THE USE OF WEB ANALYTICS ON

AN ACADEMIC LIBRARY WEBSITE

_____

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

ANINDITA PAUL

Dr. Sanda Erdelez, Dissertation Supervisor

DECEMBER 2009

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE USE OF WEB ANALYTICS ON AN ACADEMIC LIBRARY
WEBSITE

presented by Anindita Paul,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Sanda Erdelez

Professor Thomas R. Kochtanek

Professor Denice Adkins

Professor Hsin-liang Chen

Professor Steven J. Osterlind

This is dedicated to my family. I thank my father, M.P Paul for his support and guidance through this long strenuous journey and has brought me this far. I thank my mother, Purnima Paul for her trust in me, in the decisions I make. I would also like to thank my brother Avijit Paul who has constantly inspired me and continues to do so in his simplicity and sincerity towards work. Without their love, support and inspiration this would not have been possible. Though far from me, I have felt you every day and every moment of my strife through the process of achieving this milestone. Thanks Mamma and Daddy.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my committee chair Dr. Sanda Erdelez, who has helped me through the process of conducting my dissertation study by guiding me and critiquing my work. She has inspired me to stretch my ideas, pushed me to think more and encouraged me to go that extra mile. Without her support and ideas this dissertation would have been rather difficult to complete.

I thank the members of my committee Dr. Thomas Kochtanek, Dr. Denice Adkins, Dr. Hsin-liang Chen and Dr. Steven Osterlind for their guidance, time and effort. I would also like to thank Dr. Mary Bixby for editing my work and Judy Maseles for allowing me access to the MU Library Google Analytics implementation. A lot of my understanding about research related work came by working with Drs. Joi Moore, Kim Thompson and Greg Alexander. I would also take this opportunity to thank Dr. John Wedman who has provided me adequate support to balance my time for dissertation.

The journey through this process can never be complete with the help and support of friends and colleagues. Gary Westergren has been instrumental in providing me the support and guidance throughout as a manager, friend and a well-wisher. His constant words of encouragement and complements provided a refreshing break to long hours of monotonous hard work. I would also like to thank Amy Adam for her patience in helping me with the paperwork.

Friends are always there to support for good and bad times, any number of words to thank them could not make up for their kindness, support, and time that they have put in. Borchuluun Yadamsuren, Dr. Pulak Ghosh, Dr. Sumana Chattopadhyay are one of those friends who cannot be thanked enough. The long winded discussion on my research over coffee, lentil soup, and walks around campus with Borchuluun has helped me immensely

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The instant availability of information has lead to evolving web user needs and changes in their information behavior. Users have increased options available to them when it comes to electronic information access. Most of the library users' experiences are shifting from traditional face-to-face interactions to the virtual environment. Libraries are considering user perspectives when evaluating library services and constantly developing and improving services for better access to their resources. This study focuses on the use of *web analytics* that captures and presents data concerning the online usage of the website for the purposes of academic libraries.

Traditionally, the evaluation of library services has taken an internal perspective that excludes the user (Nicholson, 2004). Most user studies in libraries are conducted using focus groups and surveys that have helped shift from the traditional internal to the external forms of evaluation. However, such a shift has emphasized demands on time and resources for making decisions.

User feedback through focus groups provides advantages in terms of the in-depth discussions (Elsevier, 2008; Starkweather & Wallin, 1999; Crowley et al., 2002). On the other hand, relying solely on focus groups may limit the number of users and cannot be generalized to the larger population (Starkweather & Wallin, 1999). Surveys can increase generalizability by acquiring information from the larger population (Tsakonas & Papatheodoru, 2006). However, survey instruments pose a problem when it concerns to how respondents understand and interpret the questions or recall what they did (Wang et al., 2000).

1.1   *Decision-making in Libraries*

Libraries are reaching out to patrons through electronic services. Caswell (2004) looked at the features provided by library gateways to explore the "potential of catalogs in providing

support as a database of record." However, in order to understand whether electronic services are useful to intended users, evidence of their actual behavior is important. Evidence would be required either to confirm the intended effect of the electronic services or to further recommend any changes to the services. Hiller and Self (2004) stressed the usability assessment of library websites that would help administrators to understand the functionality of the websites. They also recommend outcome assessment for library websites.

Libraries serve diverse groups of users who bring with them different experiences, knowledge and expectations. Library decision-making needs to be based on a review of the usage statistics of the library website to identify areas of concern or area that need improvement. *Evidence Based Librarianship* (EBL) for decision-making in libraries has efficacy (Lakos, 2007; Eldredge, 2006). Eldredge (2006) calls for a systematic stepwise procedure in the EBL decision-making. His proposed procedure involves reviewing research studies in *librarianship* and drawing evidence based on the findings of these studies. However, Nicholson (2006a) highlights the lack of research articles in *librarianship* that could power the traditional EBL. Lakos (2007) calls for library directors to think beyond traditional means and to consider analytics as a tool for the evaluation of services.

Hernon & McClure (1990) state the three purposes for engaging in evaluation are to 1) monitor the degree to which the organization or program meets stated goals and objectives, and to readjust the program; 2) to provide a "once only" assessment of how well the organization or individual programs meet stated goals and objectives; and 3) to monitor the manner in which organizational resources are allocated. They note that formative evaluation in libraries takes more effort than summative evaluation because formative evaluation involves stated goals and requires ongoing monitoring that involves regular data collection and analysis procedures.

With the shift from system to user perspective, libraries started to consider users' behavior and emphasized more user-friendly systems (Dervin and Nilan, 1986). The web has influenced the library users' online behavior. In order to meet the expectations of their users,

libraries are now building electronic services. These services need to be further developed and improved based on the users' actions on the website. There is a need for a more robust tool for libraries that would provide library decision-makers evidence to support quick, effective decisions.

1.2   *Information Behavior and its Measurement*

Human Information Behavior (HIB) literature has a rich collection of studies and theories that elucidate information retrieving behavior of users. These studies have implications for broad issues such as general information interaction of a user with his or her environment to specific instances such as user search behavior on the web when performing a task (Theories of Information Behavior, 2005). With the advent of internet, the information users' behaviors have undergone considerable change. New theories have been proposed that are specifically related to users' actions on the web such as *berrypicking* (Bates, 2005), *optimal foraging* (Jacoby, 2005), web information behaviors of organizational workers (Detlor, 2005) or World Wide Web information seeking (Turnbull, 2005). However, the notions developed on users' behavior through the different theories in the past have limitations in their generalizability to the larger population.

Crowley et al. (2002) found that academic library users are usually aware of the rich information that is stored in the library; nevertheless they have difficulty accessing it. One of the issues that the users confronted was the terminology used on the library website. The end-user satisfaction is affected because users have to deal with library terminologies directly without the mediation of a reference librarian. An important element of the traditional library system is that it was accessed inside the library, and the librarian was available to help the users facilitate their search process. The online systems can be accessed from any location with just a click of a mouse, but the assistance that is provided in a traditional setting is not available. Some online library systems, in accordance with the traditional systems, have provided online helpdesk or chat

reference for assisting users. Web users have learned to browse and search for information on their own.

Cockrell and Jayne (2002) pointed out that the user's lack of knowledge of the source that they are trying to find could hamper the users' search experience. For example, undergraduate students give up a search on the library website too soon and lack persistence in their information search process. The undergraduate students need less effort to conduct a search on the web using a search engine as compared to the library website. Information experience on the web is accounted for by the consistency and the affordances of the websites. Wang et al. (2000) concluded in their study that the cognitive style of the different users is the other influencing factor of user behavior.

Griffiths and Brophy (2005) further talk about the structural variations in branches of different disciplines and the inability of students to recognize it. According to their findings, students usually act based on incomplete information obtained through colleagues and acquaintances. Students' preconceived notions and instant reactions also affect their information seeking behavior.

In a more recent study Head (2007) found that students usually rely on the library website for their information needs and that they seek help from librarians and their professors. As the web is expanding to include more and more information, users' abilities to locate quality information are decreasing. In order to efficiently utilize time, students rely on confirmed quality and credible sources of information such as that obtained from the library.

However, students might compromise quality if it is difficult to seek information conveniently. Tsakonas and Papatheodorou (2006) found that usefulness and usability are two important aspects of a user's interaction. Students see quality information as useful, but usability of the website is important to get quality information. The authors suggested that users would appreciate more flexibility in control of information that they receive. Brantley et al. (2006) found in their study that flexibility at an interface level does not really make much of a difference as

long as users can get access to shortcuts to resources that reduce the number of different layers of navigation that would complicate access.

Studies reveal that user behavior is influenced by a number of factors such as the users' past experiences, knowledge of the search task, knowledge of the resources available and efficiency of tools available for accessing information. Given the evolution through which library electronic services are going, it becomes important to constantly monitor the users' activities on the website and to identify and correct issues that are detrimental to users' experience on the web. Libraries provide different electronic services through eJournals, online catalogs, online databases, inter-library loan and so on. The library website also acts as a medium for information of the services provided by the library such as hours of operations, copy services and other announcements related to it. These different elements of information and links to information sources make the library website more than just an information portal. The library website is crucial as a gateway that provides access to a great amount of quality information. Any information about users' behavior on the website would be helpful for not only researchers but also library practitioners.

1.3   *Research and Practice Using Web Analytics*

Web analytics addresses the library's decision-making needs and the lack of available tools for measuring information behavior in human information behavior (HIB) research. Once implemented, analytics provide a constant way to keep track of the actual behavior of users of the system by capturing the different metrics available at any point in time. The visual format in which the data is presented makes it possible to identify any patterns. In addition, analysis of the data presented could be done to understand users' actions on the web. The reports that are generated are based on huge amounts of captured data. Such reports provide the required generalizability as evidence for library decision makers to implement changes to the library services.

Using web analytics, researchers could provide managers a greater understanding of their services and users, and managers could provide researchers the means to support research activities. Buckland (2003, p. 677) calls library staff's attention to create a research agenda that includes a research question that is "non-trivial, intriguing to whoever is to work on them and [that are] expected to have significant consequences for practical decisions…and/or our understanding of our field." Further, Buckland (2003) also calls for important research (for libraries) that yields a substantial return, on a small investment, with little risk and in a short term. He further emphasizes on the meaningfulness of library services as one of the challenges that libraries face, and encourages researchers to provide new ways and techniques that address old problems.

Web analytics is widely used in commercial settings for making business decisions and improving the customer experience (Jacoby and Luqi, 2007; Sen et al., 2006; Srinivasan et al., 2004). Phippen et al. (2004) present a definition of web analytics that has emerged through its application to e-commerce and for-profit organizations. They cite Aberdeen group's definition of "advanced" web analytics as a tool for "monitoring and reporting of website usage so that enterprises can better understand the complex interactions between web site visitor actions and web site offers … [and] leverage insight to optimize the site for increased customer loyalty and sales."

The broad accessibility of the web makes it a desirable form for commercial entities to reach the target customers. Many companies have invested large resources into their web presence. A subsequent step for the commercial sector has been to evaluate how their websites can be improved to tap the web's full potential for growth and to generate more profits.

A major approach of evaluating the online services by the commercial sector has been through log analysis. Log analysis involves collecting server logs of users' activities on the website. Over time more advanced forms of log analysis have been developed and used by industry practitioners. The web analytics tool provides the server log data of users' actions in the

form of metrics and readily available graphs. The users of the academic library are primarily

students – undergraduates or graduates, and faculty, or staff.

Several past studies have focused on understanding users' behavior on the website (Tyler

and Ledford, 2006; Eirinaki and Vazirgiannis, 2003; Rubin, 2004). Prior to web analytics most

web analysis was based on server log analysis, and appeared to be a supplementary tool to

qualitative analysis (Borgman et al., 1996; Bracke, 2004). Web analytics is a step further from the

traditional log analysis in that it simplifies and presents the data in a user-friendly manner.

Researchers have used variants of log analysis in library studies. Log analysis,

specifically, has been used to understand patterns of user behavior on online catalogs and

databases (e.g., Lau & Goh, 2006; Yi et al. 2006) to understand users' online behavior. Bracke

(2004) and Zhang et al. (2004) have used web usage mining that captures data about activities

performed on the web in order to understand the academic library web related services.

One technique that is used in library studies is *Bibliomining* (see Nicholson, 2003, 2004,

2005, 2006a, 2006b). Thelwall et al. (2005) discuss the relationship between the fields of

Bibliometrics and Informetrics that cover quantitative aspects of information and recorded

information. Tague-Sutcliffe (as cited in Thelwall et al., 2005) defines Bibliometrics as "the study

of the quantitative aspects of production, dissemination and use of recorded information" and

Informetrics as the "study of quantitative aspects of information in any form." Thelwall et al.

(2005) also introduce the concept of Webometrics which is "the study of quantitative aspects of

the construction and use of information resources, structures and technologies on the web

drawing on Bibliometric and Informetric approaches." Bibliomining is a shift of Bibliometrics to

the online environment that combines data warehousing and data mining (Nicholson, 2006a).

According to Nicholson (2003), bibliomining is the application of statistical and pattern

recognition tools to large amounts of data associated with library systems in order to aid decision-

making or justify services. Web analytics also helps with pattern recognition of the logs of

visitors and can be said to be a part of Webometrics. However it has not been applied to the

library environment extensively. Bibliomining, which has been applied in the past, is data mining for libraries, whereas web analytics presents the data in a manner that makes it easier for librarians to interpret the web data.

Also, privacy of user information in libraries has been stressed in the past. Nicholson (2006b) talks about this advantage in bibliomining that upholds the privacy of the user. He adds that bibliomining data focus on materials accessed and the services used and not the identity of the user. Similar to bibliomining, web analytics also provides aggregate information about visitors of the website without violating their privacy.

