid (most relevant), afraid (least relevant), disgusting (most relevant), and disgusting (least relevant). The sample size was chosen on the basis of a balance between the researcher's time availability and the estimated amount of time needed to locate the photographs and perform the analysis. The most relevant and least relevant pictures for each tag were chosen in order to observe whether there were statistically significant differences between them. For the quantitative portion of this study, it was hypothesized that (1) there are statistically significant differences between most relevant and least relevant pictures, (2) there are statistically significant differences among the pictures as categorized by their respective emotions, and (3) there are statistically significant differences for the number of views and the number of favourites among the emotions.

Because the flickr.photos.search API call does not provide Uniform Resource Locators (URLs) with which to access Flickr photographs, Flickr's end user search interface was employed to locate each photograph's page on the website. As each picture was located, brief metadata about it were recorded into a common spreadsheet, including a meaningful unique identifier, the photographic document URL, title, number of views, and the number of times it had been "favourited" by Flickr users. Completing this process afforded a first glimpse into the visual, textual, and social content of the photographic document. A sample spreadsheet entry for one photographic document can be seen in figure 1, and the corresponding photograph can be viewed in figure 2.

After all images had been located and recorded in the spreadsheet, considerable additional time was spent analyzing the photographic docu-

| # | URL | title | events/things in the picture | what makes it emotional? | date shot | # views | # favorites |
|---|---|---|---|---|---|---|---|
| DL30 | http://www.flickr.com/photos/lizzon/459404487/ | hanging out at Angela's place | woman cooking | she is making Hamburger Helper; photographer thinks it's disgusting | 31/3/07 | 115 | 0 |

*Figure 1:* Example of spreadsheet entry for one photograph

ments. At this time, two columns were added to the spreadsheet for each image. The first column included a pre-iconographical description of events and significant items in the image, adhering to the image's basic level in an attempt to best replicate how users might most frequently identify its facets. This description was written without regard to the associated textual elements. The additional column indicated what elements of the picture likely caused a photographer or viewer to tag it with the emotion in question; this choice was determined by viewing the picture in combination with reading the title, other tags, the photographer's detailed description, and viewer comments. Tremendous care was taken to make the notes as objective as possible.

Using discourse analysis (Beck and Manuel 2008; Rose 2007; Wildemuth and Perryman 2009), the images, their associated textual elements, and the contents of the spreadsheet were extensively evaluated qualitatively and iteratively by the researcher. Rose (2007), whose book describes interpretation methods for visual works, defines *discourse* as "a particular knowledge about the world which shapes how the world is understood and how things are done in it" (142). According to Rose, discourses are socially produced, and images are largely social. In her discourse analysis typology, "discourse analysis I" focuses specifically on discourse as explicated through images and text as well as "intertextuality," or "the way that the meanings of any one discursive image or text depend not only on that one text or image, but also on the meanings carried out by other images and texts" (142). Since Flickr is a social networking website and therefore socially constructed, it made sense to adapt Rose's method in this study to understand each photograph and its related user-supplied text as one document. Evaluating the intertextuality also afforded progress toward an understanding of how the photographer and public viewers might construct each photographic document's affective meaning within their own worlds. As described in the next section, a set of visual and textual themes emerged from this process, which was meant to identify the textual and visual elements that might motivate Flickr users to supply particular emotion-based tags for their pictures.

*Figure 2:* Photograph corresponding to figure 1 spreadsheet example

Since descriptive statistical measures revealed a widely skewed distribution among all collected quantitative data, non-parametric methods (Vaughan 2001) were employed to answer research question 1. Spearman's rho, the non-parametric version of the Pearson correlation coefficient, checked for statistically significant correlations between the number of

views and the number of favourites within each of the 10 categories. The Mann-Whitney test, equivalent to the parametric independent $t$ test, looked for statistically significant differences between the most relevant and least relevant pictures within each of the five emotion-based Flickr tags utilized in this study, in the context of the number of views as well as in the number of favourites. Finally, the Kruskal-Wallis test, a non-parametric approach similar to the parametric one-way ANOVA test, examined the data for statistically significant differences among emotions, in regard to the number of views as well as the number of favourites.

