Comparing Image Users and Uses with Web Analytics

Xin Wang
School of Information Science &
Learning Technologies
University of Missouri
101 London Hall
Columbia, MO 65211, USA
Email: xwbt8@mail.mizzou.edu

Thomas Kochtanek
School of Information Science &
Learning Technologies
University of Missouri
303 Townsend Hall
Columbia, MO 65211, USA
Email: kochtanekt@missouri.edu

James Bryce Borwick
School of Information Science &
Learning Technologies
University of Missouri
101 London Hall
Columbia, MO 65211, USA
Email: BorwickJa@missouri.edu

ABSTRACT
This article uses a technique called Web Analytics to examine the behavior of end users interacting with an online, web-based image database that is part of a digital library. A quantitative analysis was conducted to compare the behavior of two types of visitors accessing the database (new visitors versus returning visitors). Significant behavior differences of the two groups of visitors are manifested on four aspects: visits, pages/visit, average time on the site, and bounce rate. New visitors behaved like “tourists” on the site, while returning visitors showed much more engagement with the image collections.

Keywords
Image Information behavior, image digital library, web analytics, image information users.

INTRODUCTION
An increasing number of web-based image collections and retrieval systems have been developed offering users easy access to visual resources originally stored in libraries and museums and providing a variety of value-added system services for enhancing users’ experiences in searching, using, and manipulating images (Matusiak, 2006). User studies on image information systems have begun since the 1990s. A range of topics have been explored: image indexing (e.g., Jorgensen et. al, 2001), image attributes (e.g., O’Connor, 1999), image queries (e.g., Jansen, 2008;), image needs and uses (e.g., McCay-Peet & Toms, 2009;), etc. However, the issue of image information search behavior has not been adequately addressed (Jorgensen, 2003) except some pioneer studies such as Spink and Goodrum (2001). Additionally, traditional social science research methods such as individual interviews, surveys, think-aloud interviews, and focus-group interviews have predominated in the literature of image information behavior. With the fast-growing web-based image retrieval applications created by organizations (e.g., libraries, museums, and historical societies), these aforementioned methods fall short in allowing host organizations to monitor the actual behavior of the web visitors longitudinally (Weischedel & Huizingh, 2006). To improve these image retrieval systems, it is imperative to explore new methods and techniques of tracking users’ activities and discovering their preferences.

This study employed a technique called Web Analytics to study image information searchers’ behavior on the website of this image database. Using this approach, Web visitors’ behavior can be recorded, collected, and analyzed without significant human and financial investment. The purpose of this study is to examine whether new visitors (who have only one visit on the site) and returning visitors (who have at least two visits on the site) of an online image database behave significantly different in terms of four key performance identifiers (KPIs): visits, pages/visit, the average time on site, and the bounce rate. These four KPIs were core metrics used in past studies (e.g., Nicholas, 2006). The following is the descriptions of the four KPIs.

- **Visits:** Also called a “session” or “user session”. A session duration is the time from when a visitor logs on to the site to when s/he leaves.
- **Pages/Visit:** Refers to an instance of a page being loaded by a browser by each visit.
- **Average Time on Site:** Refers to the total time on site for all visits divided by the total number of visits.
- **Bounce Rate:** Represents the percentage of single-page visits or visits in which the person left a site from the entrance (landing) page.

METHOD
Context
The Truman Library Photographs Database is an image digital library project initiated by the Harry S. Truman Museum & Library in 2002. This goal of this project is to digitize and describe images in its collection of 100,000 historical photographs and make them available to the public via the web. The end-user can build a structured search or browse the collection. The collection can be searched by various fields (people, places, photographer, subject, parent collections, institutional creator, etc.) and by date range (start and end date). Users can also filter searches to exclude copyrighted photos from the results. Browsing is supported by subject, name, date, and thumbnail gallery.
Data Collection

The Web Analytics tool used for this project is Google Analytics which was installed on this digital library website during mid-February of 2010. Google Analytics is a “client-side data collection” approach based on the pagetagging technique that embeds one line of JavaScript code in the footer of each page on the website. The Advanced Segmentation function in Google Analytics was employed to extract clickstream data pertaining to selected visitor segments. To serve the purpose of this study, we selected two segmentations: new visitors and returning visitors. At the first phase of this study, approximately four months’ data were exported from Google Analytics and then incorporated to SPSS software for statistical analysis.

Data Analysis

A MANOVA (Multivariate Analysis of Variance) analysis was performed with visitor type as independent variables and visits, pages/visit, average time on site, and bounce rate as dependent variables. Figure 1 displays means with standard error bars for the four KPIs by visitor type. Visits generated by new visitors ($M = 254.05, SD = 73.09$) were more than that of returning visitors ($M = 55.19, SD = 19.01$); pages/visit of new visitors ($M = 3.63, SD = 1.03$) are less than pages/visit of returning visitors ($M = 11.56$, $SD = 7.25$); average time spent on the site (calculated by seconds) of new visitors ($M = 90, SD = 23.13$) are lower than that of returning visitors ($M = 310.93, SD = 145.27$); the bounce rate of new visitors ($M=60\%, SD=4\%$) are higher than that of returning visitors ($M = 45\%, SD = 7\%$). MANOVA was conducted with visitor type as an independent variable on the four key performance indicators (KPIs) as dependent variables to compare the visitor type effect on visitors’ web performance. There was a significant multivariate effect of visitor type on the set of KPIs, $F_{(4, 209)} = 296.418, p < .001$, which is a large effect ($\eta^2 = .850$). The follow-up univariate ANOVAs found significant differences on all four dependent variables: visits, $F (1, 212) = 741.924, p < .001$, partial $\eta^2 = .778$; pages/visit, $F (1, 212) = 125.387$, partial $\eta^2 = .372$; average time on site, $F (1, 212) = 240.884$, partial $\eta^2 = .532$; Bounce Rate $F (1, 212) = 300.168$, p < .001, partial $\eta^2 = .586$.

CONCLUSION

New visitors are an important sign of traffic in a web application. Returning visitors are similarly important, which are signs of “content-effectiveness”. In this exploratory study, we compared new visitors and returning visitors’ behavior based on four KPIs. The results indicated that visitor type was indeed a factor that influences visitors’ performance in this system. The results also indicated that new visitors generated more visits on this image database but they did not stay long on the site and did not explore much of the contents on the site. On the contrary, returning visitors showed significantly more interactions toward this image database, which implied that returning visitors are likely to have more “loyalty” to this site. Since returning visitors are very likely purposeful searchers, their high engagement with the site are probably caused by their interests or information needs. This finding inspires us to further examine “why” the significantly different behavior exists between these two types of visitors. Google Analytics has a series of content and navigation related functions such as “Top content”, “Content Drilldown”, “Site Search”, “Funnel Navigation” and “Site Overlay” that enable us to further explore what contents attract visitors to become “loyal visitors” and why new visitors are like “tourists” who stayed a shorter time on the site, viewed fewer numbers of images, and left the site without visiting any other pages of the site.

REFERENCES


