Analyzing Users' Retrieval Behaviours and Image Queries of a Photojournalism Image Database / Analyse des comportements d'extraction de données et de requêtes d'images dans une base de données d'images de photojournalisme

Hsin-liang Chen, Thomas Kochtanek, Christopher Sean Burns, Rick Shaw

Canadian Journal of Information and Library Science, Volume 34, Number 3, September/septembre 2010, pp. 249-270 (Article)

Published by University of Toronto Press
DOI: *https://doi.org/10.1353/ils.2010.0003*

Analyzing Users' Retrieval Behaviours and Image Queries of a Photojournalism Image Database

Analyse des comportements d'extraction de données et de requêtes d'images dans une base de données d'images de photojournalisme

Hsin-liang Chen*
School of Library and Information Science
Indiana University
Indianapolis, IN 46202, U.S.A.
chenhsin@iupui.edu
*Corresponding author

Thomas Kochtanek
School of Information Science and Learning Technologies
University of Missouri
Columbia, MO 65211

Christopher Sean Burns
School of Information Science and Learning Technologies
University of Missouri

Rick Shaw
Picture of the Year International (POYi)
Columbia, MO 65211

Résumé : Cet article décrit une recherche entreprise dans le but de comprendre les besoins des utilisateurs finaux d'une importante base de données historique d'images de photojournalisme, disponible en ligne. Le protocole suivi comprenait l'examen des habitudes de recherche par rapport aux habitudes de navigation, les éléments de métadonnées Dublin Core utilisés lors de recherches avancées, et les tendances dans les types de termes de recherche. Les résultats indiquent que peu de visiteurs du site utilisent les fonctionnalités de recherche, et que très peu en utilisent les fonctions avancées. L'analyse des caractéristiques des termes de recherche s'appuie sur les travaux de Jörgensen (1998) pour proposer une nouvelle façon de considérer l'étiquetage social dans une implantation future de la base de données d'image de photojournalisme.

Mots-clés : extraction d'images, requête d'images, photojournalisme, études sur les utilisateurs, analyses des Web logs

Abstract: This paper describes research undertaken to understand end user needs within the scope of a large, online, historically centred photojournalism image database. The framework used included an examination of browsing vs. searching habits, of Dublin Core metadata elements used when conducting advanced searches, and of trends in types of search terms. Results suggest few visitors to the site use the site's search functionality and very few use advanced search. An analysis of search term characteristics, built on the work of Jörgensen (1998), suggests a new way to approach social tagging in a future implementation of the photojournalism image database.

Keywords: image retrieval, image queries, photojournalism, user studies, Web log analysis

Introduction

New approaches in managing image collections have increasingly taken advantage of emerging technologies over the last two decades and image-retrieval research attracts researchers and practitioners from a variety of disciplines (Chen 2001; Chen and Rasmussen 1999; Jörgensen 2003). One important research goal arising from these developments is to understand the needs of end users (Jörgensen 2003). As a result, image collections are required for system evaluation and, to "ensure both reliability and validity of testing results, the content should be grounded in the reality of the image community who make heavy use of digital images in their jobs, as well as providing suitable testing material for a variety of techniques" (Jörgensen 2003, 267). Historical photographs are included in Jörgensen's list of testable image databases.

The object of this study was the archive site of Pictures of the Year International (POYi).[1] POYi is a renowned annual photojournalism contest that began in 1943. The image collection and contest managed by POYi is a program of the Donald W. Reynolds Journalism Institute at the University of Missouri School of Journalism. The archive site was born from a collaborative project focusing on the development of a site to promote images collected over more than sixty-five years. All images had been scanned in previous years or were originally digital. This project took place in 2007–8 between the University of Missouri School of Information Science & Learning Technologies (SISLT) and the School of Journalism.

The project was conducted in two phases. The first phase (December 2007 through September 2008) created a prototype of the new image

database (the archive site). Using open source technologies, including a content management system, the archive site was constructed by Thomas Kochtanek and his graduate students at SISLT. The main purpose of the collaboration was to build a site with dynamic Web-based design principles associated with end-user searching and browsing of indexed metadata. The focus was, therefore, to move beyond a linear version of static webpages to include elements of dynamic database-driven search-and-retrieval functionality, as well as to test new and exciting social computing features. The second phase assessed how well the metadata used within this new implementation met the needs of the audience, with the expectation that such an examination would enhance the quality of the metadata elements and the site's search and browse functions. To make this assessment, we addressed two key issues recommended by the Library of Congress's Future of Bibliographic Control report (2008). Within the context of this project, these issues include: testing the use of metadata elements for photojournalism images on the Web; collecting evidence to support system enhancement of the POYi archive site; and collaborating with end users on organizing and accessing photojournalism images. The third issue will be conducted in a later phase of the project. This paper, however, is the result of an assessment of the first two issues.

Literature review

Graham (1998/9) surveyed 60 art libraries in the UK, including the important areas of image collections, cataloguing and indexing practices, content-based image retrieval (CBIR) systems, and the use of images. Graham's study reported on the current management of image collections and techniques for image and video retrieval in the UK, and found a great variety of description models at the 60 art libraries. Eakins and Graham (1999) studied the state of the art in CBIR systems within the UK and submitted suggestions to government agencies, users, and managers of image collections, as well as to CBIR software developers. They recommended that all parties explore the possibilities that CBIR systems could provide to image management.