## Results and discussion

### Research question 2: Visual and textual elements that communicate emotion

The discourse analysis revealed seven visual themes and six textual themes that communicate emotion in the dataset. The visual themes are *facial expression*, *colour*, *light contrast*, *symbolic*, *inanimate observation*, *action*, and *social norms*. *Storytelling*, *jokes*, *inside story*, *text-as-image*, *antithesis*, and *personal opinion* comprise the textual themes. Definitions and examples of the themes can be found in tables 1 and 2. The themes are not necessarily to be applied in a descriptive indexing scheme, but rather are to be used as a typology for the visual and textual communicative methods present in the dataset.

In identifying which themes might communicate the emotion in question for any given picture, the themes significantly overlap. In other words, it is highly likely that more than one theme applies to a photograph. The communicative powers of the visual and textual themes present in each picture enhanced each other. In some cases, however, the interpretation benefits resulting from the visual cues outweighed the textual cues, and vice versa.

Despite a serious risk of overgeneralization, it could be said that the identified textual and visual methods within the photographic documents prevailed within the various emotions, and that might lead to the assignment of certain themes. For example, photographs depicting bodily fluids, toilets, rotting food, and other unpleasant sights appeared often in "disgusting" pictures; these could receive the themes Textual (personal

Table 1. Visual themes resulting from discourse analysis of all 300 photographic documents

| Theme Name | Definition | Examples from the Data |
|---|---|---|
| Facial expression | The subject(s) in the picture "look(s)" like a particular emotion | A frowning child looks "sad." |
| | | People smiling look "happy." |
| Colour | The use of either black and white or other colours in the image to enhance its effect | A woman hiding her face while standing in an empty room, shot in black and white, is "afraid." |
| | | A field containing bright flowers shot on a sunny day is "happy." |
| Light contrast | Light and dark imagery in the photograph | Dark plants and a bright sky appear in a "sad" photo called "A Dark Sad Day Is Brightened by the Sun." |
| | | A dark sky at dawn with the sun peeking through is "happy." |
| Symbolic | The actual content of the photograph stands for something else | A woman's face with a black handprint painted on it represents domestic abuse and is "sad." |
| | | A flying bird symbolizes the photographer's "happy" birthday. |
| Inanimate observation | The presence of something not alive, representing a particular emotion | An abandoned hotel building is "disgusting." |
| | | A small wooden figure with people standing around it represents "afraid." |
| Action | People or animals doing something in the picture | People jumping on the beach are "happy." |
| | | Crying children are "sad." |
| Social norms | The items in the picture evoke an emotion based on what we collectively may think culturally/ socially | Blood in a public washroom sink is "disgusting." |
| | | A photograph of an impoverished person is "sad." |

Table 2. Textual themes resulting from discourse analysis of all 300 photographic documents

| Theme Name | Definition | Examples from the Data |
|---|---|---|
| Storytelling | The emotion is communicated by the story in the photographer's description. | Description explains that the subject in the photo was "angry" when the photographer took the picture, and why. |
| | | Description outlines a bad experience at a restaurant that made the food "disgusting." |
| Jokes | Jokes are communicated with humour-inducing techniques, such as irony and sarcasm. | An egg with a face painted on its shell that is on an egg slicer is "afraid." |
| | | A "happy" picture shows a cat reading a book. |
| Inside story | The reason behind the tag is undetectable by the casual viewer, but likely has an inside story attached to it. | A group of people happily standing under the city limits sign for Truth or Consequences, New Mexico, is "angry." |
| | | A photo taken from the passenger seat of Painted Desert, Arizona, is "angry." |
| Text-as-image | Text has been photographed. | An "afraid" picture has cut-out photos of people with pipe cleaners in the centre; words in the picture say, "In [A]ppledore we burn crooks & wrong doers at the stake!" |
| | | An "angry" picture has the word *poop* spelled out with blocks as well as a book called *Angry Little Girls! A Little Book on Friendship.* |
| Antithesis | The picture shows the opposite of the tagged emotion. | A cat with an angry, irritated expression is in a "happy" photo. |
| | | A picture symbolizing lost love is tagged with "Happy Valentine's Day." |
| Personal opinion | The personal opinion about the subject of the picture, as expressed by the tagger | A photo of a very large fast food hamburger is "disgusting." |
| | | Photographer thought the emotion in her self-portrait looked "sad." |

