Beaudoin, J. (2008). A preliminary investigation of image indexing: The influence of domain knowledge, indexer experience and image characteristics. 19th Annual ASIS SIG/CR Classification Research Workshop doi: 10.7152/acro.v19i1.12852

A preliminary investigation of image indexing: The influence of domain knowledge, indexer experience and image characteristics

Joan Beaudoin, Ph.D. Candidate
College of Information Science and Technology, Drexel University
3141 Chestnut Street, Philadelphia, PA 19104-2875
Email: Joan.Beaudoin@drexel.edu

Subject tags: indexing, image indexing, interindexer consistency, indexing experience, subject expertise

Introduction

This study investigates the application of conceptual terms to images by individuals with various educational and occupational backgrounds. While the inherent complexities of applying terms to images are broadly acknowledged, few studies have addressed the issue of how subject expertise or practical image indexing experience may impact the work. This study begins work in this direction by examining the terms applied to a series of images by individuals with different levels of domain knowledge and practical indexing experience. In addition to the indexers' varying backgrounds, the study examined how the images' modes of representation and interpretation influenced the application of terms.

Background

Image indexing research is a relatively youthful area with the majority of its literature produced within the last few decades. The image indexing literature generally falls into several broad research areas. The first of these concerns an individual's level of visual literacy. The importance of subject knowledge to the practice of image indexing is not always recognized. Individuals' capacity to understand what is being viewed is not uniform, and so one's ability to see does not ensure one's ability to read an image (Turner, 1993). Even with a high degree of domain knowledge, the meaning of an image can present some interesting problems. This is a topic that Sarah Sharford Layne has examined. Using the theories of the art historian Erwin Panosfky she investigated the multiple layers of meaning that can be present within a single image (Shatford Layne, 1986). For example, words can be used to describe what is represented within an image (what the image is of) or the image's underlying meaning (its aboutness), and each of these may be described with varying levels of detail. Corinne Jörgensen (2003) has researched the various types of information people use to describe and retrieve images. Another area of research into image indexing has focused attention on the needs of image users. Armitage and Enser's (1997) research into this topic revealed that image users' needs are every bit as complex as those found in the parallel universe of printed media.

One of the most limited research areas concerning image indexing is interindexer consistency. Two studies to investigate this topic are those of Markey (1984) and Wells-Angerer (2005). Markey's investigation looked at the indexing terms applied by thirty-nine individuals to one hundred images of Medieval works on three different categories (objects, expressional, events). A low percentage of agreement of terms was reported by Markey, with an average of 7% for exact term matches, and 13% for conceptual matches in indexing terms. In a study assessing the influence of indexer subject knowledge on image retrieval rates of online museum collections Wells-Angerer (2005) investigated the terms applied to ten works of art by thirty participants falling into three categories of image indexers (expert, knowledgeable, novice). Wells-Angerer found the terms applied by indexers with the highest level of knowledge about the objects in the collections (scholars, curators and collection staff) had retrieval success rates of approximately 16%. Indexer retrieval rates for those who had less subject knowledge were considerably lower, at approximately 5% (Wells-Angerer, 2005). The results of Wells-Angerer's investigation indicate that indexer experience and subject expertise ought to be considered in discussions of interindexer consistency. Markey's study has been used on several occasions to support the hypothesis that image indexing produces low returns for the effort involved in the work. This is remarkable as Markey (1984) states that "[t]he use of inexperienced indexers and non-subject

ISSN: 2324-9773

specialists in this study may have diminished interindexer consistency scores." The limited number of studies investigating the practices of image indexers, and the conflicting results of these two studies, indicate additional research is warranted in the area of image indexing. Thus, the present study was undertaken in order to explore some of the issues at work which influence image indexing.

Research Questions

Several research questions were developed to drive the study and these are as follows:

- Do image indexer experience and subject expertise affect interindexer consistency?
- What types of terms (generic description, identification, interpretive) exhibit the highest interindexer consistency among indexers?
- What influence does image type have on indexing?

