SAN JOSÉ STATE UNIVERSITY | SCHOOL OF LIBRARY AND INFORMATION SCIENCE

SLIS Student Research Journal

Volume 1 | Issue 2 Article 5

12-21-2011

A picture is worth a thousand words: The perplexing problem of indexing images

Lindsay L. Marlow
State University of New York at Buffalo, llmarlow@buffalo.edu

Amy Miller
State University of New York at Buffalo, anmiller@buffalo.edu

Follow this and additional works at: http://scholarworks.sjsu.edu/slissrj

Part of the Library and Information Science Commons

Acknowledgements
We would like to thank Dr. Valerie Nesset and Dr. Brenda Battleson for their support, positive feedback, and guidance during this project and throughout our time in Library School.

Recommended Citation
Marlow, L. L., & Miller, A. (2011). A picture is worth a thousand words: The perplexing problem of indexing images. *Student Research Journal, 1*(2). Retrieved from: http://scholarworks.sjsu.edu/slissrj/vol1/iss2/5

This article is brought to you by the open access Journals at SJSU ScholarWorks. It has been accepted for inclusion in SLIS Student Research Journal by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

A picture is worth a thousand words:
The perplexing problem of indexing images

During the past 20 years, technological advances have drastically changed everyday processes. These changes have manifested in a sharply increased use of the Internet that has in turn ushered in an age of digitization. Large-scale projects across the world are rapidly digitizing materials and storing them in digital libraries. These projects have created large collections of materials readily accessible to millions that were previously only available to users locally. The great strides created in access are revolutionary, but the proliferation of digital technology also creates issues with information retrieval. One format ubiquitous to most digital and traditional collections is the image. Whether in hardcopy or digital format, images pose challenges in the areas of image retrieval, indexing systems, and options for user interaction (Matusiak, 2006; Neugebauer, 2010). CONTENTdm® is a valuable tool used for adding images to digital libraries. It assists the indexer in indexing different types of multimedia through the use of a controlled vocabulary system and metadata fields (Vermillion, 2007). Currently, there is no viable mechanism to allow users to search and retrieve images using visual means; thus, all indexing, search, and retrieval is based on text (Chai, Zhang, & Jin, 2007). This paper is only concerned with descriptive metadata. Traditionally, indexers have used standards developed for text-based media such as books, periodicals, and documents (Ménard, 2009b). These standards are not entirely satisfactory for images due to the complexity and richness of visual media, language ambiguities, and the limitations of human indexing (Matusiak, 2006). The purpose of this paper is to examine the current research surrounding image indexing, identify the implications to the indexing profession, propose a potential solution to increase successful image retrieval, and establish areas in need of further research.

Literature Review

The primary problem in indexing images is their rich and inherently subjective format. Every user and every indexer sees different things when they look at an image, giving it multiple meanings (Chai, Zhang, & Jin, 2007; Neugebauer, 2010). Therein lies the trouble for the indexer. It is extremely difficult to find terms that both correctly describe the image and will also be recognized by users. Traditionally, indexers assign descriptors based on two criteria: *ofness*, the concrete and objective entities, and *aboutness*, the abstract and subjective inferences (Ménard, 2009a). Indexers in the digital age also need to address the equally complex problem of including self-awareness of the cognitive functions of the user's mind in their indexing (Greisdorf & O'Connor, 2002).