opinion) as well as Textual (social norms). Many "sad," "angry," and "afraid" photographs displayed people and animals "looking" sad, angry, or afraid, hence the possible assignment of Visual (facial expression). Pictures of the negative emotions also tended to use black-and-white photography to convey mood, so the Visual (colour) theme could be used. Visual (colour) also applied to "happy" pictures, but bright colours were often used in them. Again, Visual (facial expression) was sighted often in the "happy" sample, typically with people laughing and smiling. There was little, if any difference in overarching themes or communicative methods between the most relevant and least relevant photographs for each emotion. It should be noted that this is, obviously, by no means a comprehensive description of the themes' applications; such a list would be beyond the space limitations and scope of this paper.

The process of reviewing the pictures and their associated text brought into play Barthes's theory about text and image working together to convey denotation and connotation, as well as words' elucidation of pictures. In almost all cases, it was a fairly simple task to determine and describe the photograph's denotation (or Panofsky's pre-iconographical level of description). For some pictures, the emotional connotation was easy to determine with a quick glance at the photograph itself (for example, a smiling, jumping person on a beach is "happy," or an old, dirty toilet is "disgusting"). However, it was frequently necessary to examine textual elements of the photographic documents, such as the title, tags, description, and visitor comments, in order to determine the emotional connotation. Pulling an example from the dataset, a close-up photograph of sushi would not necessarily equate with "happy" at most individuals' basic levels of interpretation, but it did to a particular photographer who likes sushi. Additionally, the textual elements helped clarify and expand suspicions that were held after just the inspection of the photograph itself. As an example, it was natural to respond to one picture with the thought, "He looks sad." Afterward, the sadness was confirmed and further clarified after reading that the man is homeless.

*RQ2: Potential social impact on relevance ranking*

As previously mentioned, it was hypothesized that there are statistically significant differences (1) between most relevant and least relevant pictures, (2) among the pictures as categorized by their respective emotions, and (3) for the number of views and the number of favourites among the emotions.

Table 3. Descriptive statistics of number of views and number of favourites, by level of emotion and relevance (N = 30)

| Tag and Category | Minimum Number | Maximum Number | Median |
|---|---|---|---|
| Angry views (most relevant) | 2,378 | 251,007 | 7700.50 |
| Angry favourites (most relevant) | 0 | 2,729 | 142.00 |
| Angry views (least relevant) | 35 | 4,241 | 126.50 |
| Angry favourites (least relevant) | 0 | 5 | 2.50 |
| Sad views (most relevant) | 3,240 | 75,421 | 12391.00 |
| Sad favourites (most relevant) | 0 | 1,440 | 267.00 |
| Sad views (least relevant) | 96 | 1,862 | 390.00 |
| Sad favourites (least relevant) | 3 | 39 | 12.50 |
| Happy views (most relevant) | 1,862 | 96,535 | 14318.50 |
| Happy favourites (most relevant) | 16 | 1,901 | 221.50 |
| Happy views (least relevant) | 129 | 4,426 | 456.00 |
| Happy favourites (least relevant) | 8 | 103 | 23.00 |
| Afraid views (most relevant) | 233 | 19,315 | 2931.00 |
| Afraid favourites (most relevant) | 13 | 580 | 55.00 |
| Afraid views (least relevant) | 2 | 332 | 19.50 |
| Afraid favourites (least relevant) | 0 | 0 | 0.00 |
| Disgusting views (most relevant) | 511 | 109,031 | 3145.00 |
| Disgusting favourites (most relevant) | 5 | 236 | 24.00 |
| Disgusting views (least relevant) | 4 | 237 | 35.00 |
| Disgusting favourites (least relevant) | 0 | 0 | 0.00 |

Descriptive statistics for the dataset appear in table 3. In this table, the lowest number of views, the highest number of views, the lowest number of favourites, and the highest number of favourites for each category are reported. The dataset is skewed, so the medians have been reported instead of the means. For reference, per the API calls sent to Flickr during data collection in early August 2009, Flickr held 757,924 photos tagged with "happy," 147,063 "sad" pictures, 33,729 "angry" photographs, 7,451 "disgusting" pictures, and 3,876 "afraid" photos. The exact number of photographs posted on Flickr is unknown, but a current estimate places the figure at 3 billion photos (Fisher 2009). The number