Machine-based image processing became an important component of image management in the last decade (Chen 2001; Chen and Rasmussen 1999; Eakins and Graham 1999; Jörgensen 2003). CBIR systems were devised to analyze and retrieve images on the basis of attributes such as

colour, location, and distance. However, CBIR systems remain fragmented and have favoured devoting more effort to development of interdisciplinary approaches (Berretti, Bimbo, and Pala 2000; Jörgensen 2003; Smeulders et al. 2000). These studies suggest that future research must focus on usage types, aims, purposes, knowledge domains, and human–machine interfaces. Thus, what is needed is an examination of the relationship between image-management-system functionality and end user image-seeking behaviour. In order to understand how to enhance the archive site's search functions from the perspective of an end user, this was our focus in the second phase of the POYi archive site.

Cooniss and her colleagues (Cooniss, Ashford, and Graham 2000; Cooniss, Davis, and Graham 2003) studied end users' visual information needs and search behaviours in the Visual Information Seeking Oriented Research (VISOR) project. In the first stage of the project, they found that contextual factors were important for both individuals and organizations, and that end users showed mixed reactions to digital technologies. In the second research stage, the research team developed a user-oriented evaluation framework for electronic image retrieval.

Enser and McGregor (1992) analyzed 2,722 image requests at the Hulton Deutsch Collection Ltd., which houses over 50 distinct collections of images, the great majority of which are negatives and prints. With approximately half a million engravings, woodcuts, drawings, cartoons, and maps, the combined size of the collections is about 10 million images. Enser and McGregor classified these requests into four categories: unique, unique with refiners, non-unique, and non-unique with refiners. The property of uniqueness is readily applied to a request for the visual representation of an entity, the desired, particular occurrence of which can be differentiated from every other occurrence of the same entity type (17). They discovered that nearly 70% of the queries were for a unique person, object, or event, and 34% of these queries were refined (mostly by time). These earlier reports and studies help establish the foundation of future efforts to extend image retrieval and metadata use to support end-user access to image-based content.

Emerging Web technologies have encouraged system designers to include functions that support social collaboration within information systems. End users can use these functions to provide feedback, share ideas, and organize information (Casey and Savastinuk 2006; Miller 2006). Social tagging (also known as folksonomy) is a popular Web 2.0 trend (Trant

2008). The ability to tag a digital object allows users to index online images using their own terms and keywords (Neal 2007). Currently, several photo-sharing websites, such as Flickr.com, provide users with the ability to manage and share their photos using tags. We intend to focus on social tagging in the archive site during the next phase of the project.

POYi image collection and the archive site

Pictures of the Year International was established in the early 1940s by the Missouri School of Journalism "to empower the world's best documentary photography, to provide a visual portrayal of society, and to foster an understanding of the issues facing civilization" (POYi, "About"). A prototype of the POYi image collection, the archive site, was developed by Thomas Kochtanek and his graduate students. The archive site contains approximately 38,000 photographs and offers basic and advanced search. It was built using Omeka, an open source content-management system sponsored by the Center for History and New Media and George Mason University and it builds upon or uses other open source software including Linux, Apache HTTP server, MySQL, PHP, and ImageMagick (Omeka, "Preparing to Install"). Omeka supports an unqualified Dublin Core database structure to "create complex narratives and share rich collections" and is "designed for scholars, museums, libraries, archives, and enthusiasts" (Omeka, "Home"). Although Omeka recently released version 1.0 of their application, the archive uses a previous beta version. The 1.0 release included a complete redesign of the database architecture.

The main and original POYi website is a static site where users are encouraged to browse the limited collections available by Year of Award. No search capabilities are present on this site. The archive site, which operates side by side with the main site, was developed as a prototype to test database-driven design features. Thus it supports both searching and browsing. The metadata were provided by the POYi staff. Given the enormous amount of data, which included many possible elements for over 38,000 photographs, and the limited time given to the project team to develop the archive site in the first phase of the project, the goal was to quickly map the provided metadata to the Dublin Core (DC) metadata elements.

Figure 1: POYi Homepage

Figure 1 depicts the opening screen, which includes Search and Browse Tips along with descriptions of the collection, including Items, Collections, and Exhibits. Figure 2 depicts the Basic and Advanced Search interface, which supports searching by specific fields, collection, type, and metadata element. Drop-down menus are included for ease of selection within those categories. Figure 3 depicts a typical search result and includes some of the fielded information, such as Publisher, Creator,