An important aspect of using web analytics for the academic library website is easy accessibility of analytic data. Information about the users who visit the library website can be narrowed down to mostly academic users such as students – undergraduates or graduates and faculty. In some cases these libraries have separate branches that serve specialties like Law, Journalism, Engineering and other disciplines, which make it easier to closely monitor the academic users' behavior either by reviewing literature or conducting a qualitative study. The findings from these studies can then be connected to the patterns discovered from web analytics.

Pagano (2006) urged libraries to use web analytics. The necessity today is to make the libraries' electronic resources more user-friendly. The resistance faced by libraries, as mentioned earlier, is due to lack of skills and a work culture that forms an impediment to web analytics implementation. Pace (2004) encouraged libraries to use analytics for interpretation of numbers, which would help by adding on to the skill set rather than posing a challenge. Chi (2002) talks about how easy it is to interpret, and to identify bottlenecks on a website through visualization of usage data.

A limitation of web analytics is that the aggregate data reported tells 'what' but does not tell 'why' about the users' behavior. For instance, analytics may report that users are spending a long time on a particular web page, but analysis is needed to understand the reason behind such behavior. Not having context when interpreting the data may be misleading. Norguet et al. (2006)

state as a drawback of analytics that the data are not summarized in the format needed by library management for decision-making. Other concerns involve assessing the scalability of the techniques and the difficulty in handling evolving data (Nasraoui et al., 2005). Jasra (2006), in his comparison report of two web analytics solutions, stated that not all analytics solutions are 100% accurate. A trade off has to be made to get the benefits of the capabilities offered by analytics packages.

One way to increase the understanding of the information provided by web analytics is by relating it to the HIB literature that has formulated various theories and models of user behavior. These theories can be used to attempt any explanation of the patterns of user related behavior. Also, knowing the profiles of visitors who access the system helps in interpretation of their behavior. In the current study, the system under review is the academic library website of the University of Missouri. The academic library is accessed primarily by academic users. Literature on academic user behavior could further support analysis.

Library decision-making is often constrained by time and resources. Libraries sometimes do not have the required resources to do a full-fledged investigation that involves a well-designed research study. In some other cases library staff do not have adequate time to scan through the library literature for evidence needed for decisions that they need to make. Also, the literature available might not suit the specific situation of the library. At other times where there is lack of generalizabiltiy regarding users' information behavior, analytics provides the data required to further investigate a wider population. An attempt has been made in this study to provide a common ground through a web analytics implementation for both researchers and managers in order to be mutually beneficial. This study investigates what the metrics say about the user's behavior that could be used for library decision-making. Such a connection of metrics with user behavior would help when there is a time constraint, though the need for a supplementary study cannot be dismissed.

1.4   *Research Questions*

The research questions that guide this study are:

1. How can web analytics data inform library decision-making?

2. How can web analytics data be used to interpret the information behavior of the website visitor?

This study focuses on the users of academic library of the *University of Missouri* (MU). The web analytics tool used in the study is the *Google analytics* tool. Data were collected using an interactive group interview of six members of the MU library's web usability committee and a Google analytics implementation of the MU library. A mixed-method approach with two different types of qualitative analysis was used in this study. Qualitative analysis of interview transcripts and a second qualitative analysis of data obtained using Google analytics elicited the data for this investigation.

Chapter 2

Literature Review

The research goal of the study were to investigate the use of web analytics as a tool for library decision making, and the use of web analytics as a tool to understand users' information behavior on a website. The study uses approaches from the areas of *Human Information Behavior* (HIB), library services and practice, and web analytics to address these goals. The HIB approach covers how users' behavior in the library's online environment can be identified in the analytics reports. The library service approach takes into consideration the library setting, the services provided, and library practices that might affect interpretation of user behavior using analytics. The web analytics approach takes into account the applicability of web analytics reports to understand the online visitors' behavior in the library context. A library's service goal plays a crucial role in improving the online experience for its users. Libraries evaluate their services from time to time in order to meet their service goal. In order to ensure proper services the library should incorporate the use of advanced tools in the evaluation process. Businesses have used web analytics and developed it for tracking user activities in an online environment. Such a tool could benefit libraries in their evaluation practices. However, in doing so there are important considerations regarding the library environment, such as the library's purpose, willingness to accept such a tool, and feasibility of the metric interpretations for the library. This chapter starts with a review of library setting that relates to service evaluation, followed by studies from the information behavior literature, and finally by a discussion of web analytics and its use as an evaluation tool for academic libraries and a research tool for *Human Information Behavior* studies.

2.1   *Library Services Evaluation*

Many studies call for a holistic view for assessing the quality of library services (Nicholson, 2004; Saraf & Mezbah-ul-islam, 2002). These holistic measures ask libraries to look at the user's perspective and incorporate these perspectives in the decision-making process.

Eldredge (2006) discusses the need for a systematic process of evaluation by libraries through *Evidence Based Librarianship* (EBL) that would make the library decision-making process more efficient. He performs a literature review of studies in librarianship that use evidence for decision-making and suggests a stepwise procedure to include the professional experience of librarians to judge the relevance and appropriateness of the best evidence. He acknowledges the importance of research literature and states that the library profession has increased in its research knowledge-base. However, inappropriate research methods can affect the quality of the studies, and applying these methods in a different library environment can lead to conflicting results. Nevertheless, increase in knowledge-base also provides the rigor needed that cannot be found in traditional literature reviews. Applying such stepwise procedural decisions in specific circumstances might create problems. The best posited evidence might not be suitable to the political strategies or the specific cultural or social context. The process, which is based on assessing the best evidence and what is seen as important by the librarian, is abstract.

Nicholson (2006a) points at the lack of appropriate research that librarians could use as a drawback of traditional EBL. Also, the time taken to collect evidence sometimes results in lesser number of publications and hence reduces the power of traditional EBL. Eldredge (2006) acknowledges the drawbacks and hence calls for certain fair and truthful practices that should be followed to minimize the downsides of traditional EBL. Web analytics implementation has measurable data readily available for use. Libraries do not need to spend a lot of time collecting evidence like in traditional studies.

Lakos (2007) interviews 20 library directors in his study. In his paper on EBL management, he provided a snapshot of the issues that libraries face in trying to plan and make

decisions using data but not actually using the data systematically or effectively. He discusses the increased options available in the market for assisting decision makers. The "information economy" has made available an increasing range of innovative information services and new modes of resource discovery, new opportunities for resource and information delivery, and new modes of inventory management. The library management worried about the resistance to new technology due to financial and cultural concerns. Lakos calls for libraries to look beyond the traditional style of doing things and to consider ways and means to enhance their decision-making capabilities by using products that are available in the market.

Hiller and Self (2004) discuss in their theoretical paper the importance of operating libraries as "user- or customer centered" with data that can be collected and used for decision-making rather than relying on subjective impressions and opinions. Further, they stress the importance of user needs, information seeking and using behaviors, user satisfaction, and providing the organizational focus and support on positive user outcomes. They discuss the usability of library websites and the significance of websites in getting user feedback from a representative sample of the website users. Library websites are seen as providing increased usefulness to the users. An assessment of the functionality of the website plays an important role in explaining the efficiency of services provided to the library users.

Hiller and Self (2004) also discuss the importance of 'outcome assessment' that determines the impact on library services brought by service related changes. Besides, they state the importance of a 'robust support infrastructure' for libraries that would provide guidance and expertise for developing and sustaining data based decision-making. They caution libraries about solely using quantitative data for decision-making, and call for consideration of context that would help in more meaningful interpretation of data. The commercial sector has implemented web analytics for increasing their Return on Investment. E-commerce applications have also extensively used web analytics. Libraries have been more than just a service provider, as a place and source of information. With the shift from libraries to the web as an information source,

libraries need to reassess and reposition their services as a quality information source for the web users.

New technologies that are capable of evaluating the library infrastructure evolved with the advent of online information. Libraries have employed some of these technologies, such as tools that ask the users feedback of their online information experience, and other tools that study users' activities on the online interface when they perform tasks and so on. Libraries have started to apply certain measures to understand the online users' information seeking behavior. Nicholson (2006a, 2006b) talks about "bibliomining" as an approach to improve the traditional EBL, and urges libraries to use it in supporting practice and research. "Bibliomining" helps overcome the drawbacks of traditional tools by creating a data warehouse.

Nicholson (2006a, 2005) also discusses the need for libraries to consider artifact based evaluation, which can answer some questions that the library might have without going for hypothesis based research. According to him, artifact based evaluation in libraries can be compared to archaeology where inferences are made based on historical data, with the additional advantage of access to the individuals represented by the data. Nicholson stresses the need for a balance between the research and practitioner perspective in interpreting the data captured by online logs.

Further, discussing the capabilities of "bibliomining" data, Nicholson (2003, 2006a) underscores the importance of collecting data at a non-aggregate level and expresses concern with the inability to test for statistical assumptions of the aggregate data captured using "bibliomining." However, he also acknowledges the importance of maintaining privacy and suggests aggregation of data rather than looking at the individual log items. According to him, OLAP or Online Analytical Processing is a tool that would help in exploration and evaluation by extracting interesting patterns of selected variables without having to spend time on manually analyzing vast amount of data. Nicholson, in his paper, mentions visualization as a way to understand patterns quickly which would not be apparent otherwise. The paper emphasizes the

importance of striking a balance between internal validity for practitioners and external validity for researchers through the available collection of rich library data captured using "bibliomining." By providing aggregate data the privacy of the user is maintained. However, not allowing access to individual details might create issues with the external validity.

Nicholson (2006a) points to the time and cost concerns in implementing bibliomining projects. He also calls for libraries to standardize data collection across different vendor systems for easy matching. He suggests moving beyond a single library to a group of libraries which would enable cross examination of the evidence, though he also raises concerns about the data sharing capabilities between the libraries. According to Nicholson, identifying different research questions helps in understanding the usage of the library services. These research questions should be based on patterns of the systems that could be separated or integrated.

Research on the library user's perceptions and use of information sources has implications on the library management's decisions (e.g., Kim, 2006a, 2006b). However, since the expansion of the use of internet, a wave of technology innovations have rapidly changed online users' behavior and expectations. New technology tools need to be updated in order to be effective and efficient for the end-user, at the same time assessments of any new behavioral adoptions need to be constantly performed. Updated information on users' activities and behavioral adoptions on the website through latest tools and techniques would help library management remain proactive to users' developing and changing needs, and would help them to make better decisions that would help them to retain users.

2.2   *Users' Behavior in the Library's Online Environment*

Human Information Behavior has concerned itself with concepts such as people's information context, information needs, information seeking behaviors, patterns of information access, retrieval and dissemination, human information processing, and information use (Todd, 2003). Users' knowledge, experiences, preferences, attitudes towards online systems along with

the external influences such as the environment where they are accessing information, affect users' information behavior. Moreover, the web has enabled quick and easy access to information. With the influence of the web on people's everyday lives, information behavior of users of the system has undergone considerable change.

Studies have analyzed the library users' online behavior. Cockrell and Jayne's (2002) usability study employs think aloud technique on a library website with 49 subjects – 10 faculty, 10 staff and 29 undergraduates. They find that the knowledge of journal or magazine titles and the distinction between the two affect the users' behavior on library websites. Lack of knowledge about titles hinders their information seeking process. Library users looked at the OPAC as the only way to find information on a library website as they would do in a search engine. They were unaware or did not think of using any other online resources provided by the library. Whereas more experienced users were persistent in their search tasks, undergraduates who were less experienced gave up too soon. The page layout and terminology also posed problems with the search tasks. However, since the tasks performed by the participants in the study were not their own, it seemed that the subjects were not motivated enough to try the tasks as they would naturally do. There were indications of the lack of search knowledge of users, and their unwillingness to learn ways to successfully search. The study provides the authors with meaningful insight into the aspects of the website that need to be addressed.

Cockerell and Jayne (2002) also identify other findings about user's dependency on the OPAC, presence of external links on the library website, and other issues where users overlooked information available on the page. Users' unfamiliarity with the tasks at hand could have resulted in such findings. Capturing users' real behavior on a website can provide information on the users' natural activities and problems faced when performing these activities. Users' search knowledge can be understood by looking up the search terms that they use on the website for seeking information. By looking at the sources through which the users access the page, librarians can infer users' information seeking models. Web analytics report provides some of these

capabilities that enable capturing the actual users' real behavior on the website. These reports include information on the paths that the users take to get to a site. For example, a proportion of users who came directly to the website was referred from another website or used the search engine to come to the website.

Looking at students' information seeking behavior on a library website, Griffiths and Brophy (2005) conclude that students are confused about quality, and as a result, use fewer academic resources. The authors' review of studies points out that students are unable to navigate across branches of hierarchies. Because of their failure to find information, students prefer a simpler navigation structure that is offered by search engines. The authors provide an example of students who have restricted knowledge of the structure of branches. For instance, chemistry as a discipline may have an easier structure to follow than social science. They call our time the age of "information satisfying," when something is good enough for the purpose rather than seeking to optimize the result (p.551).

Griffiths and Brophy (2005), further, summarize that students find it difficult to locate information and resources; their use of academic resources is low; students may trade quality of results for effort and time spent searching. The visitors' use of search engines influences their perception and expectations of other electronic resources. The authors also point out that students base their judgment on incomplete information, reputation among colleagues and acquaintances, and preconception or instant reactions.

Library users come from all disciplines across all demographic profiles. Students comprise a vast proportion of the library's users. With the use of web analytics, libraries have a way to separate the user profiles by geographic locations, system properties such as browsers and/or Operating System and other system characteristics as color, resolution and so on; however they do not provide information on personal characteristics of the visitors such as their gender, age or user type but provide an aggregate count of the visitors. These visitors access the academic library as authorized users either as students, faculty or staff associated with the

academic institution. Additional information regarding the environment is necessary when interpreting the data obtained through analytics (Penniman & Dominick, 1980).