Table 4. Correlation between number of views and number of favourites (Spearman's rho)

| Category | $r_s$ | Correlation? | 2-tailed *p*-value |
|---|---|---|---|
| Angry (most relevant) | .776 | Yes | .01 |
| Angry (least relevant) | .004 | No | n/a |
| Sad (most relevant) | .723 | Yes | .01 |
| Sad (least relevant) | .259 | No | n/a |
| Happy (most relevant) | .630 | Yes | .01 |
| Happy (least relevant) | .430 | Yes | .05 |
| Afraid (most relevant) | .664 | Yes | .01 |
| Afraid (least relevant) | n/a | No | n/a |
| Disgusting (most relevant) | .147 | No | n/a |
| Disgusting (least relevant) | n/a | No | n/a |

of photographs with the emotion-based tags explored in this study is low compared to the total number of pictures on Flickr, although Kipp (2008) pointed out that many non-emotion-based tags are actually affective. However, it is obvious that "happy" pictures are much more popular than pictures described with negative emotions. One explanation is the theory that positive affect increases levels of dopamine—a neurotransmitter responsible for many brain functions (Ashby, Isen, and Turken 1999).

Table 4 displays the Spearman's rho tests for correlation between the number of views and the number of favourites within each category of pictures. Statistically significant correlations were found within the following categories: Angry (most relevant), Sad (most relevant), Happy (most relevant), Happy (least relevant), and Afraid (most relevant). Except for Disgusting (most relevant), all "most relevant" categories had a statistically significant correlation, suggesting that the higher the relevance of a picture, the more likely it is to be favourited. Happy (least relevant) had a smaller statistically significant correlation, compared to the others. Since no pictures in the Afraid (least relevant) or Disgusting (least relevant) categories had been marked as favourites, the statistic could not be calculated for them.

Table 5 shows the results of the Mann-Whitney *U* test for the number of views, comparing most relevant and least relevant photographs within

Table 5. Difference between number of views, comparing most relevant and least relevant photographs within each emotion (Mann-Whitney)

| Emotion | Mann-Whitney $U$ | 2-Tailed $p$-value |
|---|---|---|
| Angry | 10.000 | .000 |
| Sad | .000 | .000 |
| Happy | 8.000 | .000 |
| Afraid | 1.000 | .000 |
| Disgusting | .000 | .000 |

Table 6. Difference between number of favourites, comparing most relevant and least relevant photographs within each emotion (Mann-Whitney)

| Emotion | Mann-Whitney $U$ | 2-Tailed $p$-Value |
|---|---|---|
| Angry | 28.500 | .000 |
| Sad | 43.500 | .000 |
| Happy | 57.000 | .000 |
| Afraid | .000 | .000 |
| Disgusting | .000 | .000 |

each emotion. This test demonstrated that there is a statistically significant difference for all emotions. In the "sad" and "disgusting" samples, all most relevant pictures held a higher rank than all least relevant pictures, giving those emotions a $U$ of .000. Table 6 lists the same test for the number of favourites. Again, there is a statistically significant difference for all emotions. As mentioned above, no pictures in the Afraid (least relevant) or Disgusting (least relevant) categories had been marked as favourites, so the result of $U$ is .000.

The Kruskal-Wallis test, used to find the difference in the number of views among emotions, can be found in Table 7. Not surprisingly, "happy" pictures had the highest mean rank, followed by "sad," "angry," "disgusting," and "afraid." The results indicate a statistically significant difference between the emotions; $l^2(4) = 43.707$ and $p = .000$ (2-tailed). Similarly, table 8 shows the same test for the difference in the number of views. The mean ranks were almost the same as for the

Table 7.  Difference in number of views among emotions (Kruskal-Wallis)

| Emotion | Mean Rank | $\chi^2$ | df | 2-Tailed p-Value |
|---|---|---|---|---|
| Angry | 157.44 | 43.707 | 4 | .000 |
| Sad | 181.45 | | | |
| Happy | 188.98 | | | |
| Afraid | 106.03 | | | |
| Disgusting | 118.61 | | | |