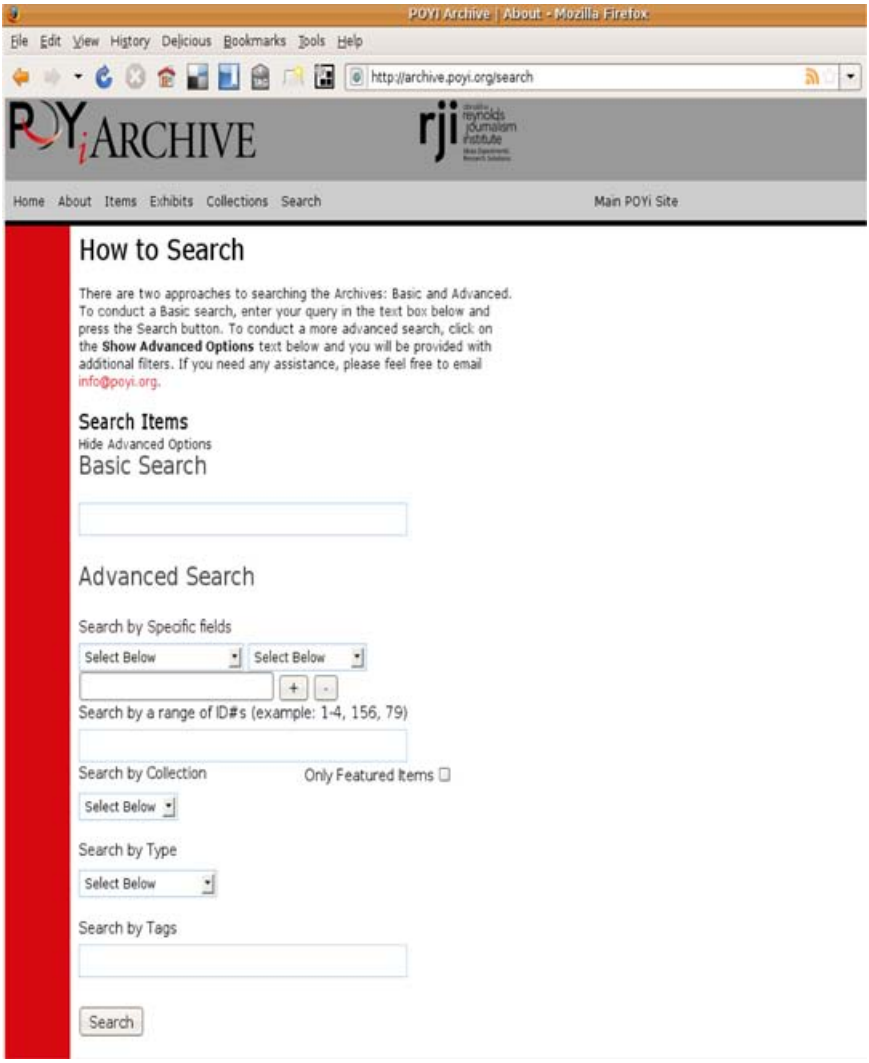

Figure 2: POYi's Search page

Relation, Spatial Coverage, Rights, and Source. Additional fields may be displayed by scrolling through the results page. Figure 4 depicts the metadata description of an image item.

The second phase of the project focused on elements from the research agenda recommended by the Library of Congress's Future of Bibliographic Control report. These recommendations are:

Figure 3: POYi's Search Results

1. To increase the efficiency of bibliographic production for all libraries through increased cooperation and increased sharing of bibliographic records, and by maximizing the use of data produced throughout the entire "supply chain" for information resources.
2. To transfer library efforts into higher-value activities, and in particular to expand the possibilities for knowledge creation by "exposing" rare and unique materials held by libraries that are hidden from view and are thus underused.

Figure 4: POYi's thumbnail image of photo with description

3. To position our technology for the future by recognizing that the World Wide Web is both our technology platform and the appropriate platform for the delivery of our standards, and to recognize that people are not the only users of the data we produce in the name of bibliographic control, but so too are machine applications that interact with those data.

The POYi collection provides a prime opportunity to examine the use of metadata elements for photographic images. In addition, the existing indexing approach for metadata elements allowed us to identify photo users' searching behaviours and needs, and to incorporate related subject analysis and future metadata standards. To accomplish this, we applied Jörgensen's (1998) attributes and 12 attribute classes to gain better insight into users' image search queries.

Jörgensen's 12 image attribute classes and 47 attributes

Jörgensen (1995) randomly selected 77 images from The Society of Illustrators 25th Annual of American Illustration, and asked participants to conduct three tasks: describing, searching, and sorting. In the describing task, 48 participants viewed six projected images and wrote a description of each. In the searching task, 18 participants were given two terms representing abstract concepts, such as happy or mysterious, and then were asked to browse the set of 77 images to find those relevant to the queries. In the sorting task, the same 18 participants sorted 77 images into groups for their own use as if the images belonged to their personal collection. She classified the verbal protocols into 12 classes of image attributes (table 1) in which 47 attributes were distributed. For example,

Table 1. Jorgensen's 12 classes of image attributes

Attribute Class	Description
Literal object	Named objects that are visually perceived, e.g., body parts, clothing
People	The presence of a human form
People-related attributes	The nature of the relationship among people, social status, or emotions
Art historical information	Information related to the production of the image, e.g., artists, medium, style
Colour	Specific colours or terms relating to aspects of colour
Visual elements	Elements such as composition, focal point, motion, shape, texture
Location	General and specific locations within the image
Description	Descriptive adjectives, e.g., wooden, elderly, size, or quantity
Abstract concepts	Attributes such as atmosphere, theme, or symbolic aspects
Content/story	A specific instance being depicted
External relationships	Relationships to attributes within or without the image, e.g., similarity
Viewer response	Personal reaction to the image

The People class had one attribute, People; and the Art Historical Information class had eight attributes, Artist, Format, Medium, Representation, Style, Technique, Time Reference, and Type (Jörgensen 1998, 174).