In a more recent study with upper division students, Head (2007) conducts discussion groups with students from humanities and social sciences. She employs content analysis of research assignment handouts used by professors for two years, and collected surveys from students. She concludes that students usually rely on library websites or course related websites for their information needs. In the process of information seeking, however, students would prefer guidance from librarians and their professors rather than doing a simple search through a search engine. According to Head, students value quality information, which goes against earlier assumptions of search engine preference over academic websites. A user who visits a library with an intention to seek quality information would be more persistent in getting the information. Regular users of the library know what to expect from the library. Thus, there is a mixed variety of users those who persistently look for information and others who do a quick search and leave the website, making it difficult to interpret server log data. Web analytics can provide an account of the pageviews of the different pages viewed by users of a website, the average time that the different users spend on a page, percent exit from a page, percent bounces on a page. These metrics can provide a general idea of the users' activities on that page; however they may still obscure some user related activities on the web.

Hargittai (2002) discusses the shortcomings of web use studies that do not take the user's context into consideration. According to her, "Although a recording of users' visit gives us information about what pages are viewed, it gives us no information on what type of information users are actually looking for and how satisfied they are with what they find" (p.1242). Relying only on analytics may lead to misinterpretation of user activities on the web site. There is a possibility of users leaving computers in the middle of a session they started and not closing out of it, which might skew interpretation. When making decisions, library administrators face time and manpower constraints, as well as lack of evidence. Libraries could use analytics to gather

evidence provided using actual numbers of real users' activities. The analytics reports can be further used to build instruments and guidelines for qualitative studies.

Tenopir et al. (2008) use a survey and think-aloud technique to study users' affective and cognitive behaviors when interacting with the ScienceDirect system. The authors infer that positive feelings of users were more associated with thoughts related to results and negative feelings were associated with thoughts related to system, search strategy and task. They also suggest that pausing behavior along with feelings and thoughts during the process can provide effective clues to problems that the user encounters. The authors measure the 'session length, pace and pause' in order to further identify temporal points to zoom in. Web analytics can measure the time a user spends on a webpage and/or the website and interpret whether the user finds the needed information. Depending on the nature of the page, a longer time taken infers obstacles to the service goal. For example, if the website has a list of pointers to the library resources, the user might be spending a lot more time scanning through the list in trying to search for the right pointer. On the other hand, if the page is informational, then spending a small amount of time on it may not be a good indicator of the quality of the page.

Brantley et al. (2006) study users' actions on a customizable library portal. The authors use task analysis method with eight prescreened participants. Their findings suggest that increasing the level of control by allowing the users to have their own selected options on the home page did not make a difference when there was lesser control. The users only cared about shortcuts offered by the portal that would enable them to navigate through lesser layers such as by putting a list of comprehensive resources on their personalized page. However, an overarching finding was that the library website needs to conform to common web features by including visual layout, shorter pages, fewer textual explanations, and making sections accessible with more understandable terminologies. Previous researchers have found that users usually do not read much on the web site when they see a text box (Crowley et al., 2002). Reducing the layers of access can simplify users' information experience. However, during the process of the search a

19

slight perception of difficulty in accessing information may change the course of users' action and they might opt out of the system. Web analytics needs to provide a way to capture the various small details of web access behavior by presenting the navigation path of the users through the website.

Wang et al. (2000) conduct an exploratory study about graduate students' factual information search on the web and showed what happens during the search using a process tracing method. They captured screenshots and verbal comments of the user while they were performing the actions and measured users' cognitive and affective states. The authors point out the lack of methods for collecting real time data. Though the process tracing technique is able to capture users' real time actions, there are some problems with vast amount of the data captured. For example, they were open to interpretation and had privacy concerns. The authors stress affordances of the website as playing an important role in understanding the information needs of a visitor to a website. Other aspects that add to the difficulty of understanding users' mental models while interacting with the web were the heterogeneous objects, poor interfaces and diverse web organization. Users would usually use search either randomly or based on prior experience without trying to understand the content scope and search mechanism.

Crowley et al. (2002) studied the users' perceptions of the library's web pages at Texas A&M University using a focus group technique. They found that users preferred a more integrated tool that can be easily used for research related information seeking. The users' lack of searching skills was apparent. Also, the need for an intuitive gateway meant moving from the traditional rules of library catalog and translating the catalog into an electronic version.

Further, Crowley et al. (2002) found that most of the users who come to the library website for information seeking are aware of the rich information stored in there but does not experience adequate ease in accessing the information, which is due to inconsistency and lack of understanding of the terminology adopted on the site. Moreover, the web users' experiences of electronic resources come from the more widely used web which shapes their understanding of

the library website. Users preferred more visibility for the catalog link and a 'What's New' column. Users' preferences for easy and quick access became obvious when they suggested that the popular databases should be brought to the top of the 'list of databases,' and that the library should provide more electronic resources that they could use more than print resources. Other preferences include, a more efficient online access page that has workable (not broken), faster links; they also would like to see more responsive staff who answers email queries, and provides help when needed. The inconsistency of the website design along with the incomprehensible terminology made its use complex and problematic. Users' expect more support from the library, which would lead to efficient usage of the online system.

Crowley et al. (2002) also find that some effort on the website for supporting the users in their information seeking process goes unnoticed. An investigation of how users are using the library website, whether they are using the available features on the website, would help libraries understand their users' behavior. The libraries need to know if users are utilizing the current features on the library website as they expect them to or do libraries need to rethink their web features.

Tsakonas and Papatheodorou (2006) survey the aspects of 'usefulness and usability.' The survey indicated a correlation between the two. They administered a questionnaire to 43 participants; the questionnaire was divided into three sections: usefulness, usability and qualitative data. The third part aimed to determine the magnitude of usefulness and usability aspects of a user desirable system. The authors conclude that both concepts of usefulness and usability hold parallel importance when evaluating systems. They identify certain system and content attributes that could affect interaction and hence affect work tasks. The authors infer that users value information relevance and flexibility, which enables them to control the quantity of information received. The usable attributes influential to the users are 'ease of use' and 'easy to navigate'. System performance also rates high on task completion. Some experienced users compromise on time to get their desired information.

21

Tsakonas and Papatheodorou (2006) base their study on the survey instrument, and draw inferences on what users think they would do as opposed to how they would actually act upon the website. Some of the situations that the users were asked to answer were hypothetical situations. Web analytics as a tool can provide glimpses of users' real online actions that can address usability aspects of the system. However, supplemental studies could investigate into the usefulness element of the services offered by libraries.

Nicholas and Huntington (2003) mention the advantage of studying a micro-setting when the data captured includes a wide range of selections. According to the authors, narrowing down logs to academic users provides robust information that transcends the traditional macro approach. The metrics provided by analytics presents information on the users activities on the website. Further investigation of these metrics can help explain user behavior on the library website.

2.3   *Web Analytics as a Tool for Measuring the Library Users' Behavior*

Web analytics is a relatively new tool that commercial industries have used. It has been predominantly used in an e-commerce setting to increase revenues. Using web analytics, companies can identify and improve problems that online customers face when they visit their websites. Some of the precursors of web analytics are E-Metrics (Dwek, 2005), Web Mining (Nasraoui et. al., 2005), Data Mining (Liu et. al.), and web usage mining (Bracke, 2004; Eirinaki and Vazirgiannis, 2003). Log analysis has been used in the past (Phippen et al., 2004), which involves analyzing raw server logs. Raw server logs consist of rows of data about user visit information, including the IP address, time of visit, pages visited, duration of visit and so on. Page counter is another measuring tool that records the number of visits on a webpage using a client-based approach. A web analytics application further processes the log files, applying statistical and data mining methods to extract patterns concerning users' navigation behavior, user

and page clusters, as well as possible correlations between web pages and user groups (Eirinaki and Vazirgiannis, 2003).

Researchers have used log files extensively for analyzing websites (e.g., Clark et. al., 2006; Chau et. al., 2005; Spink et. al., 2004), online databases (e.g., Ma, 2002; Nicholas et. al., 2000) and other online environments (eg., Jansen et. al., 2005; Peterson, 2000). They have stressed the use of logs, as logs enable an extensive quantitative analysis of the data sometimes complemented by qualitative analysis (Borgman et al., 1996).

Bracke (2004) also points to the importance of web usage mining on web logs to supplement surveys or focus group interviews. Clark et al. (2006) combine quantitative data obtained through server-side click-stream logs of user requests for pages with data obtained through questionnaire and direct observation of the users. Marchionini (2002), Cothey (2002) and Larson (1991) have used log analysis for longitudinal studies of information seeking and retrieval. Nicholas et al. (2005) mentions the importance of a comprehensive yet detailed mapping approach to analyzing data with log analysis that could relate the usage data to the user data. These authors assert that context and triangulation is an important part of analysis that uses server logs. Web analytics provides a way to extend log analysis in a more clarified visual form, hence neither requiring high analytical skills nor being time consuming.

There is evidence that libraries have been using log analysis to understand their users' online behavior. An analysis of transaction logs by Lau & Goh (2006) on a university OPAC exposed that simple and short query patterns with low Boolean usage lead to failed or inefficient search results. The variables that they analyzed were: length of the query, frequency and type of search options, and Boolean operators and its relation with search failures. Yi et. al. (2006) have also reported the use of query logs in studying databases. They studied the types of query log analysis in terms of the occurrences of single query terms, co-occurrences of words, and multi-word terms into ABC-Clio and PsycInfo databases. Their study further analyzes the relationship between the natures of the queries used with respect to the database in which they were used. The

authors find that multi-word terms can provide a better understanding of the users' behavior, and can help in categorizing the different user behaviors.

Through web usage mining, Bracke (2004) analyzes data of website visitors on an academic health sciences library website. The variables include: user type of affiliation; resource type based on the format of the resource sought; restrictions of access (for affiliated or non-affiliated users); referrer type for the URL the visitors used to link to the resource; website usage by browse; and website or catalog search or something else. Based on the analysis, the authors identified the website usage of on-campus users as more than average users. Those who were seeking journals would necessarily conduct searches on the website or catalog rather than browsing to a journal. However, the authors also point out variations based on the age of the user and affects of the web on users' behavior. They also point out other areas of improvement of the website such as search tool and resource positioning. The authors further stress the need for the library management to better understand how users seek each information format, which would help in making improvements to the website design.

Zhang et al. (2004) use web usage data in their study to understand information seeking habits of visitors in an academic library portal system. The authors stress the need for user studies to complement their findings. The data for the study are drawn from query-log analysis, and are based on "high end users" of the system only. The authors study the variables of user clicks/minute, time length of search, and seeking time per click in order to classify the users' search activities at different levels of seeking as undirected, semi-directed, or directed. Their findings indicate that the majority of their study participants were involved in undirected seeking, with the rest of the users falling under an approximately equal distribution of semi-directed and directed seeking.

Zhang et al. (2004) also find that each of the levels of seeking involved certain characteristics that defined it, such as for "directed information seeking" users tend to spend-- shorter seeking time per click, more average clicks per session (greater than 30 clicks), more

session time on an average, and users mostly monitor and browse. "Semi-directed seeking" users tend to spend approximately 3 -9 minutes of seeking time per click, moderate number of clicks per session (15-30 clicks), their average time taken per session is between 20 to 40 minutes, and they tend to use all modes of information seeking defined by the authors (start, search, browse, monitor, extract, request service, and communicate). "Undirected seeking" involved use of all modes as in semi-directed information seeking; the seeking time is higher than 5 minutes per click with average clicks per session less than the other two, and the least time taken for each session. The study reveals the usefulness of web usage data in providing detailed insights into the different characteristics of information-seekers.

Web analytics has the advantage of not having to deal with enormous data. Web analytics packages available today provide increased functionality by presenting data in a visual format (Tyler and Ledford, 2006, p.7; Eirinaki and Vazirgiannis, 2003), eliminating the need to deal with transaction logs that can complicate interpretation of patterns of web users' behavior. Pagano (2006) states the importance of commercial web analytics tools for libraries that would help them give their websites the "voice of the customer". Even earlier, Pace (2004) urges libraries to adopt the different web analytics tools that commercial set-ups use for interpreting numbers.

As mentioned before, use of web analytics as a tool has been recognized in businesses that seek to improve their internal as well as marketing productivity through an understanding of the user (Jacoby and Luqi, 2007; Sen et al., 2006; Srinivasan et al., 2004). Srinivasan et al. (2004) stress the importance of task based web statistics over the URL-based statistics to analyze the business value of a corporate intranet portal application. A review of the exact location from which a user exits a page can provide useful information on concerns with that page, especially when it is not considered as a destination page (Rubin, 2004). Libraries have a different motivation compared to e-commerce websites, but some of the goals of e-commerce websites can be translated to libraries. Similar to e-commerce websites, libraries want to provide services to the users. Libraries want their users to fulfill their intended task on the website, and have an

enjoyable experience. Libraries also want online visitors to come back to their website, which will increase *dependability* on library websites just as e-commerce websites would like to increase *retention*.

Looking at some of the possible drawbacks of using web analytics, Norguet et al. (2006) see the detailed representation of data by web analytics tools as lacking in aspects such as representation of data. Organizational top levels personnel are more interested in summary and conceptual reports that they can use for making decisions. Web traffic data are considered lacking in terms of assessing the scalability of the techniques and the difficulty in handling evolving data (Nasraoui et al., 2005). Sen et al.(2006) propose a new system of approach using web analytics called the "web forensic framework" that stresses different levels of analysis providing a structure based approach to understanding web traffic that can further clarify the data reported.

Varied web analytics solution choices are available in the market that mine and analyze web logs for research purposes (Eirnaki and Vazirgiannis, 2003, Pace, 2004). Jasra (2006), in his comparison report of two web analytics solutions, states that not all analytics solutions are 100% accurate. A trade off has to be done to get the benefits of the capabilities offered by analytics packages.