This awareness is essential because the mind of the viewer develops the impressions rendered from the subjective theme of the image. This is best described through Greisdorf and O'Connor's (2002) two cognitive viewpoints. The first cognitive viewpoint is the two-step process of visual retrieval completed by the viewer. The first step consists of creating the visual response by sensory stimuli and matching it to a syntactic equivalent. This means that the viewer is able to describe the image in a series or string of words. If the user has not seen the image before, he or she must conclude what the image is of and about. In the second step, the viewer evaluates the image based on the information need (Greisdorf & O'Connor, 2002). The user decides if the image is related to the topic, if the meaning is understood, and if the image can be used to satisfy the information need. The other cognitive viewpoint involves hierarchical levels of perception. This is the idea that humans evaluate and give meaning to images based on three levels (Greisdorf & O'Connor, 2002). The first level is the primitive feature; this includes color, shape, and texture of the image. The second level, the objects level, is a detailed look that involves noticing people, location, and actions within an image. The third and most complex level is inductive interpretations. This is where the image viewer's inherent subjectivity takes form. Either the viewer sees a symbolic value, or an emotional cue is triggered from the image. The problems for the indexer are as follows: not knowing at which level to index, determining how many levels to index, and predicting what the emotional response would be for individual users. Griesdorf and O'Connor's cognitive hierarchical levels of perception can be compared to Panofsky's (1955) three levels of meaning in a work of art.

Panofsky's seminal work (1955) identifies three levels of meaning: pre-iconography, iconography, and iconology. Pre-iconography is the most basic level of understanding consisting of the primary or natural subject matter. Iconography is used for cultural knowledge, including factual and expressional concepts. Iconology is the term used for the technical, cultural, and intrinsic content of the work, in addition to the method of interpretation based on synthesis of these elements (Panofsky, 1955). The levels are similar to the model proposed by Greisdorf and O'Connor (2002); however, the latter research applies to all images, whereas artwork, specifically Renaissance Art, was the focus of Panofsky's research.

Traditional methods of indexing images

The aforementioned authors have attempted to capture and define the inherent subjectivity of the image format. Three traditional approaches to indexing images are currently used to address this research: human indexing, controlled vocabularies, and computer extraction. During human indexing, a human indexer

selects the terms she or he feels best describes the image. This is thought to be a more accurate approach to indexing because it captures the intellectual process behind an image. Human indexers are able to capture emotional and contextual cues that otherwise would be missed by some controlled vocabularies and most computer algorithms. However, human indexing has several disadvantages. It is highly subjective, labor-intensive, and fraught with debate upon the level at which an image should be indexed (Chai, Zhang, & Jin, 2007; Matusiak, 2006; Neugebauer, 2010).

Controlled vocabulary includes classification schemes and thesauri that are developed to promote uniformity and to increase the probability of matching indexing language with search language. This process improves retrieval. Controlled vocabularies are limiting in that they represent concepts in an artificial way by using terms that are correct at the linguistic level but are infrequently incorporated in real life by users. For example, a controlled vocabulary would use a generic term such as *facial tissue* and not *Kleenex*®, since *Kleenex*® is a brand name. However, many users might search for the term *Kleenex*®, a recognized brand name, instead of the more general term *facial tissue*, thereby retrieving fewer results from their search. Furthermore, controlled vocabularies are expensive to create and constant maintenance is needed in order for the controlled vocabulary to remain viable (Matusiak, 2006; Ménard, 2009b).

Computer extraction uses a software program that is designed to automatically identify and extract primitive features from the image and to assign descriptors. This system offers the promise of eliminating bias and assigning descriptors without the inherent subjectivity of human indexing. However, there is currently no system in mass production that fully satisfies end-users. Automated annotation is more efficient but less accurate. This is because there is no existing algorithm to account for semantic relationships—defining elements into verbs and adjectives—or to capture the intellectual processes behind an image. The only assistance computer extraction methods can provide at the moment is with the identification of primitive shapes and textures within an image and often this is lacking (Chai, Zhang, & Jin, 2007; Matusiak, 2006; Neugebauer, 2010).