Table 8.  Difference in number of favourites among emotions (Kruskal-Wallis)

| Emotion | Mean Rank | $\chi^2$ | df | 2-Tailed p-Value |
|---|---|---|---|---|
| Angry | 152.41 | 64.386 | 4 | .000 |
| Sad | 186.53 | | | |
| Happy | 200.83 | | | |
| Afraid | 117.81 | | | |
| Disgusting | 94.93 | | | |

number of views, except that "afraid" was higher than "disgusting." Again, a statistically significant difference between the emotions exists; $\chi^2(4) = 64.386$ and $p = .000$ (2-tailed).

On the basis of statistical tests performed on the current sample, $H_0$ is rejected for all three hypotheses. There are statistically significant differences between most relevant and least relevant pictures, among the pictures as categorized by their respective emotions, and for the number of views and the number of favourites among the emotions. The most relevant pictures possess the highest number of views, as well as the highest number of favourites. It is also clear that there are more "happy" pictures on Flickr than "sad," "angry," "disgusting," or "afraid" pictures, and they generate the highest amount of social interaction, as measured by views and favourites. In response to the second research question, it seems highly plausible that either social interaction affects relevance, or relevance affects social interaction. However, since no current explanation of Flickr relevance could be located, and the two-year-old explanation was unclear regarding relevance calculations (Flickr 2007), it is difficult to answer this research question with confidence. However, it is clear that the most relevant pictures obtain more views and

favourites, and the least relevant pictures obtain many fewer views and favourites.

## Additional findings

As an unexpected corollary to these results, many difficulties with Flickr's search engine were identified while locating the sample pictures on Flickr. For example, it was typically impossible to locate a photograph by its photo ID, a unique number identifying the picture retrieved in the flickr.photos.search API call. In many cases, a Google search by the picture's photo ID located the picture much more easily than a search using Flickr's search engine. What would logically be considered the most "relevant" picture from the standpoint of precision and recall (based on a search for the exact title, for example) was not always in the top-ranked results. Occasionally, the photo being sought did not appear in the search results list individually, but appeared in one of many user-created photo mosaics.[1] The photographic document in question could then be located by clicking on its individual title, typically listed in the mosaic's description. Additionally, the search engine experienced frequent technical problems during data collection, greeting users with the message, "Hold your clicks a moment please ... Flickr has the hiccups. We're looking into the problem right now." This is apparently a frequent problem, as demonstrated in the Flickr Help Forums.[2] Concerns about the precision, recall, and relevance ranking in the search mechanism as well as its technical reliability may give information professionals pause regarding Flickr's status as a feasible information retrieval tool. However, Cox (2008) makes an intriguing point: "It is probable that most users are not searching Flickr with a certain 'information need' in mind. Rather, they are browsing for direct visual pleasure. Precision and recall are largely irrelevant, as a result ... navigation in Flickr is by browsing" (496).

## Conclusions and future research

This study examined image Emotional Information Retrieval by evaluating the interplay among textual denotation and connotation, image meaning, and social participation as measured by views and favourites in highly relevant and highly non-relevant Flickr photographic documents with emotion-based tags. A set of themes was proposed for suggesting the methods in which text and image communicate emotion. Through

these themes, Barthes's theory that words elucidate pictures could be observed. Non-parametric measures demonstrated statistically significant differences between most relevant and least relevant pictures, between the number of views and number of favourites within each emotion, and between pictures with different emotion-based tags. These measures also demonstrated a potential but unexplainable link between social interaction with photographic documents and Flickr's relevance ranking system. Although it was not a planned research question within this study, concerns regarding Flickr's relevance ranking as well as in its search and retrieval mechanisms were noted.

The study's most apparent weaknesses, including the inherent subjectivity of the discourse analysis and Flickr's unclear relevance rankings, are intrinsic to the topic under consideration. To the author's best knowledge, no analysis method yet developed in concept-based image retrieval is completely objective. Stacking the task of discerning emotional content on top of the already-subjective image description process is certainly a cause for experimental concern, but it is also necessary in order to advance the field. This study does not include the analysis of non-emotion-tagged photographic documents that cause emotion in its viewers, which is an area for future consideration: "We cannot state that an image is not emotional-laden if only some (or none) users have tagged it" (Schmidt and Stock 2009, 873).