There is not much research that studies the effectivity or use of Google analytics as a tool for library websites. Fang (2007) uses Google analytics for evaluating library website content. In another study, Fang and Crawford (2008) apply analytics on the usage of the online catalog of a library website and provide recommendations based on the reports. Some of the features they use for understanding the users' behavior were keyword comparison, visualized summaries, trend reporting, defined funnel navigation, content by titles, site overlay, visitor segmentation, and data export. They measure a remarkable change in user traffic after the redesign based on the results obtained. Jasra (2006) compares Google analytics with VisiStat and infers Google analytics as having an edge, in terms of the appeal of the reports it generates and the customizable views. According to Jasra, a probable concern for Google analytics is the lower numbers reported

compared to VisiStat; he also states the shortcoming of analytics solution lies in its inability to be completely accurate. Implementing Google analytics for library websites would provide for libraries a first step to using web analytics as a solution to measure users' behavior on the library's online resources and help support any decisions at the management level.

The need of libraries for enhanced decision-making for its online system calls for the application of tools available in the market. With developing online access, server-side data of online users is readily available. However, interpretation of logs of user data can be difficult and time consuming. Commercial sector has successfully used web analytics. Some studies point towards drawbacks of web analytics tools such as restricted access to the techniques of data collection and interpretation of data reported. There still is a lack of research on the use of web analytics for information behavior research and library practice.

Chapter 3

Research Methodology

The data for this study were collected using a mixed method approach. The mixed method approach included two types of qualitative analysis on the interactive group interview and the data obtained using web analytics. This chapter provides a detailed account of the research design, sample, data collection, and data analysis followed in the study. The chapter also presents the steps that the researcher applied to ensure research quality. The research questions for the study are as follows:

RQ1- How can web analytics data inform library decision-making?

RQ2 - How can web analytics data be used to interpret the information behavior of the website visitor?

RQ1 attempts to understand the Library's impression about the MU library's Google analytics implementation and past use of it as a tool for informing the Library's decision-making. RQ2 attempts to understand how analytics metrics and features can be applied to interpret users' behavior. Inputs from RQ1 were considered to inform the use of analytics for addressing RQ2.

3.1 Research Design

The current study is based on a naturalistic inquiry that follows a mixed method approach using two kinds of qualitative analysis. The study follows a sequential design with an interview followed by interpretation of data obtained by analytics. The nature of the study to investigate analytics in a library setting by interviewing the Library's web committee was seen as an attempt for a naturalistic inquiry. Lincoln & Guba (1985, p. 39) mentioned that one of the characteristics of naturalistic inquiry is to collect data in a natural setting that suggests realities in totality that cannot be fragmented.

The current study was conducted in the University of Missouri (MU) library setting. The analytics tool studied is provided by Google and was already, in a simplified way, implemented by the Library. In the naturalist setting the researcher mostly prefers qualitative over quantitative methods because they "are more sensitive to and adaptable to the many mutually shaping influences and value patterns that may be encountered" (Lincoln & Guba, 1985, p.40). In this study, the researcher uses interviewing as a qualitative tool followed by a qualitative exploration of analytics reports.

The study was conducted in MU's main Library. There were three back-to-back stages in the study. A description of each stage follows.

Stage 1 included a preliminary analysis of analytics data captured for the Library website for Spring 2008 semester. The purpose of this stage was to provide the MU library committee an insight into the data provided by Google analytics. Another purpose of the reports was to provide context for discussion about the Library's analytics implementation. The guiding questions used by the researcher to interpret analytics were based on Kyrnin (n.d.):

- Do most people visit at specific times?

- Which pages are the most popular?

- Do your readers browse more than one page before leaving?

- What is the average length of time your readers stay?

- Do your readers come from search engines?

- What pages are primary exit pages for your site?

- Who are linking to you?

Stage 2 consisted of doing an Interactive Group Interview involving presentation of stage 1 report to the Library web committee and solicitation of their feedback about the usefulness of web analytics for the Library purposes. Power point slides were used to display the analytics reports during the meeting with the web committee.

The usefulness of the *interactive group interview* technique was to engage participants in the discussion. Patton (2001, p. 400) suggests this technique is best used for situations which involve "program staff or clients as colleagues in… program evaluation…making it interactive and cooperative rather than one-sided and antagonistic." The technique is intended to be non-intrusive with the involvement of stakeholders, developer and the evaluator present as a team and meant to be reactive (Tikunoff, as cited in Patton, 2001). The interview used to generate an open-ended conversation with librarians to collect their views on the Library's analytics implementation topics and analytics use for Library decision-making. Patton (2001) cites Tikunoff's experiences in using such an interactive technique and mentions that such interactive designs may increase the validity of data and reduce reactivity by making evaluation more visible and open, thereby making participants or clients less resistant or suspicious.

Apart from the above, Lincoln and Guba (1985) provide a sketch of the characteristics of a naturalistic paradigm which stresses the importance of context in a 'natural setting.' They use Willem's definition that suggests that naturalism is a "function of what the investigator does" and provides two dimensions that an investigator's activity can be described by the "degree of the investigator's influence upon, or manipulation of, the antecedent conditions of the behavior studied, [and]…the degree to which units are imposed by the investigator upon the behavior studied (p. 8)." A second important characteristic of the naturalistic paradigm involves the 'human instrument' that involves gathering data from the inquirer as well as other humans involved. More precisely stated by Lincoln and Guba (1985, p. 39):

> N [the naturalist] elects to use him- or herself as well as other humans as the primary data-gathering instrument…because it would be virtually impossible to devise a priori a nonhuman instruments with sufficient adaptability to encompass and adjust to the variety of realities that will be encountered; because of the understanding that all instruments interact with respondents and objects but that only the human instrument is capable of grasping and evaluating the meaning of that differential interaction.."

Lincoln and Guba (1985) mentioned other characteristics of an 'operational naturalistic inquiry' some of which are (p. 39) - inductive over deductive data analysis, an emergent design

rather than a preordinately constructed design, negotiated outcomes with the human sources from whom the data has chiefly been drawn (and hence reconstructed their construction of meaning) and the tentative application. He stated that tentative inquiry are "tentative about making broad application of the findings because realities are multiple and different…[and] may not be duplicated… The extent to which the findings may be applicable depends upon the empirical similarity of sending and receiving contexts (p. 42)".

Stage 3 involved doing a detailed analysis of the analytics reports for interpreting user behavior. In this stage inputs from the interactive group interview of the MU library's web committee was considered to select the various metrics from the MU library's Google analytics implementation. These metrics were further investigated to understand the Library website visitors' activities. Specifically, the meaning of the metrics was interpreted in terms of their representation of the users' behavioral elements. Interpretation was also informed by the literature on the library users' online behavior.

The interactive group interview was conducted with six members of the Library's web usability committee on August 22, 2008. The interview was conducted within the MU library premises. The Library web usability committee is one of the three committees under the MU library's web advisory group that drives the Library's online effort. The other two groups are the content coordination group and design group. The web usability group is made up of eight members from the MU library and the branch libraries. The details of the members profile is mentioned in Table 1. The web usability group will be interchangeably referred to as web committee or MU library web committee in this study.

Table 1 -*Description of Respondents of the Interactive Group Interview*

	Gender	Role	Experience	Any past use of web analytics
Respondent #1	Female	Head Information Services, Health Sciences Library	28 years	Yes
Respondent #2	Female	Library Information Specialist, Engineering Library	8 years	No
Respondent #3	Male	Catalog Librarian; Chair Web Usability Group	12 years	Yes
Respondent #4	Female	Head, Catalog Management Unit, Catalog Department	35 years	No
Respondent #5	Female	Engineering Librarian and Web Development Administrator	28 years	Yes
Respondent #6	Female	Instruction, Subject Bibliography/Reference; serves on web usability committee	10 years	No

The Google analytics tool was implemented in the Library in March 2007. This tool is available for free from Google (Google, 2007). It gives libraries an affordable solution that could be used for a comprehensive understanding of the online user.

3.2   *Data Collection and Analysis*

Data were collected using 'Interactive Group Interview' and Google analytics. Description of each method and the analysis procedure is explained below.

3.2.1   Interactive Group Interview

In stage 2 the interactive group interview of the Library's web usability committee was conducted with six members. Discussions in the interactive group interview were centered on the analytics reports from stage 1. The researcher recorded the interview sessions using MORAE by Techsmith Corporation. MORAE captured the display screen of the computer as well as a view of the interactive group interview participants seated around the table. Capturing the picture of the interview setting allowed keeping a track of the context of the discussion with the Library web committee.

The nature of this session was open-ended, intended to generate a conversation. Some of the guide questions for the interview were as follows:

1) What was the last time any changes to web analytics tracking code were made and why?

2) When was the last time a web analytics report was generated and/ or was viewed?

3) What other uses have been made of the web analytics implementation by the Library in the past, if any?

4) What are the specific metrics from web analytics that have been considered for use and how have they been interpreted in the past?

5) Have there been any implementations of the web analytics findings?

6) What are the typical activities that are carried out by the web administrator on the Library website?

7) Does the report presented (from stage 1) answer the questions that are asked by the Library management regarding the electronic services provided by the Library?

8) Do you see the analytics report as providing the required evidence in the stage 1 report for implementing any changes to the website?

9) What, according to you, are the weaknesses of the report?

10) What are the deficiencies that can be corrected by web analytics implementation (in other words, what tracking codes can be changed in the analytics implementation)?

The data collected using the *interactive group interview* technique with the six subjects were analyzed qualitatively. Different grounded theory techniques were applied for analysis. The researcher assembled the raw data from the interview, transcribed the user feedback, analyzed the transcripts, and kept extensive records from the reflections of the interview. Further clarifications were made with the members as and when needed. Patton (2001) accounts the importance of such extensive records that pull together and organize voluminous data into a comprehensive, primary resource package.

The transcript of the group interview was carefully phrased to retain its original meaning, broken into simple sentences, and analyzed (Appendix B). The researcher then identified participants' statements that needed more clarification. Some of these clarifications were based on the Library's use of web analytics while others were to clarify the meaning of certain statements. The members clarified some of the sentences that were rephrased from the main transcript in order to retain the original intended meaning. For example, users' statement– "We have used it to see what are the higher [rates] per visit. What else the users are coming through to see, so we could prioritize resources and make sure they don't have any problems…" was rephrased into –Libraries have used analytics to see –what are the visit rates, what parts of the website have higher pageviews, what are the top areas that the user visits on the website, and

libraries would like to know what problems do the users face and try to solve those. A question to ask the member was – "What are the specific metrics were you referring to by higher rates per visit?" Also, a set of analytics metrics that applied to specific statements were assigned by the researcher. The metrics were also assigned to assess its use in interpreting user behavior.

The qualitative analysis consisted of a series of different coding techniques. At first, the researcher used open coding on the qualitative data collected through the interview of the Library committee. Open coding involved doing naturalistic inquiry as detailed by Lincoln and Guba (1985). In the open coding the researcher categorized the rephrased sentences primarily based on intuitiveness of the feel-alikeness, and look-alikeness. The categories that were formed were generic and vast, to capture all the characteristics of the items coded under that category. In other words, the items under the respective categories were based on a broad and general relation with the categories they were placed under. The researcher further verified the categorization by constant comparison between the items under the main categories. During comparison, the researcher looked at each item and its relation with the other items under the category. A memo was stated for each of the categories defining the characteristics of the category provided in Appendix C.

Appendix D provides an example of the classification of the statements under different theme categories, and subthemes. The categories formed were:

- Decisions about services

- Decisions about web design

- The Library's interest in user behavior

- The Library's management and function

Specific issues that the Library encountered with web analytics were also documented.

Since the first level categorization was based on intuitiveness and look-alikeness, further sub-categorization of data was conducted in order to make close-bound connections between the

category and subcategory. A positive aspect of grounded theory data analysis technique lies in its flexibility. Charmaz (2006) views grounded theory as a set of principles and practices rather than prescriptions or packages. She further stresses how grounded theory methods can complement qualitative analysis methods "rather than stand in opposition to them" (p.9). Since the first round of analysis seemed to have left gaps in explaining the data collected, a further analysis was performed. Charmaz (2006, p. 48) states,

> Initial grounded theory coding can prompt you to see areas in which you lack needed data. Realizing that your data has gaps or holes is part of the analytic process. It is inevitable when you adopt an emergent method of conducting research….The advantage of grounded theory strategies is that you may learn about gaps and holes in your data from the earlier stages of research. Then you can locate sources of needed data and gather them.

The researcher did a second level categorization where closely bound sub-categories were formed. An instance of such a sub-categorization is the statements categorized under *decisions about web design*. Two user statements under this theme – "It might or might not be necessary to link pages from the gateway as users could be bookmarking certain pages" and "which links are redundant," were further sub-categorized under *restructuring website content to improve navigation*. Another statement on "What were the search terms people inputting" was subcategorized under *providing a satisfying search experience*.

The sub-categorization provided better identification of the rephrased respondent statements with the sub-categories. Charmaz (2006, p. 48) talks about the importance of flexibility in grounded theory technique and states that "Codes are [also] provisional in the sense that you may reword them to improve the fit. Part of the fit is the degree to which they capture and condense meanings and actions. Compelling codes capture the phenomenon and grab the reader."

The researcher tried to keep an account of similar subcategories so that the similar items could be identified at the end of the second level categorization. Similar subcategories were then grouped irrespective of the main categories they belong to. For example the subcategories

popular parts of the website and a similar subcategory *significance of parts of the whole website* were grouped together. A further assessment was made in order to make the subcategories the same, as the researcher decided that the sub-categories *popular parts of the website* and *significance of parts of the whole website* can be changed to the subcategory *providing for specific user needs*. Such similar sub-categories were then reviewed in the whole data and readjusted for its wording in order to make it identical to the other subcategories.