Each of the aforementioned methods have merit; however, independently, they fall short of user retrieval needs. Without descriptive and comprehensive indexing, images have the potential to remain inaccessible, effectively hidden from users (Matusiak, 2006). This problem is particularly acute in the Internet realm, due to the lack of assistance from information professionals. The literature defines two methods for image indexing, concept-based and content-based (Chai, Zhang, & Jin, 2007; Ménard, 2009b; Neugebauer, 2010). Concept-based indexing is performed by human indexers who examine characteristics of the image and identify and describe semantic content. This type of indexing is generally more descriptive, but is prone to subjectivity issues. The process of translating the

content of an image into verbal expressions poses significant challenges to indexers. The resulting descriptors frequently do not meet user needs nor do they provide effective retrieval. Content-based indexing is often an automated process where features of the image, such as color, shape, or texture, are identified, extracted, and made into descriptors. Machine-driven indexing can miss key relationships and fail to describe the intellectual processes behind images. Thus far, a content based-image retrieval system has yet to be produced that satisfies the end-user (Ménard, 2009b). This may be due to the disconnect between what users articulate for text-based queries and what the computer extracts. Since they do not precisely describe the information users need, a gap is created between low-level visual descriptors and users' semantic expectations. A combination of approaches, in addition to the incorporation of user-generated tagging, is supported by current research on the topic (Chai, Zhang, & Jin, 2007; Matusiak, 2006; Ménard, 2009a; Ménard, 2009b; Neugebauer, 2010).

It is of little use to speak of the inherent problems with indexing images and current research in the field without relating this information to a larger context. In order to improve image search and retrieval, a synthesis of the aspects of the problem along with proposed solutions must be developed. Possible solutions should be tested in order to ascertain the optimum answer for both indexers and users, hopefully providing an opportunity for better image indexing and retrieval.

Incorporating Social Tagging into Image Indexing

A new method of image indexing relying on social tags has replaced traditional methods in many public user driven sites such as Flickr, Tumblr, and Delicious. The use of social tagging allows users to ascribe uncontrolled tags or labels to an item. Social tagging is increasingly used in many digital collections, including those available freely on the Internet. Tags solve the problem of vocabulary control because they provide additional access points apart from conventional ones such as a user-generated term of *trains* opposed to the Library of Congress Subject Headings' (LCSH) use of the term *rail transport*. Tags are useful in part due to their use of natural language. This increases the variability in the keywords assigned to items, ranging from very general tags to more specific tags. While this wide variability can be an advantage, it also serves as a disadvantage because it often results in a lack of control. This lack of control can allow incorrect tags or an excessive number of tags to be assigned to an image. This may result in the creation of too many access points, making retrieval difficult. The act of social tagging is also individualized since it is usually done for private images. Social tagging is primarily used in the personal realm for items that are owned by or important to the user. It is not known if users are willing to invest their own

personal time and effort to describe images in an altruistic manner and for free. This could decrease the chances of accurate tags being assigned (Chai, Zhang, & Jin, 2007; Matusiak, 2006; Ménard, 2009a). As a result, social tagging has not yet been implemented in a way that best fits the needs of all users. Case studies aimed at determining if users would assign accurate tags if they had no personal connection to the material's content would help to further clarify social tagging in relation to images and digital collections.

A case study from O'Connor, O'Connor, and Abbas (1999) helps to further illustrate the limitations of traditional methods and tagging, while supporting a collaborative approach. This study comprised a survey of 120 Master's students in a Library and Information Science program. The participants were asked to respond to an image depicting a duck on water. Each respondent ascribed unique descriptors for the subject of the image and gave phrases defining how the image made them feel. The responses users gave would qualify as social tagging because the descriptors or phrases would not necessarily be found in an authoritative controlled vocabulary, such as LCSH. User responses for the subject terms included: *duck, water, mallard, goose, placid lake, water scene, paddling, reflection, evening, summer,* and *waterfowl*. For the emotional response, users responded with the following terms, among others: *glorious, restful, I hope it's not hunting season, serene, solitary, relaxing, pretty, calm waters, I would love to go swimming too, refreshing,* and *quiet water with a smug duck* (O'Connor, O'Connor, & Abbas, 1999, p. 687). It is evident that the variety of descriptors ascribed to this one image illustrates the need for a collaborative approach among both indexers and users in the process of indexing images for information retrieval.