We used these sets of image attributes to structure an understanding of search terms used on the POYi photo database, the archive site. Chen (2001) studied 29 art history college students' image queries by comparing the features of their queries to those identified in studies by Enser and McGregor (1992), Fidel (1997), and Jörgensen (1995). According to Chen's study, Jörgensen's attributes provide more details for understanding end users' image queries. Therefore, we applied these attributes to the search terms logged from the archive site.

Research questions

We address several key issues recommended by the Library of Congress's Future of Bibliographic Control report (2008): testing the use of metadata elements for photojournalism images on the Web; collecting evidence to support system enhancement of the POYi archive site; and collaborating with end users on organizing and accessing photojournalism images. The last item will be addressed in a later phase of the project. Based on the first two items, our research questions are:

1. Do end users search or browse the collection to locate and discover images?
2. What are the most frequently used / least used metadata elements searched by users?
3. What are the common characteristics of search terms that end users select?

Methodology

We collected data using Google Analytics—a free Web analytics tool. The Web Analytics Association (2007) defines Web analytics as "the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage." In contrast to log analysis that collects data from the Web server, Web analytics collects data about the interaction from the client side (user's computer) and

therefore provides more accurate and specific data about end user behaviour. Dyrli (2006, 72) compared several Web analytics tools—VisiStat, StatCounter, ClickTracks, and Google Analytics—and recommended Google Analytics. Fang (2007) used Google Analytics to track online visitors' behaviours at a university law library's website and made beneficial changes to the site based on the data from Google Analytics. Google Analytics contains a set of tools and methods for analyzing user behaviour on a website (Nicholas, Huntington, and Jamali 2007; Ravid et al. 2007; Zuccala et al. 2007). Out of privacy concerns, Google Analytics provides a limited set of details about each visit. Therefore, detailed information about users, such as their IP addresses and what each IP address viewed on the site, is not available. This is a barrier to understanding an individual visitor's browsing and search activities. Web analytic tools integrate data collection with data analysis and reporting. The investigators installed Google Analytics on November 12, 2008. Variables examined were users' search keywords, visit length, viewed pages, and exit pages.

To use Google Analytics, a small piece of JavaScript code, supplied by Google, is inserted into the website's footer. When a page on the site is visited, Google Analytics records the actions. Google Analytics also collects user input from the onsite basic and advanced search forms. Since the archive site converts queries from these forms into URL query strings, and all site URLs visited are logged by Google Analytics, these URL query strings, which contain both basic and advance search terms as well as the metadata elements used with the advanced search terms, form the unit of analysis. Figure 5 lists examples of these query strings.

52.	/browse?search=walla&submit_search=Search	2
53.	/browse?search="Frank+lloyd+wright"&submit_searcl	1
54.	/browse?search="McGahee"&submit_search=Search	1
55.	/browse?search="post+office"&submit_search=Searcl	1
56.	/browse?search=&advanced[0][field]=&advanced[0][typ	1

Figure 5: Basic and advanced search string examples shown in Google Analytics

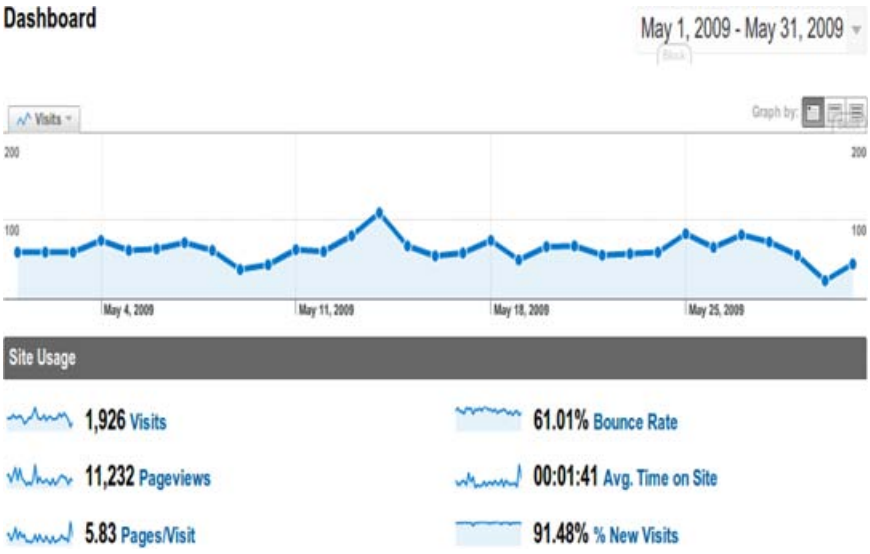

Figure 6: Google Analytics Main Dashboard

Google Analytics provides several functions, graphs, and timelines to analyze a rich array of data. The Dashboard is the first graph users see after accessing one of their accounts (figure 6). It provides information about the total number of visits, page views, pages per visit, bounce rate, average time on site, and percentage of new visits. The timeline and corresponding data may be adjusted by changing the dates. In addition to collecting data about user activities on the monitored site, Google Analytics is able to collect search terms from external search engines, such as Google and Yahoo, that lead users to the site. These search terms were not analyzed in this study.

Results and discussion

Research question 1: Do end users search or browse to find images?