Axial coding was done in the next step. Strauss and Corbin (1990) define axial coding as "A set of procedures whereby data are put back together in new ways after open coding, by making connections between categories…[by] utilizing a coding paradigm involving conditions, context, action/interactional strategies and consequences (p. 96)." Further Strauss and Corbin suggest that researchers engaging in analysis alternate between open and axial coding. The authors talk about the 'Paradigm Model' that provides a way for linking subcategories to their category in a set of relationships. The relationships specified include causal conditions, phenomenon, context, intervening conditions, action/interactional strategies, and consequences. The axial coding that was used in the current study followed these steps (Strauss & Corbin, 1990)

a) The hypothetical relating of subcategories to a category by means of statements denoting the nature of the relationships between them and the phenomenon.

b) The verification of those hypotheses against actual data.

c) The continued search for the properties of categories and subcategories, and the dimensional locations of data (events, happenings and so on) indicative of them.

The researcher drew relations between the categories and subcategories. These relationships were further verified based on the content of each of the categories and subcategories. The specific dimensions (Strauss and Corbin, 1990) of the relations were underlined in the data (Appendix E). According to Strauss and Corbin (1990), dimensions in axial coding are locations of characteristics of a category. Memos were further built upon their initial descriptions based on the relationships between the categories and subcategories as illustrated in

37

Appendix C. Category of themes was then grouped by the subcategories. Finally, the researcher described the analytics features and metrics, where applicable, that could be used to address the dimensions under each of the subcategories along with the coded records.

3.2.2 Google Analytics

In the stage 3 of the study, web analytics metrics were selected based on the data from the *interactive group interview* with the Library's web committee, conducted in stage 2. The Library committee's needs for decision-making and its interest to be aware of certain aspects of library users' behavior were taken as a rationale for interpreting the metrics for users' behavior. The interview discussions provided meaning to the analytics interpretation in the MU Library context. Google analytics was used as the web analytics tool for this study. Google Analytics "is the enterprise-class web analytics solution that gives…rich insights into [the] website traffic and marketing effectiveness. Powerful, flexible and easy-to-use features [allows to] see and analyze…traffic data in an entirely new way" (Google Analytics, 2009).

Google analytics scripts, once implemented on a website, capture visitors' activities. The metrics that are captured and the different features offered by analytics are mentioned in Appendix A. The metrics that were used in this study were from the three main sections of analytics – *Visitors, Traffic Sources, and Content*. *Visitors* section provides information on visitors to the website. The sub-sections under *visitors* used for this study were the visit patterns, machine settings of the visitors as well as network properties through which the visitors access the website. *Traffic sources* option provides the information paths that the visitor takes to reach the website. The visitors could use direct access, access by clicking through another website, or use a search engine. A sub-section on the keywords that the visitors use are also provided. The third section on *content* provides information on visitors' use of the pages of the website in terms of the most used pages or groups of pages, pages that the visitor entered to and exited from and

the navigation paths that they take within the pages of the website. The specific metrics for each

of these sections that were used were:

1. Visitors – Provides visitor specific information.

    a. Visitor Trending – Provides report about the visitor in terms of the visit

       quality and technology.

        i. Visits - Number of visits on the site made by the visitors.

        ii. Absolute unique visitors – Number of unique visitors on the site.

        iii. Pageviews - Total number of pages viewed by the visitors.

        iv. Average Pageviews - Total number of pages viewed in a given

           day divided by the number of visitors that day. It is considered

           one way of measuring visit quality. A high value suggests that

           visitor interacts extensively.

        v. Time on Site - Time spent on site.

        vi. Bounce Rate - Percentage of single page visits on the website.

    b. Browser Capabilities – Provides information on the visitors machine

       settings such as the browsers, Operating Systems, and combination of the

       two that visitors use, their set screen color preferences, monitor

       resolutions, versions of Flash installed, and support for Java is available

       or not.

    c. Network Properties- Presents information on the network visitors use to

       connect to the website. For this study the Network Location metric was

       used that provides the name of the internet service provider visitors use

       to access the website

2. Traffic Sources - Provides information on the paths of access to the website. The

   metrics used for the study and the definitions are as follows:

a. Direct Traffic – Visitors who clicked a bookmark to come to the site or typed the site URL into their browser.

b. Referring Sites - Visitors who are referred from other sites.

c. Search Engines - The search engine entries to the site.

d. All Traffic Sources - Visitors from search engines, sites, and tagged links.

e. Keywords – Provides information on all metrics for each keyword used to access the site.

3. Content - Presents reports on the website information in terms of the use of the pages. The following metrics under this report were used:

a. Top Content - Most commonly viewed pages.

b. Content by Title - Most commonly viewed groups of pages as denoted by similar titles.

c. Top landing pages - Pages that entice visitors to click further.

d. Top Exit Pages - Pages from where people exit the site.

e. Site Overlay - This feature allows navigation of the site as a visitor would, allowing the analyzer to view metrics for the different links on the site.

f. Navigation Summary – The report shows how often a page in the website was an entrance page, what pages were viewed before this page, how many visitors left the site from this page, and what pages were viewed after this page.

Google analytics view display features that allow users to manipulate the reports were used for comparing and contrasting the figures for the metrics. For example, the reports were mostly viewed using the table and percentage views (Figure F1). Graphs were viewed mostly by day except for one instance when it was viewed by hour (Figure F2). Google analytics also

provides comparison options between two metrics and between the metrics and the site (Figure

F3). The comparison feature between the metrics and the site was used for one report in this study

(Figure F3 -b). Reports were also using the contribution of a metric to the total *visits, pageviews,*

time on site, new visits, and bounce rate in the percentage view (Figure F4). Dimensions of a

report allow viewing segments of the report by different metrics (Figure F5); some of these

metrics that relate to a non-commercial setting are *Source, Keywords, Visitor Type, Landing*

Page, Language, Browser, Operating System, Screen Colors, Screen Resolution, Flash Version,

and *Java Support.* Appendix A provides a summary of definition for each of these metrics. The

metrics *visitor type* and *landing pages*, as required, were used to segment a few reports.

Analysis of the Google analytics data was conducted by investigating the aforementioned

reports. The graphs and reports were physically explored and interpretations of the figures for the

metrics were extrapolated to understand users' information behavior. The researcher specifically

was looking for descriptions of:

- patterns in users' behavior

- what might influence users' behavior

- evidence of users' information resource use

Knowledge gained from the discussions with the MU library web committee, and review of

academic users' information behavior further helped inform the researcher what to look for.

Patterns of users' behavior were investigated for the *visitor trending* sub-section under

visitors and the *traffic sources* section. These metrics were selected to understand the quality of

visits of the users on the Library website. For example, the Library web committee discussed that

they would be interested to know the peak times that users visited the Library website so that they

can decide their online reference staffing. The visits graph under visitor trending was investigated

to understand user's peak time of visits (Figure F2). Another example is the committee's desire to

know if users recall the location of the resources they visit. The traffic sources section provided

information on the visitors' different paths of access that was further segmented by visitor types

41

of new or returning visitors. The traffic source report helped investigate visitors' use of the sources with an indication if these visitors were new or returning to the Library website.

Influences on visitors' behavior were investigated by looking at the technology the visitors use. The *browser capabilities* and *network location* under the *visitors* section were explored. The Library committee stated their interest in the screen resolutions of the users monitor that could enable the Library to provide appropriate images on the website. Visitors' screen resolution will influence whether the visitor could view these images or not. Also, the committee was curious to know if the users are coming from on-campus or off-campus. While analytics provides *geographic* location of users, more specific differentiation between users' access from on-campus was identified by looking at the network service providers of the visitors under *network location*.

Evidence of resource use was investigated from the reports under *content* in analytics. The Library committee wanted to know what resources provided are being used by the Library visitors. The *top content* report was inspected to understand the resource use by the visitors. The Library committee was also interested to know if users are leaving a page or going back from that page. To understand the users' flow of actions through the website, the navigation summary was probed. The sequence of users' actions from one page to another was studied for more interpretation.

The data for this study, therefore, was collected using the interactive group interview of the Library's web committee, and the Google analytics tool. The two types of qualitative analysis using coding techniques for the interview data and investigation of analytics data further informed the research questions. Though a part of the data is quantitative as provided by Google analytics, the interpretation and investigation was qualitative in nature. Table 2 segregates the data analysis by the data collected and the research questions:

Table 2 - *Data Collection Source and Data Analysis Strategy*

Research Questions	Data Collection Source	Data Analysis Strategy
1. How can web analytics data inform library decision-making?	• Interactive group interview	– Coding the informant feedback from the interactive group interview. Interpretation of the metrics by administrators and its application in the Library environment informed how the data obtained using analytics can be interpreted in a library context
2. How can web analytics data be used to interpret the information behavior of the website visitor?	• Interactive group interview	– Coding the informant feedback from the interactive group interview. Understanding the user behavior elements that interest the Library.
	• Google analytics	– Examining the metrics- Visitors, Traffic sources, Content

3.3   Reliability and Validity

Feedback from the interactive group interview participants provided the qualitative basis of the study. The respondents' feedback was used to understand the Library context and the use of analytics. The analytics' reported data, though presented in quantitative format, were analyzed qualitatively.

Some of the issues with validity of the study are related to the Library's analytics implementation. The MU library has computers for access to its resources that are placed within the Library's premises. Tracking profiles of the users who access the Library website from these public computers and other computers in campus labs is difficult. Tracking other users who access computers outside the premises is complicated because of the dynamic IPs assigned to the machines by their Internet service providers. Nicholas and Huntington (2003) recommended micro-mining and segmented log file analysis that can help solve some of the issues with large logs of data by focusing on narrow range of IPs that belong to a particular profile. In the current study, users' affiliation to the academic institution and the knowledge of the profile of users that access the Library resources provide additional information that is used in the interpretation of the data. Owing to the nature of academic libraries it can be safely assumed that MU Library users are either students (undergraduates and graduates), staff, or faculty. Further, the study draws from the literature on academic user information behavior. For example, Head (2007) pointed out that "upper division" academic users who visit the library's electronic resources would be driven by their need for quality information. Users of the web, on the other hand, might visit the websites for all sorts of activities starting from information seeking, for entertainment or for transacting and so on. The directed goal of the users of an academic library's online resource further reduces a layer of uncertainty regarding users' behavioral intention when visiting the library's gateway.

A major part of the interpretation of users' behavior is done using the information presented by Google analytics. The validity of the study is contingent upon the underlying codes

that generate the reports for each of the metrics. Examining the validity of each of the metrics is one way to assure the validity of the study. However, such a validity check would give rise to the need for comparative studies trying to understand the correctness of the values generated by the analytics application (e.g., Jasra, 2006), which is beyond the scope of the current study. The study examines the importance and usefulness of analytics for academic libraries, given the information captured by the analytics tool from the Library website. The primary goal of the study is to understand the users' online behavior on the Library website using web analytics and evaluate the usefulness of this tool for the Library.

Lincoln and Guba (1985, p. 301) establish credibility as a trustworthiness criteria that should be considered in naturalistic inquiry. They provide instances of activities that could increase the probability of credibility – prolonged engagement and/or triangulation. The stage 1 of the study consisted of interactive group session that provided feedback from multiple subjects at the same time, which helped in reaffirming some of the facts about the use of analytics by the Library. Member checking, as another option for establishing credibility (Lincoln & Guba, 1985), involves testing of interpretation and conclusions with the members from whom the data were originally collected. The members of the respondent group were further consulted for checking data interpretations and follow up questions that arose from the interpretations. Lincoln and Guba (1985) discuss the purposes of such a member checking technique that provides an "opportunity to assess intentionality…, [provides] immediate opportunity to correct errors of fact and challenge…wrong interpretations, [provides] opportunity to volunteer additional information, puts respondents on record as having said certain things" (p. 314). Further Lincoln & Guba (1985) distinguish between member checking and triangulation in that the former deals with overall credibility and the latter is directed at accuracy of specific items. 'Transferability', according to Lincoln and Guba (1985), can be ascertained by maintaining a thick description of the study. They state that "it is [the naturalists] responsibility to provide the database that makes transferability judgments possible on the part of potential appliers" (p. 316).

The qualitative element of the study is naturalistic to the extent that the research takes place in a real world setting and there is no attempt by the researcher to manipulate the phenomenon of interest (Patton, 2001). Libraries have been urged to use web analytics as a tool for analyzing the web traffic just as e-commerce websites have in the past (Pagano, 2006; Pace, 2004; Bracke, 2004). Lakos (2007) recently discussed at length the need for libraries to use new analytical tools that would support *Evidence Based Librarianship*. Also, the analytics data that were studied were already set up by the Library before the study was conducted, and thus recognized as important to the Library management.

Furthermore, Nicholson (2006b) suggest the differences between the researcher (concern about external validity), and the practitioner (concern about the internal validity) can be bridged by proposing a data-based collaboration. One of the purposes of this study was to understand how the information obtained using analytics can be used as a library management tool. The internal validity of the study can be ascertained by assessing the usefulness of the analytics data to the management.

The notion of 'transferability' in naturalistic studies has been emphasized through a description of the time and context in which the study was held. According to Lincoln and Guba (1985) the "naturalist cannot specify the external validity of an inquiry… [rather] can provide only the thick description necessary to enable someone interested in making a transfer to reach a conclusion about whether transfer can be contemplated as a possibility(p.316)." The qualitative analysis of the data obtained through analytics will be used to interpret within a particular domain of academic users' behavior on the MU library website.

A possible way of addressing the construct validity of the study can be achieved by examining multiple sources of evidence. In this study the implementation of web analytics as a library management tool was determined by the Library's web usability committee's feedback. The specific feedback from the interactive group interview was carefully coded for issues that libraries face with a web analytics implementation and its use. Other factors that were looked for

46

in the interactive group interview were indications of internal politics that hampered the implementation and strategic issues that required halting or affecting the implementation in any way.