Additionally, since Flickr's relevance-ranking algorithm is not revealed to the public, it is not possible to compare the relevance-related results of this study with Flickr's relevance determination. Perhaps as a method of closing the knowledge gap created by this missing information, the quantitative portion of this study set out to determine whether the number of views and the number of favourites might play a role in Flickr's determination of relevance, and the findings of this research seem to indicate that they do so. It is highly likely that Flickr search results are also ranked using basic information-retrieval techniques, such as checking for the presence or absence of query terms within the photographic document's textual elements. However, this study's main focus was to explore the interplay of social, textual, and visual elements on emotion identified within each document, not the mechanics of Flickr's search algorithms. As discussed, there are very apparent differences between the result lists of the flickr.photos.search API call and of Flickr's end-user search engine. At the same time, however, without further research, it is not possible to know whether these search engine anomalies are bothersome to Flickr users.

The area of tagging as a phenomenon as well as a topic of research is very new; as a result, any study on it will by definition be exploratory. Therefore, this study cannot draw definitive conclusions, but it can present some suggestions for discussion. A strong implication is the need for the integration of concept-based and content-based image retrieval (e.g., Neumann and Gegenfurtner 2006); unfortunately, researchers in the two camps do not cite each other's work (Enser 2000; Neal 2006; Persson 2002). For example, with effort, algorithms could be developed that guide the user toward "happy" pictures on the basis of their descriptive metadata as well as their perceptual features. Additionally, the utilization of Flickr's highly active social aspects could be utilized more effectively, but more research is needed in order to understand how to leverage them. Currently, users tend to congregate around their chosen groups on the site (Cox 2008), and this study suggested that there is a positive correlation among favourites, views, and relevance. However, these interactions could be better employed in enhancing the user experience in interface and search enhancements still left to be developed. Evaluations of social features such as user comments, patterns of "contact" choices, and "group" behaviours are needed.

Other future research possibilities include running a similar experiment with a larger sample size, and perhaps a random sample should be taken in order to compare results with the purposefully chosen sample in this study. A wider variety of emotion-based tags and a more complex model of emotion could be utilized in future work (Rubin, Stanton, and Liddy 2004). Additionally, to continue the discussion of whether predictable models of traditional precision and recall are truly needed on Flickr, comparison studies of users' browsing behaviours and searching behaviours, both affect-based and non-affect-based, should be conducted. An eventual research goal of the author is to create an intelligent agent-driven, artificial intelligence system that will guide, and be guided by, system users in the identification of related images based on both content- and concept-based affect similarity.

## Acknowledgements

and his thoughtful insights. Finally, the anonymous reviewers provided excellent suggestions for improvement.

## Notes

1. See http://www.flickr.com/photos/grrrl/383175308/ for an example.
2. See http://www.flickr.com/help/forum/en-us/90609/ as an example.

## References

Angus, Emma, Mike Thelwall, and David Stuart. 2008. General patterns of tag usage among university groups in Flickr. *Online Information Review* 32 (1): 89–101.

Ashby, F. Gregory, Alice M. Isen, and U. Turken. 1999. A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review* 106 (3): 529–50.

Barthes, Roland. 1964. Rhetoric of the image. In *Image, music, text*. Essays selected and translated by Stephen Heath, 32–51. Repr. Glasgow: Williams Collins, 1977.

——. 1981. *Camera lucida: Reflections on photography*. Trans. Richard Howard. London: Flamingo, 1984.

Beaudoin, Joan. 2007. Flickr image tagging: Patterns made visible. *Bulletin of the American Society for Information Science and Technology* 34 (1): 26–9.

Beck, Susan E., and Kate Manuel. 2008. *Practical research methods for librarians and information professionals*. New York: Neal-Schuman.

Borlund, Pia. 2003. The concept of relevance in IR. *Journal of the American Society for Information Science and Technology* 54 (10): 913–25.

Cox, Andrew M. 2008. Flickr: A case study of Web 2.0. *Aslib Proceedings: New Information Perspectives* 60 (5): 493–516.