We analyzed data collected between December 1, 2008, and May 31, 2009. There is some discrepancy in the data Google Analytics collects. When adding individual months, the total number of visits sums to 10,039. When using Google Analytics' Dashboard to exhibit data for the six-month range, the total is 9,998 visits. Since we collected data on a monthly basis, we decided to use the sum of the monthly numbers rather than the sum provided by the Dashboard.

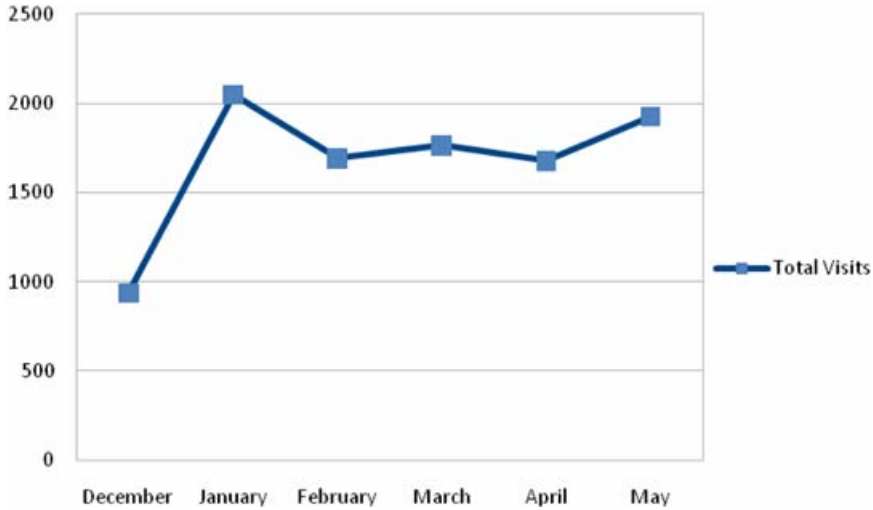

Figure 7: Visits on the site, December 2009 to May 2009

The total number of visits increased from the month of December to January but tapered off and held relatively steady during the following three months. May witnessed another spike in visits. The total number of visits reached 10,039 over the six-month period of examination. The median number of visits was 1727.5 per month, with a mean of 1673.17 per month. Figure 7 graphs the visits for each month.

To answer our research question, we performed several basic calculations. During the six months, the site received 10,039 visits and 53,739 page views per visit, so we calculated 5.58 webpages per visit (53,730/10,039 = 5.58). Since the site received 53,739 page views and 744 searches (basic and advanced), we calculated 72.23 page views per search (53,730/744 = 72.23). Thus, for every 72.23 page views, one search was conducted (1.38%). Or, of the 10,039 visits, since the archive site experienced only 744 basic and advanced searches, we calculated a rate of 13.5 visits to the site per search conducted (7.41%). Additionally, the average time on site for this period was one minute, 39 seconds. The numbers, therefore, suggest that most visitors to the site browse through a few pages and leave quickly, without conducting many on-site searches.

Research question 2: What are the most frequently used / least used metadata elements searched by users?

The archive site's search allowed users to conduct advanced search queries using the Dublin Core metadata elements. Only 72 advanced queries were conducted over the six-month period. Given this statistically insignificant number, our second research question remains unanswerable. The low number suggests a need to improve either the site's design or its search capabilities. However, when the metadata elements in advanced search were chosen, visitors used the Creator and Description elements most often (28 queries for each). The Collection element was the next most popular (10 queries), followed by the Title (9 queries) and Publisher (8 queries) elements.

Research question 3: What are the common characteristics of search terms that end users selected?

To answer our third research question, one researcher described the search terms collected from the archive site's on-site search page using Jörgensen's (1998) list of 47 attributes belonging to the 12 classes. Since the advanced search queries were so few and because the query strings were complex and may been difficult to interpret, we decided to classify basic searches only.

To help establish instrument reliability, a library and information science graduate student analyzed the search terms. Background information about the project and Jörgensen's attributes were provided to the student along with the archive site's search terms stripped of the other researchers' attributes. The graduate student repeated the process of assigning attributes to the terms independently and separately. Once the student had assigned attributes to the search terms, these attributes were returned to the researchers for analysis.

The analysis involved comparing the search terms by assigning a numerical code ranging from one to four. Code 1 was assigned to all instances where the researcher and the graduate student agreed on an attribute (1 = Agreed). Code 2 was assigned to all instances where the researcher and the graduate student disagreed on an attribute (2 = Disagreed). Code 3 was assigned to all instances where we believed both the researcher's and the graduate student's attributes could be applicable to a search query (3 = Both). Code 4 was assigned to those queries where