Reliability can be achieved by operationalizing as many steps as possible that can be audited later (Yin, 2003). This study has been operationalized to involve working with the metrics, drawing out any explanation from the literature on user studies, coding the feedback obtained from the librarians, and creating a list of information that is of importance to the Library management that they expect analytics to provide. Another way to run a reliability check is by replicating past studies; however, the lack of any full-blown study on analytics implementation in any library restricted any likelihood of replication. Proper documentation and detailed stepwise account of this study has been done for any future replication, though circumstantial factors could still be a possible deterrent.

In summary, the study ensured research quality through the following ways:

1. Use of studies in user behavior –*Human Information Behavior* literature to explain the interpretations of the analytics data.

2. Research plan iteration –The research design involved investigating the analytics data before and after inputs from the Library web committee. Such a measure was taken to ensure that the Library context is maintained while implementing analytics.

3. Interactive group interview –A higher engagement of the Library management was achieved through an *interactive group interview* technique. The respondents in the interview were kept motivated to learn about the value of analytics as they have their vested interest.

4. Informant feedback –The researcher checked with the informants about questions, clarifications of the feedback obtained.

The interview analysis was conducted with qualitative coding. The researcher provided a detailed account of the coding steps and illustrated them with examples. In addition, the stepwise process that was undertaken to form the various themes and subthemes has been described and illustrated in detail. The next chapter provides the findings from the analysis segregated by each of the research questions. The themes found through the analysis of the interview are provided for the research question 1, and the findings from the analysis of analytics metrics are discussed under the research question 2.

Chapter 4

Findings

This chapter provides the findings obtained from the study of Google analytics implementation for the MU library website. The data were collected from the interview with the Library web committee members and from the Google analytics implementation of the Library website. This study used open-ended and axial coding techniques to analyze the interview transcripts. Investigation of the analytics metrics and reports were done by careful observation and interpretation of the metrics. The study employed findings from the analysis of the interview transcript to respond to research question one. The inquiry focused on findings from the analytics metrics to address research question two. The chapter will name MU library as the "Library," and the MU library web advisory committee as the "committee" or the "web committee."

RQ1- How can web analytics data inform library decision-making?

Findings from the analysis of the interactive group interview transcripts that relate to Library decision-making are reported here. Excerpts of quotes by the respondents, where necessary, are provided and in some cases paraphrased for the purposes of clarity. For ensuring privacy of respondents, this document does not include their names. Instead they are represented here as Respondent one, two, three, four, five, and six. Analysis of the interview transcripts revealed four major themes that are discussed here. The primary themes were –

1. Decisions about services
2. Decisions about web design
3. The Library's interest in user behavior
4. The Library's management and function

These themes were further divided into subthemes. There is a discussion of the findings from the interviews for each theme along with the subthemes for each.

4.1   Decisions about Services

The theme *decisions about services* emerged from aspects of decision-making by the MU library to improve services that are provided to its patrons. MU library needs to take decisions to offer better services to its patrons. An important part of the services involves that the patrons get the required information from the resources provided to them by the MU library. The MU library website provides a major portion of the information through online databases, and eJournals. Other Library services can be accessed through the website include the catalog, ILL, user account services that provide check out information online, online renewal of books, and access to course electronic reserves. The MU library would like to provide information that is convenient and available online, easy and quick to find. The team formed by the MU library web committee ensures the smooth operation of the MU library website and therefore makes decisions to that effect. Themes and subthemes associated with *decisions about services* are as follows:

- Supporting staffing decisions

- Providing machine compatible resources

- Offering useful subscriptions to electronic journals

- Providing a satisfying search experience

- Providing for specific user needs

4.1.1   Supporting staffing decisions

The members of the MU library web committee mentioned the need to ensure proper staffing to assist patrons. With the online interface, the patrons have an option to access the Library's resources through the Library website. Knowing the time of day the website was used more frequently would help the Library support the users' activity on the website. As a part of

ensuring proper services, the Library would like to know the time of day when they experience high traffic. The Library could increase their preparedness to provide efficient services to their online visitors by looking at the variation in traffic flow and create a back-up system resources and staff to handle high volume of visitors. The members of the MU library web committee mentioned that analytics data can be used for understanding the access times of the traffic to the website, and hence support their staffing related decisions. Respondent six stressed the use of analytics to support staffing needs,

> …If we were seeing big uptakes and if it, say, dropped off at certain time of the evening that will help us make a case either for or against expanding reference desk staffing hours. Certain times…maybe we need to get some chat 24/7 [as] these people are using our webpage around the clock and somebody can contact the reference.

4.1.2 Providing machine compatible resources

The respondents mentioned the importance of knowing users' machine preferences. Since the online visitors use different technology to access the Library resources, knowledge of the technology can help to ensure that the Library services conform to these specifications. Other than that, the constant upgrades in technology further enhance the need for keeping track of whether the Library services can be accessed by online visitor machines. Respondent one stated that compatibility with the users' browser was a major concern because one of the branch libraries has resources that are mostly audio clips, and Firefox users might have issues.

It is also important that the Library is aware of the resolution of users' machines. Respondent three mentioned an instance when the Library had concerns about an image that was 600 X 800 that was not compatible with bigger monitors which users accessed. Respondents realized that the services provided by the Library need to be compatible with the users' machines so that the users can avail themselves of the Library's services smoothly and conveniently. It will help the Library implement features and services that are compatible with user machine configurations.

4.1.3   Offering useful subscriptions to electronic journals

The Library provides information services by subscribing to eJournals. The members of the web committee mentioned that due to the Library operations in the past, most of the subscription decisions have been made without evidence of its usage. Some of these decisions about eJournal use have been made a long time ago, and the Library management usually decides to adhere to it.

The application of analytics on the Library website provided them a way to understand how these different eJournals are being used. According to Respondent six, analytics implementation could help in decision-making regarding to which journals to subscribe. The members of the web committee agreed that analytics can provide detailed information on the visitors' activities on the eJournal usage. However, the respondents expressed their concerns about the noise in analytics reports. They were wary of the fact that there could be some nonhuman responses in analytics reports and that might affect the correctness of the reported metrics.

4.1.4   Providing a satisfying search experience

Most of the online visitors that access the Library website are accustomed to getting information quickly using search engines like Google. The MU library web committee was aware that users prefer to search rather than browse through the different Library website links. This was further stressed by the presence of different search preferences on the MU library website. The Library website has multiple search functions that allow users to seek information on the Library website. Besides the website search box that searches across the Library gateway, the Library also provides a menu-based search option that allows visitors to find information across the Library's collection. For users who want to search for a specific article, database or a journal there is a search box under the respective choices on the home page: "Find database," "Search for

articles on a topic," and "Find specific article or journal." These allow users to search for a database or journal.

The menu-based search options in the Library website further lead the librarians to wonder how users are accessing the different menu options when conducting a search. Respondent six stated: "…What kinds of searches people were putting in and whether they were correctly changing the drop down menu. If they were searching for what looked like an article title or book title, then were they doing their appropriate kind of search for the kind of search box they were using?"

Committee members found that understanding the different access terms that visitors use to find the information was considered useful. The databases that the Library subscribes to, allow the Library to create *aliases*. The aliases are alternative names that can be used to access the databases. Respondent five mentioned that the access terms that visitor of the website may use within the Library website can be put as database aliases. Looking at the different access terms that the visitors use provided a check for matching them with the aliases for a database. The members of the committee saw matching the access terms with the aliases; adding key terms to the aliases in case of a mismatch was a means to enhance visitors' online search experiences.

Information about the users' search behavior helped the Library to make the content more searchable and easily accessible. The Library would have liked to know if the search terms entered by people were found or not. Respondent three asked: "What sorts of things were people inputting to search?" If they did not find those elements they tried to accommodate the situation. They also used the search terms entered to make changes in the keyword search. Respondent six stated,

> It [has] been very very useful for us to see what terms people are putting in to the
> search engine on our web page for a couple of reasons: one, to see the spellings
> of various databases so we could incorporate the spellings so that it would bring
> the search directly to what they are looking for. So we were able to do an alias
> and were redirected to get people seamlessly into that page. Another [reason was]
> to see what kinds of searches people were putting in, and whether they were
> correctly changing the drop down menu to pull up [the searches]. If they were

searching for what looked like an article title or book title then were they doing their appropriate kind of search for the kind of search box they were using…[Also] which kind of searches were done most often and what kind of choices were…pulled up most.

Respondents wanted to know, if users are accessing a search engine like Google, how many of them are doing so and what keywords they used most to access. Visitors' search-related behavior would help in search engine optimization of the Library's website.

4.1.5   Providing for specific user needs

Visitors to the academic library are usually undergraduate students, graduate students, faculty, or staff. The Library's user base consists of different demographics that include users who were from different backgrounds, age groups, experiences, and needs. The access habits of these different user groups are useful in providing targeted services to these users. Sometimes knowing that some visits are from Library staff would have been useful in that the Library can eliminate those visits to get a better understanding of the real users who come to the Library website for accessing Library's services.

Respondent six stated that it would help to know where the online users are coming from, either Library staff machines or from outside the Library. The respondents were interested in knowing which of the Library visitors were students, and if they were behaving differently than the faculty. The current visitors of the Library website could be identified by if they are using "Mizzou" in their searches when they enter the Library URL to get to the Library website. The respondents thought that would be a good way to know how many of these users are students, and they can further look into the differences between other users who utilize "Missouri" in the URL.

Knowing the profile of the specific users and if they were students or faculty would have helped support the user-specific decisions of the Library such as if undergraduate students need resources more than graduate students. Identifying user visits that were from staff would provide a clear picture of the usage of the Library resources.

4.2 Decisions about Web Design

The theme of "decisions for web design" emerged from the Library's need to constantly improve its website's usability to suit its visitors. The constantly changing information environment has affected user's information-seeking behavior. Since the Library has tried to reach its patrons through the online system, providing them access to the Library through the conveniences of their home would help to keep improving the users' experience on the web. Therefore visitors would be able to get the most out of the services provided to them.

The Library website had different ways to get to information provided, such as catalogs, different search tools, closed access to their Library accounts, and so on. Apart from that, there were different types of information provided by the Library. Library services included hours of access, different workshops held by the Library, copy services, online chat reference, and so on. Other services included providing different information sources that the Library subscribes to through databases, and eJournals. There were still other kinds of services where the online visitor can access their accounts, check or print out electronic reserves of curriculum items, look at items the visitor has checked out, renew materials online, place holds on required items, and at other times request an Inter Library Loan. Providing such varied ranges of services made the respondents curious about how visitors were able to successfully accomplish the functions for which they visit the website. The subthemes that emerged from were:

- Changing website design based on analytics reports
- Restructuring website content to improve navigation

4.2.1 Changing website design based on analytics reports

The MU library has had Google analytics implemented on the Library website since March of 2007. A purpose of this implementation was that they should be able to capture users' activities on the Library website. A use of the information captured by analytics was to understand what changes in the website could help provide users a satisfactory experience.

According to the respondents, most of the interpretation on the MU library users' behavior that they have gotten has been through general observations such as visitors who go to 'Search Databases' were looking for articles.

Information on whether the user moved forward or closed out of the website would have helped in making changes to the website structure. Librarians would have liked to have seen in analytics the continuity of users' actions between pages, as shown by Respondent three: "Seeing one page at a time is useful, but seeing one thing after the other is useful some times." In other words, continuity of users' actions when visiting the Library website was considered useful for the Library's purposes. In case a user was at a page for hours, then it would have been informative if the data suggested that the user was lost and then left the page. A relationship between the times the user spent on a page with exits from those pages would have been an instance that they would expect to see in analytics. Respondents thought that the more the time the user spends on the page, the more chances of the user being lost and leaving.

Users' paths of access are important for the Library to control their movement through the pages. Respondent three stated that the Library is interested in looking at whether people went from a page to where they would like them to go, where they thought they should go, or they left the website without clicking on any link. The respondents indicated that it would have been useful for them if the analytics presented the full path to the Library resources.

Members of the MU library web committee expressed interest in knowing the use of links on the Library website as shown by analytics. In some cases, when the Library wanted to know about user activity on a specific link, it had to search for that link from the content report in analytics to get specific metrics for that page. Analytics provided an option to compare a page metrics with the website metrics. However, the respondents mentioned that a comparison of the usage of the links that led to different resources would have been useful for the Library, as the comparison would indicate preferences of one resource over the other.

The Library has taken steps to improve the visibility of the less used sections of the website. Respondent five mentioned that the Library has used analytics in the past mostly for moving links based on the clicks they received on each. They saw using analytics features to make improvements to sections that get less clicks. Respondent six mentioned by looking at the *site overlay* feature of analytics they saw that the 'Announcement' section on the home page was not getting many hits, and so they might have decided to make the section scroll. Though the Library would like to use analytics more often, they have not been able to use it as extensively.

4.2.2   Restructuring website content to improve navigation

The various resources on the Library website are made accessible through the hyperlinks on the different webpages. Links play a crucial role in allowing users' entry to the Library resources as well as to connect from one page to another. Over time the Library has accumulated lots of links across the different web pages. A major concern for navigation through the Library's pages lay in the presence of multiple links to the same resource across the website, and users who may have been lost or confused when returning to the website. On the other hand, multiple links provided multiple options for users to go to a resource.

Respondent three mentioned that most of the users got lost because of the presence of multiple links and eventually found themselves on the same page where they started. Whether the users recalled the paths they visited when they came back to the Library website was important for the respondents to know. Also, Respondent six recognized the ineffectiveness of a link if it was not being used. The Library could have decluttered a page based on any information on redundancy of links on that page. The Library wanted to provide an optimal number of links that does not clutter the page along with providing easy access for its users. Respondent five raised her concerns on the complications that might be involved when interpreting usage of the Library's resources such as the database through multiple links.