Ding, Ying, Elin K. Jacob, James Caverlee, Michael Fried, and Zhixiong Zhang. 2009. Profiling social networks: A social tagging perspective. *D-Lib Magazine* 15 (3/4). http://www.dlib.org/dlib/march09/ding/03ding.html.

Enser, Peter G.B. 1991. An indexing-free approach to the retrieval of still images. In *British Computer Society 13th Annual Information Retrieval Colloquium*, ed. A. McEnery, 41–55. London: British Computer Society.

——. 2000. Visual image retrieval: Seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science* 26 (4): 199–210.

Fisher, Adam. 2009. 50 best websites 2009—Flickr. *TIME.com.* http://www.time.com/time/specials/packages/article/0,28804,1918031_1918016,00.html.

Flickr. 2007. The help forum: Question in regard to Search/groups/most relevant. http://www.flickr.com/help/forum/44629/?search=relevant.

——. 2009. Explore/about interestingness. http://www.flickr.com/explore/interesting/.

Golder, Scott A., and Bernardo A. Huberman. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32 (2): 198–208.

Greisdorf, Howard. 2003. Relevance thresholds: A multi-stage predictive model of how users evaluate information. *Information Processing & Management* 39 (3): 403–23.

Greisdorf, Howard, and Brian O'Connor. 2002. Modelling what users see when they look at images: A cognitive viewpoint. *Journal of Documentation* 58 (1): 6–29.

——. 2008. *Structures of image collections: From Chauvet-Pont-D'Arc to Flickr.* Westport, CT: Libraries Unlimited.

Jörgensen, Corinne. 1998. Attributes of images in describing tasks. *Information Processing & Management* 34 (2/3): 161–74.

——. 2003. *Image retrieval: Theory and research.* Lanham, MD: Scarecrow.

Kipp, Margaret E.I. 2008. @toread and cool: subjective, affective and associative factors in tagging. Proceedings of the Canadian Association for Information Science. http://www.cais-acsi.ca/proceedings/2008/kipp_2008.pdf.

Korfhage, Robert R. 1997. *Information storage and retrieval.* New York: Wiley.

Krause, Michael G. (1988). Intellectual problems of indexing picture collections. *Audiovisual Librarian* 14 (4): 73–81.

Lee, Hyuk-Jin, and Diane Neal. 2007. Toward Web 2.0 music information retrieval: Utilizing emotion-based, user-assigned descriptors. In *Proceedings of the American Society for Information Science and Technology.* http://www.asis.org/digitallibrary.html.

——. In press. A new model for semantic photograph description combining basic levels and user-assigned descriptors. *Journal of Information Science.*

Markey, Karen. 1984. Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research* 6: 155–77.

Neal, Diane R. 2006. *News photography image retrieval practices: Locus of control in two contexts.* PhD diss., University of North Texas.

——. 2007. Folksonomies and image tagging: Seeing the future? *Bulletin of the American Society for Information Science and Technology* 34 (1): 7–11. http://www.asis.org/Bulletin/Oct-07/Neal_OctNov07.pdf.

——. 2008. News photographers, librarians, tags, and controlled vocabularies: Balancing the forces. *Journal of Library Metadata* 8 (3): 199–219.

Neal, Diane, Andrew Campbell, Jason Neal, Casondra Little, Anissa Stroud-Mathews, Shontadra Hill, and Cyntria Bouknight-Lyons. 2009. Musical facets, tags, and emotion: Can we agree? Paper presented at the 2009 iConference—iSociety: Research, education, engagement, Chapel Hill, NC.

Neumann, Dirk, and Karl G. Gegenfurtner. 2006. Image retrieval and perceptual similarity. *ACM Transactions on Applied Perception* 3 (1): 31–47.

O'Connor, Brian C., and Roger B. Wyatt. 2004. *Photo provocations.* Lanham, MD: Scarecrow.

O'Reilly, Tim. 2005. What is Web 2.0: Design patterns and business models for the next generation of software. *O'Reilly.* http://oreilly.com/pub/a/web2/archive/what-is-web-20.html.