The retrieval of ordinary images representing common objects is more effective when the images have been indexed using a combination of controlled and uncontrolled vocabularies (Ménard, 2009b). While not a stand-alone solution, user-generated tags have merit in the form of an enhancement to the traditional methods of indexing images and introducing uncontrolled descriptors. Tagging would allow new terms, multiple languages, and cultural influences to be reflected, in addition to the characteristics ascribed by the indexer. This combination would optimize queries and improve image retrieval (Matusiak, 2006). A process like this would foster collaborative knowledge construction, potentially reversing the isolated act of indexing, and would garner increased user involvement. Tagging would increase interactive feedback from the users of image retrieval systems, thus creating a visible gauge of their utility. Images are inherently multidisciplinary; therefore, it would follow that the best way to describe and index them would also be a concerted effort from a combination of parties: indexers who control the language and attempt to capture the intellectual information behind an image, machines that take an unbiased view of images and

ascribe characteristics, and users who define images in relation to the world as they see it (Chai, Zhang, & Jin, 2007; Matusiak, 2006; Ménard, 2009b; Neugebauer, 2010).

Marlow and Miller's Collaborative Model for Image Indexing

The literature overwhelmingly favors incorporating social tagging into traditional methods of image indexing. However, the logistics of this contemporary collaboration have yet to be defined. The authors of this paper propose a solution to the challenge of indexing images. Current systems utilize separate approaches, whereas a collaborative design would be advantageous to indexer and user alike. Further studies and additional research should focus on creating an interoperable interface that can be incorporated into various data and content management software programs to facilitate user-generated information. Current data and content management software programs used in digital libraries, such as CONTENTdm®, could be modified to include a metadata field for user-generated descriptors, also known as social tagging. The software would optimally allow a chosen group of expert users to define terms for a given image. Descriptors would then be selected based upon the consensus of the entire user base via a single click polling mechanism. Expert users would vary depending upon the class of images or the collection being indexed. The expert user title would require that these expert users have some proficiency with the subject matter or credentials to ensure they accurately tag the image(s). Further study is needed to determine if CONTENTdm® is the best platform available to implement tagging.

The proposed model is depicted in Figure 1. It can be effectively demonstrated using a website such as *New York Heritage* (http://www.nyheritage.org), which uses CONTENTdm® as their content management system. The newly created *New York Heritage* research portal merges the previous *Western New York Libraries Resources Council* (http://www.wnylegacy.org) website with collections from the eight other regions of New York. Subject specialist librarians from each of the regions represented could be selected by site administrators to assign tags as expert users. This selection would provide for the slight differences in dialect (i.e., language ambiguities) across the state. A broad selection would also blend regional history and culture, thereby creating multiple access points. Each expert librarian would assign the same number of descriptors to each image. Research will be needed to identify a method to select expert users since not all collections function in the same way as the *New York Heritage* research portal. Tags would then be pooled together and displayed within the CONTENTdm® software below the image they describe to be voted upon by the users. They would also be placed in the social tagging metadata field until the polling process is complete. Metadata would only

be accessible to system administrators. The administrators would use the content management software to oversee the entire process. They would monitor the assigned tags, supervise the polling system, and select the final social tags to be included in the metadata based on the consensus of the user base.