The Library website had various hyperlinks that lead to different or same pages. The links that lead to same pages were often labeled differently. For example the link http://mulibraries.missouri.edu/outgoing/merlinmu_gateway and the link http://laurel.lso.missouri.edu/search~S1/ were both directed to the Merlin catalog but the former one linked from the gateway. Respondent six also stated at one point that – "…We may have kinds of redundant links on our website. One [says] search for articles on a topic and another one [says] databases." However, the destination pages for both these links were the same. In order to keep an account of which of the links the visitor actually used to go to the next page, analytics provided an option to tag these links. Tagging links allowed analytics to track the different links that were used to access the same destination page.

As mentioned before, the Library staff was curious to know how users adapted the bookmark feature to access the Library pages. However, besides a general curiosity they also thought that information on bookmarking a link could sometimes provide input to the existing structure of the website. If a visitor directly entered a page through a bookmarked link, it would mean users visited that webpage frequently or they found it difficult to access the page by browsing. Respondent three stated, "You know that people are going from here to there; whether they find it is whole another story." Information on direct entries to particular pages would have helped the Library to make decisions about structuring the navigation pathways to the resources it provides.

Information on the full path traversed by the user when on the website certainly could have enhanced the Library's decision-making. Acquiring detailed information on users' activities when on the website would have helped the Library. They would have liked to be aware of how the users are maneuvering through the different links on the website to reach a particular destination page, and to compare it with their expectations of how they would like to see the users traverse.

4.3   The Library's Interest in User Behavior

The theme of *the Library's interest in user behavior* emerged from the Library's interest to know about their users' activities when on the Library website because of sheer curiosity. With the advent of Internet, online users have adapted to the developing technology. Most of the users' behavior on the web was determined by the changing information environment. This was sometimes shaped by developing technology such as availability of multiple creative ways to get information through wikis, online blogs, and digital libraries. The MU library website provided a way for the online visitors to access the Library services through the convenience of their homes. Members of the MU library web committee saw the Library's analytics implementation as a means to get the bigger picture of how their users interact with the Library website. Such an intention to know users' behavior might not have translated to any decision due to constraints or other functional characteristics that apply to the Library. The subthemes that emerged were:

- Branch usage of Library resources

- Using analytics to understand visitor activities

- Shift in user behaviors

- Entry paths to the Library website

- User activity on the Library website

4.3.1   Branch usage of Library resources

The MU library provides various resources for its users within the main Library, as well as across different Library branches that were established under specific academic departments. The distribution of resources across the main and branch library made librarians eager to know the usage of the Library's resources and its relation to the different branches, for example the number of users that come from the branch library websites to use the resources that are under the main Library website.

The MU library has different branches that are domain-specific such as the Engineering library, Law library, Health Science library, Journalism library and so on. The Library's resources could be accessed from the different branch library websites. The Library members wanted to know whether their resources were being used by the branches. Respondent five mentioned that it was of interest to see which branches sent the most traffic to the online resources. Information on usage for particular branch libraries would have also benefitted the branch library gateways, and would have helped the respondents to know how users access the resources available through the branch library gateway.

4.3.2   Using analytics to understand visitor activities

The respondents know that many people rely on its website to visit the Library's resources. The Library website acts as a gateway to the vast resources and services that it provides. Library users who visited the Library gateway either accessed it to obtain information, availed the many Library services provided, or used the information resources to which the Library subscribes. Analytics reports could help the Library understand user behavior elements through its various metrics.

Respondents found reports that enable comparison of information on users' behavior on the website to be interesting. For example the comparison reports between the use of the website and the databases , as shown in a cropped view in Figure 1, interested the members of the web committee. The respondents noticed how the database-use pattern was similar to the website use pattern, indicating that the database usage contributes considerably to the total website usage.
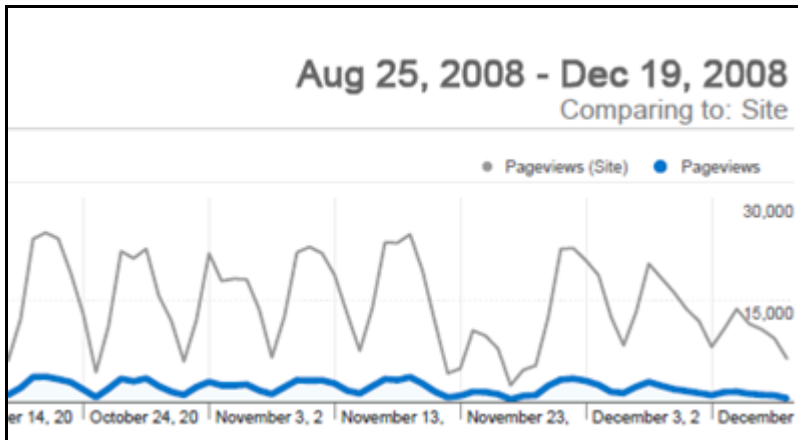
Figure 1. Comparison Graph Between Pageviews of the Databases Page at /search/databases/default.aspx with the Library Website

4.3.3 Shift in user behaviors

Technology improved ways to access information readily and conveniently through different tool options such as improved search engines or navigation bars. Users' expectations from websites changed based on their wider experiences over the internet and how they accessed information in other websites or search engines.

The Library was concerned about how users browse for information on the Library website. It would also have liked to be aware of any shift in the users' access behavior such as change in catalog use. It would have liked to know if the catalog use has decreased from previous years and how was that related to usage of other resources. Respondent three was curious, "[whether] there is a shift of usage from catalog to some other medium within the Library website, if at all."

4.3.4 Entry paths to the Library website

Any online application that is on the web was accessible. The very nature of being online was to provide convenient access. The percentage of visitors accessing through the different sites provided a picture of the users' mental models and where the referring sites fell in these models. Respondent five mentioned that it would be interesting for the Library to know the various paths,

61

such as the specific branch library gateway or other course management websites that was used to access the resources.

Since there was no one specific way that the online visitor could come to the website, it made librarians curious about what paths users take to come to the website. Respondent one stated her curiosity: "Where people are coming from, whether they are coming from the gateway or whether doing an organic search". Respondent six also stated that the Library would like to know whether its users are coming from off-campus or on-campus. Further, information on whether users were coming to the branch library websites through the gateway or are being distracted from other websites was of interest to the members of the MU web committee. How many users were actually accessing the Library website directly was of interest. Respondent six questioned if "direct access....would just be a bookmark [where] the [users] open the browser and type in our URL."

4.3.5   User activity on the Library website

Use of links as discussed before was a more discrete way of looking at users' activities. However, the members of the MU library web committee wanted to understand users' overall intentions and actions on the Library website.

The Library would have liked to know specific user behavior elements on the website, for example, if users were spending a long time on a page searching for resources, coming back to the same page or losing their way to a resource on the website. This information would have helped in regular maintenance of the website. It would have helped to know if people did more things with the databases and not much with Library services. A continuity of actions between the layers of the page would have enabled libraries to know their users thoroughly. Respondent five mentioned that it would be interesting to see how changes made to the Library website could influence users' behavior.

The Library wanted to have a detailed account of its users' actions on the website including where the user was coming from and where are they going. The accounts included the time the user spends on a page, what she was doing when on the page, if she chose to click on a link and go to the next page, to external or internal links in the system or if she left the website. A thorough description of the users' actions would help the Library to get a clear picture of the users' activities that would make them more comfortable.

4.4   *The Library's Management and Function*

The Library's management and function theme emerged from those elements that were exclusive to the Library, its nature and purpose. This study scrutinized the MU library web committee feedback from the interview for such elements that related to the Library policies, typical operational requirements, strategy, upper level decisions, and activities that affected or informed Library management at some level. The management was responsible for most of the decision-making in the Library.

The MU library's mission statement included, "[to] *acquire, preserve and make accessible scholarly resources"* (Mission, Vision and Strategies for the MU Libraries, 2006). The purposes of its online presence were not to achieve a profit motive as with most of the commercial entities that have an online presence. The purpose of making its services available online was for convenience of access. The online presence of the Library remained under the jurisdiction of the Library management that decides on various aspects of the Library's website. The subthemes that were a part of this theme were:

- Privacy of users' data

- Reservations about the appropriateness of analytics

- Issues with measuring usage

- Analytics of branch libraries

- Analytics implementation and subscription to proprietary systems

- Evidence-based decision-making in libraries

- Library disregard of analytics use

- Marketing Library services

4.4.1 Privacy of users' data

Analytics, once set on the Library website, captured the visitors' data. Though the data captured was IP-based and were collected through the scripts entered in the Library website's source code, the reports provided are aggregated without any reference to the IP-address of the visitors' machines.

The Library would have liked to get a detailed account of the visitor activity on the website; however they feared the privacy of the visitors was going to be violated. Privacy of its patrons was a priority for the Library. The respondents were conscious of the different non-user responses in the form of spammers and crawlers, which they feared were going to get across to the visitors' machines through the information captured by analytics. Respondent six observed that users would not like it if spammers and crawlers came to the Library website and acquired their emails. Besides, the respondents questioned regarding how many of the reported figures were due to unaccounted visitors. They were concerned that misrepresentation of usage due to these unaccounted visitors would affect Library decision-making.

4.4.2 Reservations about the appropriateness of analytics

The Library had too much information to manage. According to the committee, there could have been over 1000 URLs in the Library website and many of these were redundant. The Library website was built over time and involved many developers who came and left. There was no standard way to address the different pages. Respondent five mentioned that she "has her little thesaurus" that helps her to know the different links for a page.

Apart from that, the Library website was a gateway to the many resources the Library provides. Respondent three stated that there was more to the Library website than just the

gateway. Respondent six stated, "[The Library has] outside stuff that we consider inside for our purposes." Even the Library catalog was created outside the Library. The Library would not have minded if users went to these outside resources. As a result, the Library thinks analytics reports might not provide the right picture. Respondent five stated that,

> …Most [Google Analytics] was created for companies that want to keep people on their websites as long as they can show them ads. The whole purpose of the gateway [Library] is to get them off the site as fast as you can because then where they are going because it's a gateway. It's not trying to be sticky and trying to put all this information, I mean like the bounce rates on the other pages...if a company looked at the Library's bounce rates they would think it was a terrible failure, but for us high bounce rates is good, if they are getting to what they need. They don't have to come back and try.

4.4.3 Issues with measuring usage

The MU library provided facilities that allowed visitors to access the Library website from within the Library premises. There were other options that allow visitors to access the website from computer labs on campus. Users could also access the Library's website through their own laptops or through their machines at their offices and homes. The Library's internal staff visits the Library website more often to perform regular tasks related to their job roles. The Library had profiles of users visiting the Library website that had varied information access habits.

Sometimes the same users accessed from different machines; at other times different users accessed the website from the same machines. The Library would have liked to know how much of the data reported was by staff, from on-campus or off-campus users. To get an account of the unique users to the Library website was difficult. Also, as mentioned before, the interpretation of metrics for usage of the Library website was different than the actual interpretation of the metrics in analytics such as the interpretation of bounce rate and exits. A user visiting a single page and leaving from that page would not have necessarily be a negative trait of the website.

65

4.4.4 Analytics of branch libraries

The MU library had discipline-specific branches that are spread across the campus, and some of these branches had their own online access interfaces. The analytics implementation for the Fall 2008 semester was for the whole MU library system except for the Law library, Health Sciences library, and the Veterinary Sciences library. Looking at specific branch analytics provided useful information related to each of the branches of the MU library.

Respondent five cited the usefulness of analytics if it could have provided information on the branch the resources are being accessed. The Library is not a homogeneous entity. What would be important to them was how the different branch libraries fare. In order to look up the branch information the Library had to search for the branch URL from the content report to get more information about the metrics for a particular branch. However, doing this every time did not provide a detailed report of the branch gateway use.

4.4.5 Analytics implementation and subscription to proprietary systems

The MU library provided services from proprietary systems that are external to the Library. The Library visitors were directed from the Library website to these external systems. Because these proprietary systems were external to the Library, the Library could not implement analytics to measure the usage of these systems. Respondent one mentioned that the databases they subscribed to provided the Library statistics. These external proprietary systems provided usage reports to the Library on a regular basis. The Library had to rely on these reports for any decision-making, and could not measure the performance of such external systems. Not being able to set up analytics for the proprietary systems that were part of the Library services keep the Library from setting an accurate picture of the services they provide.

4.4.6 Evidence-based decision-making in libraries

EBL has been a part of traditional library decision-making. The Library recognized some use for analytics in providing evidence to support important decisions. For instance, Respondent

six stated that in case a decision was made on dropping the link to the Library website from a

referral website that is internal to the university, the Library could have been accountable by

showing the analytics figures for the number of visitors who came through that referral link.

Moreover, using analytics changed decision-making that was based on mere intuition and

general knowledge about users' behavior. Use of journals and databases could be supported by an

estimate of users' visiting these resources. A further understanding of what the user was able to

access and use could also be made by looking at user machine compatibility with the services and

features provided by the Library.

The Library made decisions without information on actual usage when they subscribed to

eJournals in the past. However, the respondents thought information on actual usage of the

Library resources could help to inform Library decisions. For example, Respondent one stated

that people primarily do things with databases and not so much for Library services; this was

certainly going to inform the design of the Library website.

4.4.7   Library disregard of analytics use

On one hand, analytics as a tool was useful for supporting Library decisions by providing

evidence. On the other hand, sometimes this evidence was overlooked because of preexisting

policies. Respondent six stated that,

> …We may have kind of redundant links on our website one goes to search for
> articles on a topic and another one goes to databases. Now we know that that [is]
> intuitive of people who go to databases, go to search for articles on a topic and …
> probably with some comparative numbers of which one of those links attracts the
> most clicks. But basically…for political reasons we are going to keep both of
> them on there.