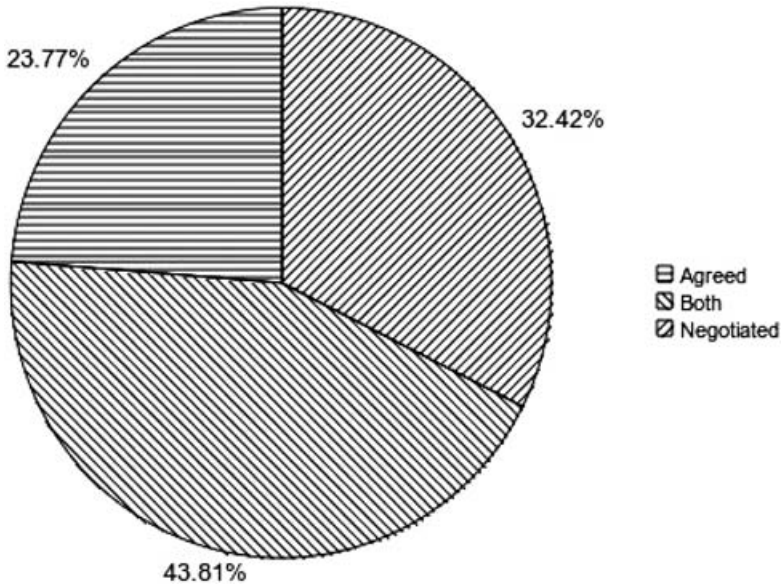

Figure 8: Comparing attributes between coders

we believed, after discussion, that only one person's attributes were applicable (4 = Negotiated). After this initial process, we re-examined the Disagreed terms and were able to settle on attributes for them. For all six months of terms, therefore, we found that we agreed 23.77% of the time, could assign both the researcher's and graduate student's attributes 43.81%, and had to negotiate between terms 32.42% of the time (figure 8).

The difficult issue we confronted in describing each of these search terms and in negotiating on attribute assignment resulted from having to interpret what the searchers intended as well as whether they were satisfied with their search as a result of using a term in their query. In order to clear at least a little of the fog, when names or vague terms were used, we queried the archive site or a Web search engine to see if we could better understand what a term referred to. Hence, if there was an object (e.g., the value of a DC element, the content of an image, etc.) on the archive site that literally mapped to the search term, then we felt comfortable applying a specific attribute to that term. For example, when we saw a specific query string such as *fenway*, which we believed referred to Fenway Park in Boston, and therefore at least could be described using

the attribute Setting, we queried the archive site (and for some terms, the Web) to see what instances of *fenway* existed on the archive site and whether our belief would be in any way validated. What we discovered was that such terms could be described easily with multiple attributes. In one result for the query *fenway* from the archive site,[2] one of us might assign the Description attribute to the term, whereas the other might assign the Setting attribute to the term. Since the term *fenway* appears in the photograph's DC description element, and the photograph is an image of two baseball players at Fenway Park, we believed both attributes were applicable, in a comparable way that a Flickr user may apply multiple tags to a single photograph or a librarian may apply multiple subject headings to a single document. For the purpose of this instrument, which is to provide the researchers with enough knowledge about the characteristics of the search terms used on the site to later enhance the site's search and retrieval capabilities, we believe such a method worked well.

In many instances the researchers had to negotiate on a query's attribute assignment (32.42%). In the majority of these cases (87% or 184 queries), the negotiation was solved by simply referring back to Jörgensen's attribute descriptions as well as searching the archive site. For example, for the search term *drezdzon*, one of us assigned the People attribute, whereas the other assigned the Artist attribute. Since Gregory Drezdzon is a sports and news photographer with two images on the archive site, we were able to conclude that this was most likely an Artist attribute and the original query was for this photographer. Jörgensen's People attribute would be applicable only if *drezdzon* was a person in an image (see table 1). For the remaining cases (13% or 28 queries), there was no overall pattern in how they were negotiated. Sometimes we negotiated on an attribute for contextual reasons. At other times we negotiated on an attribute because we believed one of us applied an attribute incorrectly. And at yet other times we negotiated on an attribute because one of us did not know what a query referred to but the other one of us had personal knowledge of the term(s). As noted, these last cases involved very few terms.

Since we were not analyzing the advanced search queries, total basic searches summed to 678. After removing the attributes that we negotiated out, we were able to apply 1,153 attributes to these terms. Following Jörgensen's list of attributes, search term attributes fall into two broad categories: Perceptual and Interpretive. Of these, 471 (40.85%)