Research Methods

Data was gathered through a web-based survey using WebSurveyor (now Vovici) From June through December 2006. The study was publicized through several listservs and blogs (VRA, VRAP, ARLISNA, ARLISnap, H-INFO, H-BIBLIO) and as a printed flyer posted around several campuses in the greater Philadelphia area. Through the online survey 140 participants provided demographic data and indexing terms for eight images. The first part of the survey consisted of a questionnaire which collected basic demographic data, the number and types of courses the participants had completed with a visual focus, their level of image indexing experience, and the frequency of their image indexing. The second part of the survey was an indexing component which collected terms assigned by the participants to a series of eight images. Each image was presented at the top of the screen with ten data entry boxes beneath.

Images of cultural works formed the focus of the study. However, several documentary style photographs were included to assess the possible influence of an image's subject accessibility and mode of representation on the terms chosen by the study's participants. In order to evaluate whether or not the images themselves influenced the indexing, a framework was developed to look at these two fundamental characteristics (Table 1). Images were chosen for the study based on their level of realism and on the accessibility of their subjects. Two images were chosen to represent each of the four groups (basic and complex levels or representations and basic and complex levels of interpretation). The titles of the images are noted in Table 1.

Image Group 1	**Image Group 2**	**Image Group 3**	**Image Group 4**
Basic level of representation	Complex level of representation	Basic level of interpretation	Complex level of interpretation
Examples: **Image 1**. *Mountainous landscape* **Image 5.** *WWII Cemetery in Normandy, France*	Examples: **Image 2.** Claes Oldenburg's *Clothespin* **Image 6.** Franz Marc's *The Sheep*	Examples: **Image 3.** Duccio's *Madonna and Child*, **Image 7.** *Great Sphinx of Giza*	Examples: **Image 4.** Goya's *Executions of the 3rd of May, 1808*, **Image 8.** Pieter Claesz's *Vanitas*
Realistic representation.	Surreal, abstract or otherwise confounding representation.	Well-known subjects in the history of Western art.	Obscure themes in the history of Western art.

Table 1. Definition of image groups and titles of images chosen for the study.

For data analysis purposes the demographic data was used to divide the participants into several groups based upon their subject expertise or image indexing experience. The groups consisted of roughly equal number of participants (Table 2). The first group of 35 participants, titled Subject Novice (SN), had completed two or fewer courses in any discipline with a strong visual focus (fine arts, art history, archaeology, architecture and so on). The second group of 32 participants, the Subject Experts (SE), had completed eleven or more courses in a discipline with a strong visual focus. The third group of 33 participants, the Image Indexers (II), was identified by the frequency with which they performed image indexing (once a week or more). In the case of this last group, the Image Indexers, the vast majority (28 of 33 participants) would also qualify as Subject Experts. As Table 2 illustrates, gender equality was highest in the first group (SN) and lowest in the third (II). The demographic data collected from the participants showed that first group (SN) and the second group (SE) had a broader range of degree attainment when compared to the third group (II).

The total number of participants for the data analysis presented here is 100 (74 female and 26 male). The data from the remaining 40 participants, who had moderate subject expertise (3 to 10 visually oriented courses) and limited or no image indexing experience, was not analyzed. It was believed that the data from the participants falling in the extreme ranges of subject expertise and those with strong image indexing experience would offer a clearer representation of what was taking place. The data collected from the 100 participants who fell into the three groups (SN, SE, and II) represented in Table 2 was analyzed using qualitative and descriptive statistics.

Group	Number	Female	Male	Education
Subject Novice	35	20	15	U 6 B 10 M 17 P 2
Subject Expert	32	25	7	U 1 B 7 M 19 P 5
Image Indexer	33	29	4	U 0 B 1 M 30 P 2

Table 2. Demographic details for the participants in the three groups studied.
(Education: U = un-degreed, B = Bachelors, M = Masters, P = Doctorate)

Results

Data analysis revealed that the participant's degree of subject expertise and indexing experience influenced their application of indexing terms. The images themselves also appear to have had an effect on the indexing. The data was examined to determine the number of terms applied by the indexers, the percentage of co-occurring terms, and the types of terms chosen.

Even at a very basic level of analysis, counting the number of terms applied to the images, it was clear that the participant's application of indexing terms varied and that they appeared to be influenced by subject knowledge, indexing experience and image type. An analysis of the average number of terms applied by each of the three groups revealed that the Subject Novices provided the fewest terms per image (5.05) in seven of the eight images. The Image Indexers were found to provide the highest number of terms per image (6.02), applying on average one more term per image than the Subject Novice participants. The Subject Expert participants applied an average of (5.33) terms per image which fell in between the other two groups. The average number of terms applied to the images by the different groups is noteworthy since it suggests that through domain knowledge individuals develop the ability to provide an increased number of terms to describe images, and that through indexing experience they develop this ability even further.