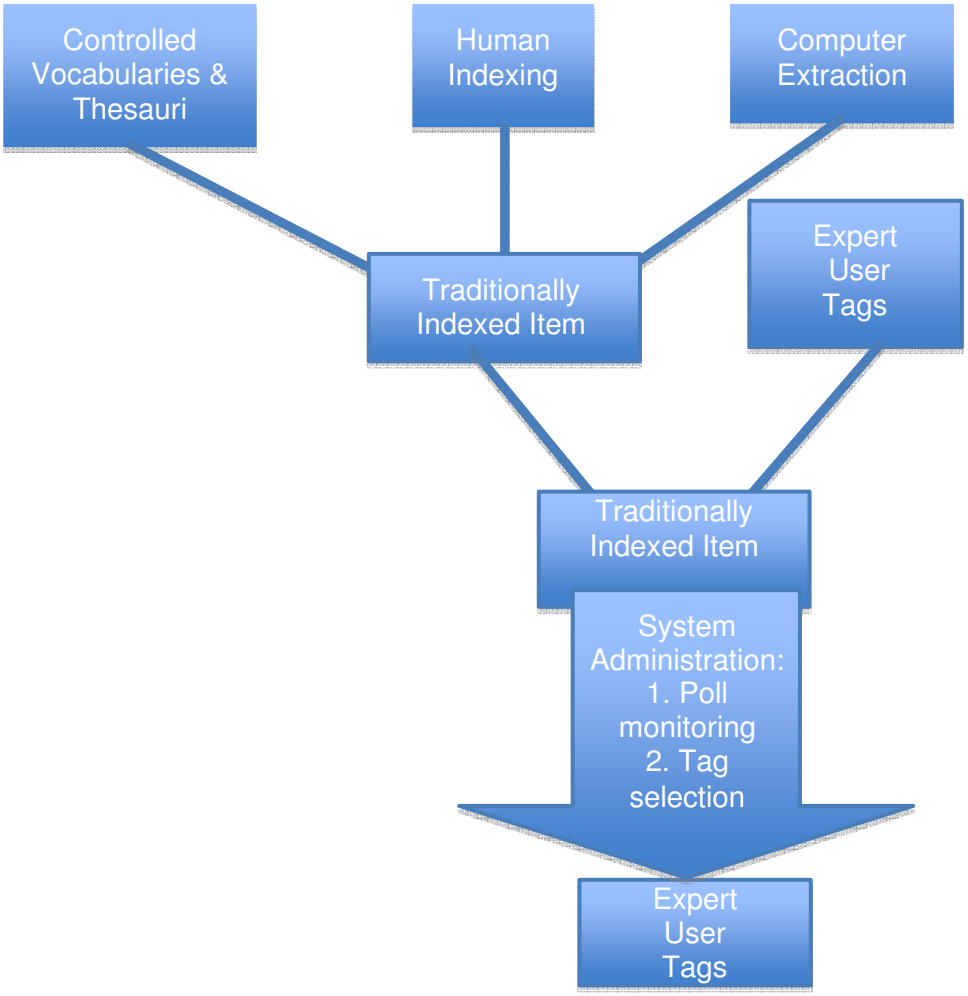

Figure 1. Marlow and Miller's collaborative model for image indexing.

The administrators would choose either a given amount of time, a certain number of clicks, or a combination of the two for the polling system selection mechanism. The administrators would incorporate the polling system with the indexer created descriptors. A one-click only link would facilitate voting by

general users. This single vote system would prevent "spagging," or spam tagging, often done for profit or to cause damage (Steele, 2009). Multiple votes by a single user would be prevented by a mechanism similar to the paywall instituted by websites such as the *New York Times* (http://www.nytimes.com/). Users' cookies would alert the website to their previous activity, hindering most attempts to inappropriately tag. One flaw with this system is the ability to delete one's cookies and function on the website as though it had never been visited before. The only viable way to prevent this action would be to integrate a username login system. However, this could possibly decrease user traffic to the website due to patrons' potential unwillingness to create a username and password, therefore creating a barrier to access. These intricacies would need to be assessed and examined through further research and case studies in order to implement the best possible system with the widest access for all users.