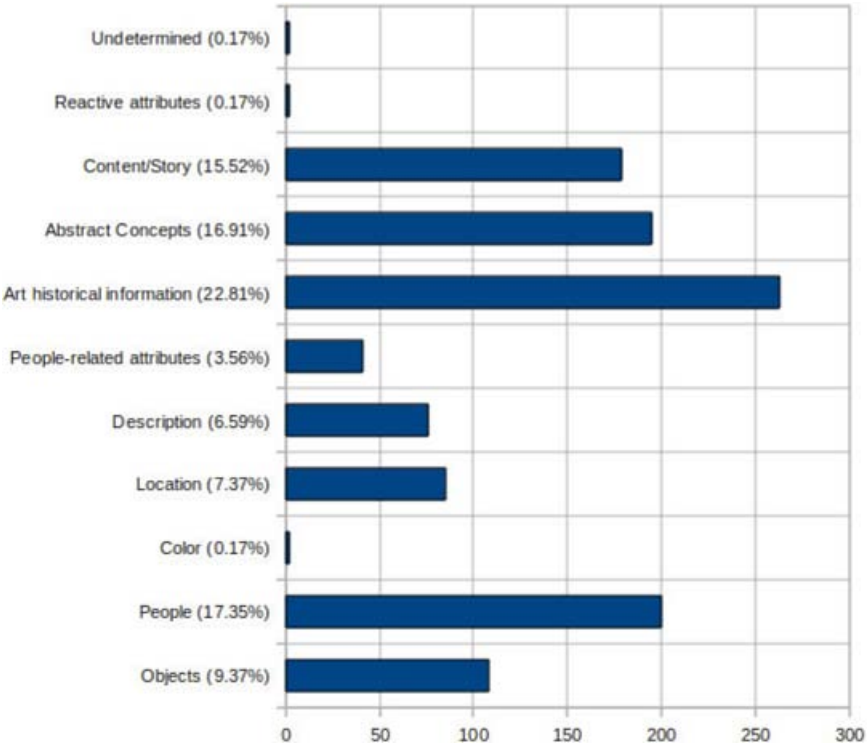

Figure 9: Portion of classes used and their percentages of total

attributes belonged to the Perceptual category and 680 (58.98%) to the Interpretive category (two search terms were vague enough to elude an attribute description). These categories are further narrowed into Attribute classes. Jörgensen supplied 12 Attribute classes, but after we applied attributes to our terms, we found we only used 10. Figure 9 lists the classes used and the percentage of attributes ascribed to each.

Attribute classes are further narrowed down to specific attributes. Although Jörgensen provides 47 total attributes, we used only 25. Figure 10 lists the attributes we used as well as the sum and the percentage of the total.

Implications

With the application of Google Analytics, the data collected easily answered the first research question, that most visitors browsed the

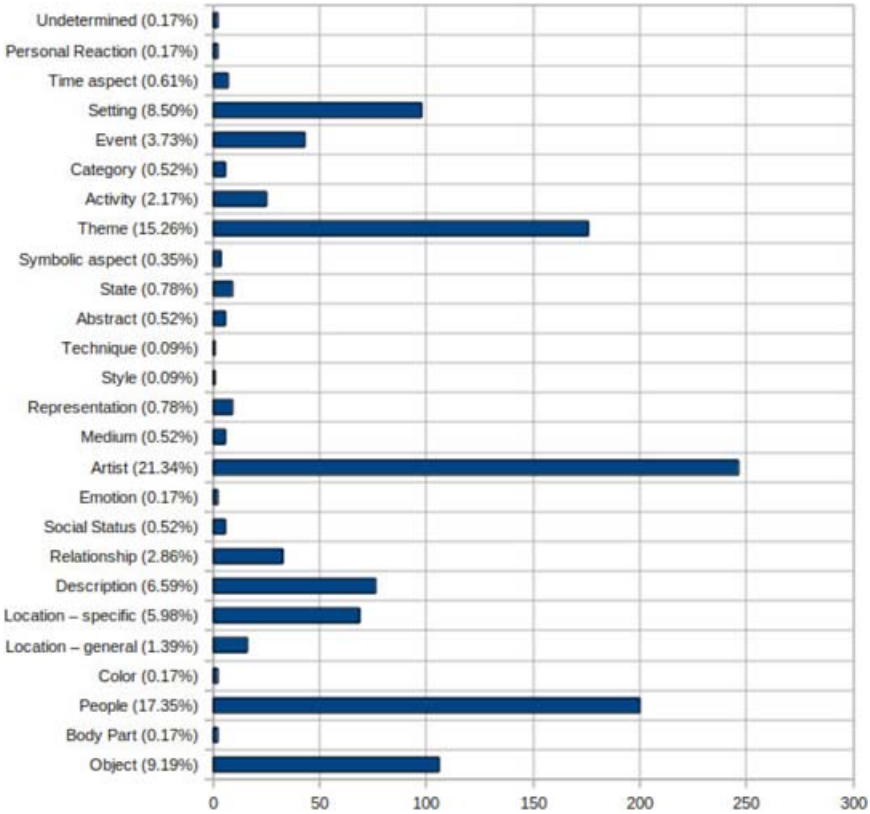

Figure 10: Attributes used with percentages and sums

POYi Archive site with nearly six pages per visit and conducted searches only 7.41% of the time. This suggests that the site might not engage visitors as much as desired, given part of POYi's mission "to foster an understanding of the issues facing civilization," and might currently function simply as a basic repository. While the site might not act as much more than an organized repository, our answer to this research question provides a baseline for improvement.

We were able to extract search terms from Google Analytics by examining the URL query strings created when visitors used the site's search function. However, while we were able to count the number of advanced search queries and could identify the DC elements used in these queries, there were few such searches and the query string was complex—often containing multiple search terms and DC elements, and requiring inter-

pretation—we decided that the process would open the door to many potential errors. Therefore we decided to abandon this analysis until further improvements to the site's advanced search can be implemented. Such improvements might include making advanced search less complicated for end users and more analyzable for researchers.

Regarding the characteristics of end users' search terms, we identified eight of the most popular (a total of 87.94%) attributes: Artist, People, Theme, Object, Setting, Description, Location-Specific, and Event. These attributes will be beneficial to us when we conduct the next phase of the project: adding social tagging functionality to the POYi archive site. These attributes might guide potential users in providing meaningful tags to the POYi images, so it might be worthwhile to suggest to visitors that they tag images using the above attributes as categories so that they may, for example, tag a photo within the Artist category, the People category, etc. In the meantime, we plan to use a quantitative analysis tool (e.g., NVivo, ATLAS.ti) to assist a consistent search term analysis.