ISSN: 2324-9773

When looking at the average number of terms applied to each of the images by the three groups of indexers exceptions were found. Interestingly the inconsistencies which were found occur with the photographs included in the study. The strongest discrepancy was found in Image 5, a photograph of a World War II military cemetery located in Normandy, France. In this instance the Subject Novices applied terms at a slightly higher rate (5.46) than either of the other two groups (SE 5.28; II 5.24). The differences between the numbers of terms applied by the indexers were less pronounced in the case of Image 1, a view of a mountainous landscape (SN 5.06; SE 4.84; II 5.12). These anomalies in the data pattern seem to be related to the image type. Both of these images have highly realistic modes of representations and common straightforward themes and so the Subject Novice indexers were found to be equally adept at applying terms when compared to the Subject Experts and the Image Indexers.

The idea that the characteristics of an image itself could influence the number of terms applied found additional support in the case of Image 6, an abstract painting by Franz Marc. This image, the only abstract image included the study, received the lowest average number of terms across all participant groups (SN 3.66, SE 3.85, II 4.45). It seems that without readily recognizable figures within the image the participants were literally at a loss for words. The images which created the opposite situation, those with the highest number of applied terms, were different for each of the three participant groups. The Subject Novices applied their highest average number of terms (5.63) to Image 4, Goya's *3rd of May¸1808*. The Subject Experts provided the most terms on average (6.09) to Image 3, Duccio's *Madonna & Child*, and the Image Indexers' highest average (7.75) was found in Image 8, Claesz' *Vanitas*. While different images received the highest number of terms, there are similarities to be found among the three images. Each work is rendered with a degree of realism, and they all contained a great number of items and details to describe. It appears all of the indexer-participants applied a higher number of terms to works with realistic representations and richly interesting and accessible themes. The difference between the three groups might be the result of the varying accessibility of the different themes. For example, the Subject Novices seem to have been drawn to the explicit emotionalism of Goya's painting. The Subject Experts applied a great number of art historical terms to the Madonna & Child painting, and the Image Indexers essentially inventoried the objects rendered in Claesz's *Vanitas*. While the limited number of images in the study makes it difficult to state with certainty what impact the various modes of representation and interpretation have on indexing, it is clear that these fundamental image characteristics affect the application of terms.

In order to discover if any semantic patterns occurred among the terms applied by the three groups, the kinds of indexing terms were also examined. Each of the terms applied by the indexers was identified as generic, specific (identification) or interpretive using qualitative methods. Generic terms were used to describe persons, places, or things in a general way. Some examples of generic description terms found in the data analysis are man, violin, mountain, landscape, and shootings. Specific terms name particular people, places, times, things and cultural concepts. Examples of specific terms found in the data analysis are the Alps, Claes Oldenburg, French, 1814, and Romanticism. It should be noted here that this group of terms was the most difficult to define since the line of demarcation between a generic term and a specific identification was sometimes found to be difficult to discern. The final term type, interpretive, describes emotional responses to the image or a work's underlying meanings ("aboutness"). Some examples of interpretive terms found in the data analysis are desolate, angry, horror, veterans, death, and solemn.

The application of these three term types and how their usage varied among the groups and images was examined. The first aspect to be investigated was the frequency of application of the various term types. The most often applied type among the Subject Novices was found to be generic while the Subject Experts and the Image Indexers typically applied a greater percentage of specific terms. The application of a higher number of specific identification terms suggests that education and/or training predisposes individuals to index images with a higher degree of specificity. None of the participant groups applied a great number of interpretive terms.

ISSN: 2324-9773

However, the Subject Novices use of them was two to four times greater than that of the other participant groups. This reveals that the Subject Novices are more likely to note emotive or interpretive content in their choice of indexing terms when compared to the terms applied by subject specialists or practicing image indexers. Image type also seems to have exerted an influence over term usage, since Image 3 (Duccio's *Madonna & Child*), Image 6 (Marc's *The Sheep*), and Image 7 (*Great Sphinx of Giza*) received the highest percentage of specific terms across all three indexer groups. The impact of image content on the indexing was also seen in the case of Image 4, Goya's *3rd of May, 1808*, where the painting's strong emotional theme led to the application of the highest number of interpretive terms across all three groups.