Ornager, Susanne. 1997. Image retrieval: Theoretical analysis and empirical user studies on accessing information in images. In *Proceedings of the 60th Annual Meeting of the American Society for Information Science*, 202–211. Medford, NJ: Information Today.

Panofsky, Erwin. 1955. *Meaning in the visual arts.* Garden City, NY: Doubleday Anchor Books.

Persson, Olle. 2002. Image indexing: A first author co-citation map. http://www8.umu.se/inforsk/Imageindexing/imageindex.htm.

Power, Mick J. 2006. The structure of emotion: An empirical comparison of six models. *Cognition & Emotion* 20 (5): 694–713.

Rafferty, Pauline, and Rob Hidderley. 2007. Flickr and democratic indexing: Dialogic approaches to indexing. *Aslib Proceedings: New Information Perspectives* 59 (4/5): 397–410.

Rorissa, Abebe. 2008. User-generated descriptions of individual images versus labels of groups of images: A comparison using basic level theory. *Information Processing & Management* 44: 1741–53.

Rorissa, Abebe, and Hemalata Iyer. 2008. Theories of cognition and image categorization: What category labels reveal about basic level theory. *Journal of the American Society for Information Science and Technology* 59 (9): 1383–92.

Rorvig, Mark E. 1986. The substitutability of images for textual description of archival material in an MS-DOS environment. In *Proceedings of the Second International Conference on the Application of Micro-Computers in Information, Documentation, and Libraries*, 407–15. New York: North-Holland.

Rorvig, Mark E., Charles H. Turner, and Jesus Moncada. 1999. The NASA Image Collection Visual Thesaurus. *Journal of the American Society for Information Science* 50 (9): 794–8.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. 1976. Basic objects in natural categories. *Cognitive Psychology* 8: 382–439.

Rose, Gillian. 2007. *Visual methodologies: An introduction to the interpretation of visual materials.* London: Sage.

Rubin, Victoria, Jeffrey M. Stanton, and Elizabeth D. Liddy. 2004. Discerning emotions in text. *Proceedings of the AAAI Spring Symposium: Exploring Attitude and Affect in Text: Theories and Applications.* Stanford, CA. http://publish.uwo.ca/~vrubin/Publications/RubinStantonLiddyAAAI2004.pdf.

Saracevic, Tefko. 2007. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology* 58 (13):1915–33.

Schamber, Linda, Michael B. Eisenberg, and Michael S. Nilan. 1990. A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management* 26 (6): 755–76.

Schmidt, Stefanie, and Wolfgang G. Stock. 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): 863–76.

Shatford, Sara. 1986. Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly* 6 (3): 39–62.

Stvilia, Besiki, and Corinne Jörgensen. 2009. User-generated collection-level metadata in an online photo-sharing system. *Library & Information Science Research* 31: 54–65.

Surowiecki, James. 2004. *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations.* New York: Doubleday.

Sutter, John D. 2009. Wikipedia: No longer the Wild West? *CNN.com.* http://www.cnn.com/2009/TECH/08/26/wikipedia.editors/index.html.

Svenonius, Elaine. 1994. Access to nonbook materials: The limits of subject indexing for visual and aural languages. *Journal of the American Society for Information Science* 45 (8): 600–6.

Vander Wal, Thomas. 2007. Folksonomy. *Vanderwal.net.* http://www.vanderwal.net/folksonomy.html.

Vaughan, Liwen. 2001. *Statistical methods for the information professional: A practical, painless approach to understanding, using, and interpreting statistics.* Medford, NJ: Information Today.

Wildemuth, Barbara M., and Carol L. Perryman. 2009. Discourse analysis. In *Applications of social research methods to questions in information and library science*, ed. Barbara M. Wildemuth, 320–8. Westport, CT: Libraries Unlimited.

Wilson, Patrick. 1968. *Two kinds of power: An essay on bibliographical control.* Berkeley, CA: University of California Press.

——. 1973. Situational relevance. *Information Storage and Retrieval* 9 (8): 457–71.

Xu, Yunjie (Calvin), and Zhiwei Chen. 2006. Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology* 57 (7): 961–73.

Yoon, JungWon. 2009. Towards a user-oriented thesaurus for non-domain-specific image collections. *Information Processing and Management* 45: 452–68.