After the conclusion of the designated polling period, site administrators would then assign the tags receiving the most votes as descriptors. These tags would be incorporated into the metadata and displayed below the image in order to create access points. Another point to consider is the popular use of social tagging clouds, as seen on websites such as Flickr (http://flickr.com/), which have been incorporated into some digital library websites. Tag clouds are visualizations that display tags frequently assigned to images or tags selected the most frequently by users accessing images. Tags garnering the most traffic are visually displayed in larger font sizes to establish their popularity. The type of cloud most appropriately used by a digital library would be the cloud that enlarges the tags most selected by users. The cloud would only be displayed on the home page of the website to increase access points to users. This, in turn, may help them to feel less intimidated by the search process of a digital library and may facilitate additional user browsing. It would not be advisable to display the cloud on the same web page as the images as it may cause users to become overwhelmed.

Conclusion

The widespread use of digital technology and the Internet ushered in the current information explosion. The pervasiveness and magnitude of information available in an instant today makes the job of the information professional paramount. A high level of organization, excellent search and retrieval, and multiple access points to information are key in the information age. Indexing of images has always been problematic because of their richness of content and innate subjectivity. This issue has been magnified due to their boundless uses in society today. A sharp increase in the growth of digital libraries is a direct consequence of our embrace of digital culture. The digital nature of these collections has granted access to a much wider audience. Previously, materials were only available to

users locally. The mere presence of this information in an online format is not enough. The content must be accessible to users or its fate is to remain forever hidden by the sheer volume of information.

Current research supports a collaborative approach incorporating controlled and uncontrolled vocabularies, along with user-supplied content. This addition could satisfy the need for additional access points to information and users who wish to take an active role in the process. Tag clouds have already been incorporated into some digital libraries; however, further steps should be taken to ensure user satisfaction. The literature supports the model laid out within this paper because of its application of user-generated content along with traditional methods of indexing. This is just one proposed collaborative method that would need to be implemented, further studied, and critically evaluated alongside other suggested processes. Additional study in computer extraction methods is also needed. Research in the area of advanced algorithms could provide additional help with assigning primitive and possibly object descriptors while avoiding subjectivity and bias. This is a growing field and its advancement could contribute to the growing collaborative nature of image indexing. The issue of indexing images will continue to be a major issue within the profession due to the irreversible subjectivity of images. The method described in this paper is one potential way to alleviate bias and the pressure placed on indexers while attempting to index images with the user in mind.

References

Chai, J. Y., Zhang, C., & Jin, R. (2007). An empirical investigation of user term feedback in text-based targeted image search. *ACM Transactions on Information Systems, 25*(1), 1-25. doi:10.1145/1198296.1198299

Greisdorf, H., & O'Connor, B. (2002). Modelling what users see when they look at images: A cognitive viewpoint. *Journal of Documentation, 58*(1), 6-29. doi:10.1108/00220410210425386.

Matusiak, K. K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems & Services, 22*(4), 283-298. doi:10.1108/10650750610706998

Ménard, E. (2009a). Image retrieval: A comparative study on the influence of indexing vocabularies. *Knowledge Organization, 36*(4), 200-213.

Ménard, E. (2009b). Images: Indexing for accessibility in a multi-lingual environment—challenges and perspectives. *The Indexer, 27*(2), 70-76.

Neugebauer, T. (2010). Image indexing. *The Indexer, 28*(3), 98-103.

O'Connor, B. C., O'Connor, M. K., & Abbas, J. M. (1999). User reactions as access mechanism: An exploration based on captions for images. *Journal of the American Society for Information Science, 50*(8), 681-697. doi:10.1002/(SICI)1097-4571(1999)50:8<681::AID-ASI6>3.0.CO;2-J

Panofsky, E. (1955). *Meaning in the visual arts: Papers in and on art history.* Garden City, NY: Double Day.

Steele, T. (2009). The new cooperative cataloging. *Library Hi Tech*, *27*(1), 68-77. doi:10.1108/07378830910942928

Vermillion, J. (2007). Indexing images. *Key Words, 15(*1), 12-14.