The types of terms used to index the images are also notable in connection to the co-occurrence of terms. Before discussing term types in tandem with co-occurrence, the basics aspects of interindexer consistency need to be made explicit. In the analyses performed for this study the terms applied by indexers were examined for exact match co-occurrences, and so "mountain" and "mountains" were counted as a match, but "mountainous," "mountain range," and other variations were not. The top co-occurring terms are those which represent the highest number of overlapping term applications for a single image by each of the three indexer groups. The average for the top performing terms across all images in the study was 61% for the Subject Novices, 63% for the Subject Experts, and 70% for the Image Indexers. The highest performing single term applied in the study was "mountain" for Image 1, the *Mountainous Landscape*. The co-occurrence rates across the three groups were uniformly high for this single term (SN 97%, SE 94%, II 94%). Looking at the abstract painting by Marc (Image 6), the term with the highest rate of co-occurrence was modest (SN 31%, SE 38%, II 33%). This was lowest performing single top term of the study and, again, this low number is suggestive of the influence basic image characteristics have on indexing. Returning to the issue of term type, the top co-occurring terms showed a pattern connected to the kinds of terms chosen by the indexers. Generic descriptors had the highest co-occurrence rates, and the top three spots for co-occurring terms were almost exclusively generic terms for all of the study's images. A few specific terms crept into the top co-occurring spots, however, and these were the identifications of the "Madonna and Child" in Image 3, Duccio's *Maddona & Child*, and "Goya" in Image 4, Goya's *3rd of May, 1808*.

An examination of the co-occurring terms revealed there were a few high performing terms applied by each group and then the co-occurring terms dropped off rapidly. This pattern can be seen in Table 3 which shows the distribution of terms applied by the three groups of indexers. The Image Indexers had the highest overall co-occurrence percentages. However, their performance was only modestly better than the other two groups of indexers. The overall average co-occurrence rates for the groups are 4% for the Subject Novices, 4% for the Subject Experts, and 5% for the Image Experts. As was mentioned previously, the Image Indexers generally applied more terms to the images than the other two groups. So, while this group showed higher rates of co-occurrence among more terms, the large number of singleton terms applied by these indexers lowered the group's overall co-occurrence rate.