Conclusion

The POYi archive site is an ongoing project as its image collection continues to grow. Findings from this study have provided us with directions for improvements and new system implementations. In the third phase of the project, the research team plans to study visitors' browsing and search preferences via an online survey. We also will seek to implement image tagging functions. This will require approval by stakeholders in the POYi content and collection. This is a collaborative effort, with the authors serving as both designers and evaluators for the resulting Web rendering, so it is expected that social tagging and possible other forms of end user input will be embraced and supported by content stakeholders and the design and development team.

Acknowledgements

Contributions in support of graduate research assistants by the Reynolds Journalism Institute are greatly appreciated. We owe many thanks to the anonymous reviewers and the graduate student who helped classify the search terms.

Notes

1. http://archive.poyi.org/
2. http://archive.poyi.org/items/show/34709

References

Berretti, Stefano, Alberto Del Bimbo, and Pietro Pala. 2000. Retrieval by shape similarity with perceptual distance and effective indexing. *IEEE Transactions on Multimedia* 2 (4): 225–39.

Casey, Michael E., and Laura C. Savastinuk. 2006. Library 2.0: Service for the next-generation library. *Library Journal* 131 (14). http://www.libraryjournal.com/article/CA6365200.html.

Chen, Hsin-liang. 2001. An analysis of image queries in the field of art history. *Journal of the American Society for Information Science and Technology* 52 (3): 260–73.

Chen, Hsin-liang, and Edie M. Rasmussen. 1999. Intellectual access to images. *Library Trends* 48 (2): 289–300.

Conniss, Lynne R., A. Julie Ashford, and Margaret E. Graham. 2000. *Information seeking behaviour in image retrieval: VISOR 1 final report*. Library and Information Commission Research Report 95.

Conniss, Lynne R., Janet E. Davis, and Margaret E. Graham. 2003. *A user-oriented evaluation framework for the development of electronic image retrieval systems in the workplace: VISOR 2 final report*. Library and Information Commission Research Report 144.

Dyrli, Odvard Egil. 2006. How effective is your Web site? Free online tools help measure site success. *District Administration* 42 (9): 72.

Eakins, John, and Margaret Graham. 1999. Content-based image retrieval. JISC. http://www.jisc.ac.uk/uploaded_documents/jtap-039.doc.

Enser, Peter G.B., and Colin G. McGregor. 1992. Analysis of visual information retrieval queries. *Report on Project G16412 to the British Library Research and Development Department*. London: British Library.

Fang, Wei. 2007. Using Google Analytics for improving library website content and design: A case study. *Library Philosophy and Practice* (June): 1–17.

Fidel, Raya. 1994. User-centered indexing. *Journal of American Society for Information Science* 45 (8): 572–6.

——. 1997. The image retrieval task: Implications for the design and evaluation of image databases. *New Review of Hypermedia and Multimedia* 3: 181–99.

Google Analytics. 2010. http://www.google.com/analytics/.

Graham, Margaret E. 1998/9. *The description and indexing of images*. Report of a survey of ARLIS members, 1998/9.

Jörgensen, C. 1995. Image attributes: An investigation (Indexing systems, retrieval systems, computerized). PhD diss., Syracuse University.

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

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

Library of Congress. 2008. On the record: *Report of the Library of Congress Working Group on the future of bibliographic control.* http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

Miller, Paul. 2006. Coming together around Library 2.0: A focus for discussion and a call to arms. *D-Lib Magazine* 12 (4). http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april06/miller/04miller.html.

Neal, Diane. 2007. Folksonomies and image tagging: Seeing the future. *Bulletin of the American Society for Information Science and Technology* 34 (1): 7–11.

Nicholas, David, Paul Huntington, and Hamid R. Jamali. 2007. Diversity in the information seeking behaviour of the virtual scholar: Institutional comparisons. *Journal of Academic Librarianship* 33 (6): 629–38.

Omeka. 2009. Homepage. http://omeka.org/.

Omeka. 2009. Preparing to install. http://omeka.org/codex/Preparing_to_Install.

Pictures of the Year International. 2009. http://www.poyi.org/.

POYi. See Pictures of the Year International.

Ravid, Gilad, Judit Bar-Ilan, Shifra Baruchson-Arbib, and Sheizaf Rafaeli. 2007. Popularity and findability through log analysis of search terms and queries: The case of a multilingual public service website. *Journal of Information Science* 33 (5): 567–83.

Smeulders, Arnold W.M., Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (12): 1349–80.

Steve: The museum social tagging project. 2008. Homepage. http://www.steve.museum/.

Trant, Jennifer. 2008. Social tagging museums: Research and reflections. *The NDAP International Conference, Taipei, Taiwan*. http://www.archimuse.com/papers/taiwan0803/trant-steve080319-taiwan.pdf.

The Web Analytics Association. 2007. Web Analytics 'big three' definitions, version 1.0. http://www.webanalyticsassociation.org/resource/resmgr/PDF_standards/WebAnalyticsDefinitionsBig3.pdf.

Zuccala, Alesia, Mike Thelwall, Charles Oppenheim, and Rajveen Dhiensa. 2007. Web intelligence analyses of digital libraries: A case study of the National electronic Library for Health (NeLH). *Journal of Documentation* 63 (4): 558–89.