ISSN: 2324-9773

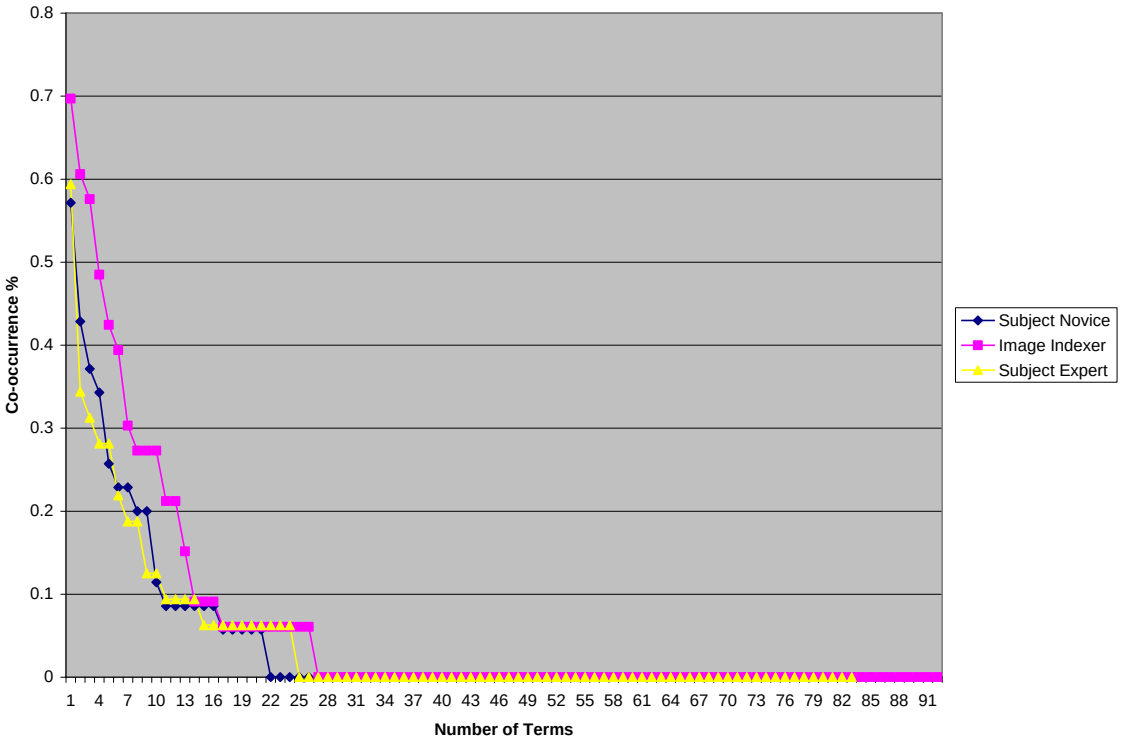

Table 3. Term distribution for Image 8, Claesz's *Vanitas*.
(Number of unique terms applied: SN = 79; SE = 83; II = 92)

Future work

This investigation revealed several interesting phenomena at work surrounding image indexing and future work is needed in order to validate and expand on upon the research that was undertaken. Additional research is clearly needed to increase our understanding of how image characteristics influence image indexing. This work should be undertaken so that these differences can be better accommodated in the indexing process. This in turn will help increase the effectiveness of image indexing. Finally, the discovery that emotive and interpretive terms were applied more readily by the Subject Novices is a finding that is calling out for more explanation.

Conclusions

Subject expertise and indexing experience were found to have an impact on the terms applied to images. The number of terms applied and the co-occurrence of terms was typically tied to the level of indexing experience and subject expertise of the participants. On the most basic level of analysis, the experienced image indexers provided on average the highest number of terms per image, with the subject experts supplying a slightly reduced number and the subject novice participants the fewest. Co-occurrence of applied terms among participant groups also followed this pattern. The images themselves were also found to have an influence on the number and types of terms applied and the rates of term co-occurrence achieved by the indexers of these images. The legibility of images with easily accessible subjects and realistic representation, while scoring well in terms of interindexer consistency were found to receive fewer term applications by the image indexers and the subject experts. This finding suggests that while interindexer consistency might be highest among skilled indexers and those with solid domain knowledge, a broader range of terms were sometimes applied to images with readily accessible subjects by those individuals who lacked training or subject expertise. Other interesting findings of the study point to the various kinds of terms applied by the three groups. The subject novices applied a greater number of generic terms to the images with the indexers and subject experts providing a

higher number of terms which identified specific aspects of an image. Finally, while the number of emotive or interpretive terms applied to the images was found to be very low across all three groups, the subject novices applied these terms more often than the other participant groups.

The results of this study provide a preliminary account of the influence of subject knowledge and indexing experience on image indexing. The findings of the study indicate subject knowledge and indexing experience have an influence on the indexing of images. This influence had both positive and negative affects. On the positive side, Image Indexers and, to a lesser extent, Subject Experts achieve higher interindexer consistency rates alongside providing rich and varied terms. On the negative side, these same indexers are less likely to apply emotive or interpretive terms and they sometimes do not fare so well when asked to index documentary style photographs. The results of the study also suggest that features inherent in an image play a pivotal role in indexing. Images with abstract representation and obscure themes posed difficulties for all three groups. An awareness of these various influences on image indexing is the first step in providing improved term application and ultimately better access to images.

References

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

Jörgensen, C. (2003). Image Retrieval. (Lanham, MD: Scarecrow Press, Inc.).

Markey, K. (1984). Interindexer consistency tests: A literature review and report of a test of consistency in indexing visual materials. *Library & Information Science Research,* 6, 155-177.

Shatford Layne, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging and Classification Quarterly,* 6(3), 39-62.

Turner, J. (1993). Subject access to pictures: Considerations in the surrogation and indexing of visual documents for storage and retrieval. *Visual Resources,* 9(3), 241-271.

Wells-Angerer, T. (2005). A study of retrieval success with original works of art comparing the subject index terms provided by experts in art museums with those provided by novice and intermediate indexer Master's paper, School of Information and Library Science, University of North Carolina, Chapel Hill.

ISSN: 2324-9773