Another instance of a Library decision that disregards analytics figures was the 'Announcement'

section that had to be present even if it does not capture many clicks.

4.4.8 Marketing Library services

The Library services' need to be marketed did not seem to be as important as for commercial websites. It is preferable usually for commercial websites to be accessed through many referral sites. However, when it came to the Library setting, the staff would have preferred to refrain from promoting the Library gateway to the referral websites, especially the external ones that sent traffic to the Library website. The Library's concern was that any promotion to the referral sites would have given these sites a reason to capitalize on the Library's expense. Respondent three emphasized that the Library would be giving marketing ideas to places that they were not directly linked to if they try to promote their website to these outside parties.

Concerning internal referral websites, the respondents agreed that the referral information would have helped if some referral site tried to remove the link to the Library website such as Journalism, for which staff could always have made their case with analytics numbers. Also, the librarians thought it would be interesting to know how Factiva would react if they knew that they were sending traffic to the Library website.

The findings from the interview discussions provided insights about how analytics and Library decision-making could have been connected. Though the discussions were aimed towards understanding how analytics could be used in Library decision-making, there were other aspects that came forth. Popular analytics use from the commercial sector could be idealized in an academic library setting as is. Implementing analytics in the Library for Library decision-making needed an inside-out approach where basic Library ideology needed to be considered first. Some of the aspects of analytics interpretation in the Library could be connected to that of the commercial sector when it came to providing service, improving web design and understanding user behavior. However, the Library setting was far more focused on the non-commercial facet as compared to the commercial enterprises where analytics was popularly used.

RQ2- How can web analytics data be used to interpret the information behavior of the website visitor?

As mentioned before in chapter three, the analytics tool used for this study was provided by *Google analytics*. Data on the users' online activities were captured using *Google analytics* by inserting the profile-specific tracking code. In this study the *Google analytics* implementation for the MU library was used, and the data for the Fall 2008 semester was investigated to understand how analytics could be used for interpreting users' online behavior.

Google analytics has a variety of metrics that capture the different elements of users' activities on the website such as *visits, pageviews, visitors' network properties* and so on, as detailed in chapter three. In order to provide a more significant meaningful interpretation, inputs from the MU library web committee, collected during the first stage of data collection were utilized to select the metrics needed to understand users' behavior. Metrics that applied to the Library's decision-making needs were used to interpret the users' behavior on the MU library website. For example, the respondents wanted to know whether users' navigation path matched with the Library's expectations of it. The *Navigation Analysis* feature of analytics that is present under the *Content* section provides a page-by-page navigation behavior of the user.

The findings discussed in this chapter are explained under the three different sections under which Google analytics segregates data on the visits to the website: *Visitors, Traffic Source,* and *Content*. Metrics under each of these sections were selected for investigation. The selection of metrics was done based on its significance to the MU library web committee discussions. Explanation of how these metrics are used to interpret users' behavior is provided using screenshots of the analytics reports. Some figures are cropped from their original views for clarity purposes. Table 3 provides an account of the Figures that are used to present the analytics reports. Also, for clarity purposes the screenshots are marked using the *Jing technology* provided by Techsmith that allows screen capturing and annotating with arrows, texts and highlights to

show the figures and features being referred to when interpreting user activity. *Google analytics* has been referred to occasionally as analytics or web analytics in this section.

Table 3 -*Description of Figures Presented Using Analytics Reports*

Figure #	Figure Name
Figure 1	Comparison Graph Between "Pageviews" of the Databases Page at /search/databases/default.aspx with the library website
Figure 2	User Visits on the Library Website for Fall 2008 Semester
Figure 3	Average Pageviews for Visitors on the Website during Fall 2008 Semester
Figure 4	Time on Site for All Visitors of the Website during Fall 2008 Semester
Figure 5	Bounce Rate Variation across Fall 2008 Semester
Figure 6	Graph by Hour over a Day for Fall 2008 Semester
Figure 7	Visitors Trend Variation Compared with Visits for the Fall 2008 Semester
Figure 8	Visitors Trend Variation with Pageviews for Fall 2008 Semester
Figure 9	Browsers that Visitors Used for the Fall 2008 Semester
Figure 10	Different Internet Explorer Browser Versions used by the Library Visitors for Fall 2008 Semester
Figure 11	Operating Systems of Library Users during Fall 2008 Semester
Figure 12	Versions of Windows Operating System Used by Visitors for Fall 2008 Semester
Figure 13	Browsers and OS Combinations Used by Visitors
Figure 14	Screen resolutions of the Visitors Machines Accessing the Library Website
Figure 15	Network Location of Visitors of the MU Library Website during Fall 2008
Figure 16	All Traffic Sources to the Library Website for Fall 2008 Semester
Figure 17	Direct Traffic Details by Visitor Type for Fall 2008 Semester
Figure 18	Referring Site Traffic Shown With the Visitor Type Dimension for Fall 2008

4.5   Visitors

The *visitors* section in analytics provides information on user visits to the website in terms of *visitor trending*, *visitor loyalty*, *browser capabilities*, *network properties*, and *user defined*. The metrics investigated in this section were chosen based on the discussion of the MU library web committee. The MU library web committee wanted to know information about the visitors of the MU Library website. Specifically, this information related to whether the visitors were present round the clock, whether the peak times that they visited fell off at a particular time, if there were dips in usage at certain times, and what time of the day or year they usually visited the most. More information on how long the users are spending on the Library website was also of interest to them. They also discussed the need to separate the visitor profile as to whether they were students, faculty, or staff. The web committee would have also been interested in knowing about the users' machine configurations that involved the browsers and connection speed of the users' machine. A further interest was knowing from where these users are coming, off-campus or on-campus. The discussion was the basis to investigate the three specific categories of the *visitors* section of analytics mentioned as follows:

- *Visitor Trending* provided information on how effective the website is in terms of keeping website visitors engaged.

- *Browser Capabilities* showed how different aspects of the site design and structure could affect how the visitors see and use the website.

- *Network Properties* provided important measurements about the network the visitor was using such as network location, host names and connection speed.

4.5.1   Visitor trending

The *visitor trending* report in analytics provided information on "how effective the website is in keeping people engaged" (Ledford and Tyler, 2007, p.183). It grouped all trend

reports under the headings of *visits, absolute unique visitors, pageviews, average pageviews, time on site, bounce rate.* The metrics discussed:

- Visits

- Pageviews

- Average Pageviews

- Time On Site

- Bounce Rate

Appendix A provides a definition of these metrics. Findings from these metrics are discussed here.

The visitor overview report in Figure 2 provides data concerning the Library *visits* for the Fall 2008 semester. As marked, Figure2 -a shows the visits reached to a high of almost 7000 visits/day during the semester. Figure 2 -b,c, and d of the *visitor overview* report show that the visits fell down during the third, seventh, and fourteenth week of the fall semester with a highest fall during the fourteenth week. The dip in the 14th week could be attributed to the weeklong Thanksgiving recess. However, the drops of visits during the third and seventh week need to be investigated. The Labor Day weekend could be a possible reason for the fall in traffic in the third week. However, the visits dropped for the weekend only and caught up the following week when the visits went to a high of approximately 6000 visits/day.
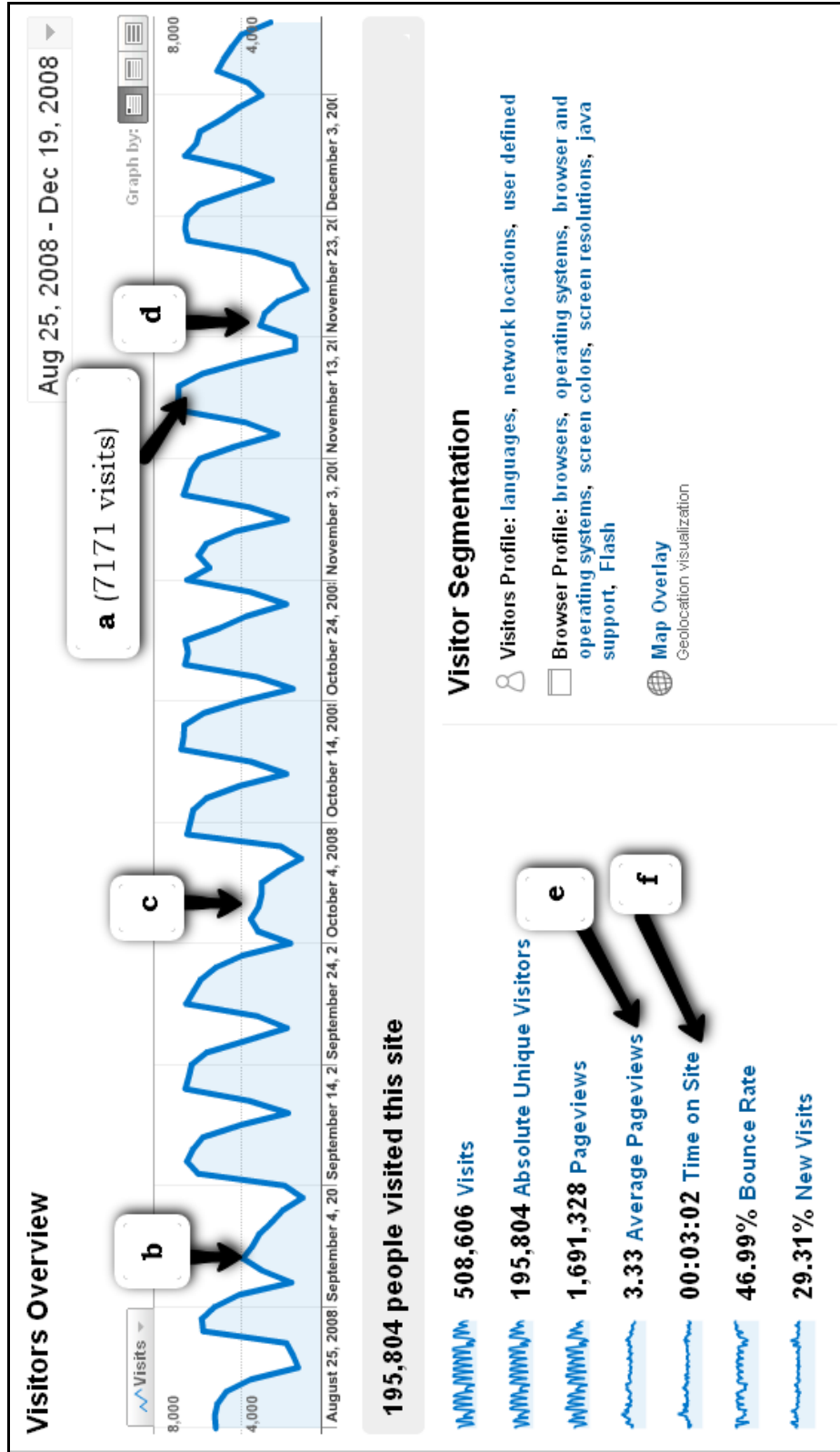
Figure 2. User Visits on the Library Website for Fall 2008 Semester

Figure 3 and Figure4 depicts that the *average pageviews* and *time on site* remained constant of about 3.3 pages per visit and three minutes respectively. The *time on site*, however went up to over four minutes from September 9 to September 11, 2009 (Figure 4 -a). The *bounce rate* on the Library website, however, remained at an average high (Figure 5). Such an indication of high *bounce rate* might not be as negative a sign, as most of the links on the Library website direct the users out to other websites and resources.

Figure 3. Average Pageviews for Visitors on the Website during Fall 2008 Semester

Figure 4. Time on Site for All Visitors of the Website during Fall 2008 Semester

Bounce Rate for all visitors

Figure 5. Bounce Rate Variation across Fall 2008 Semester

The *graph by hour* utility as explained in chapter three provides a way to see the hourly visits for the Fall semester. The visits were at a peak of approximately 42,000 (Figure 6 -a) visits during the day from 9 am to 3 pm. The visits decreased after 3 pm until about a 1000. Most of the user visits were during the day time between 9 AM to 6 PM with an increase of usage during early mornings and a decreasing trend during the evenings.
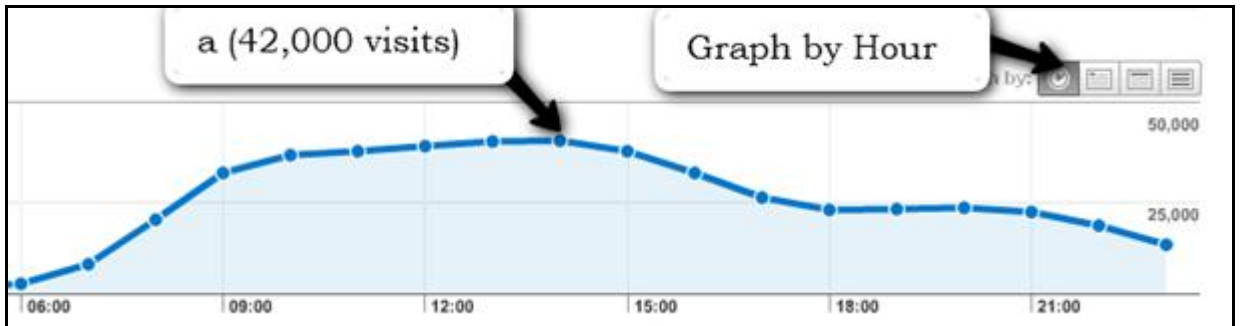
Figure 6. Graph by Hour over a Day for Fall 2008 Semester

The chart in Figure 7 shows a cropped view of the report that compares the two metrics, visitors with visits (Figure 7 -a,b). According to the report the daily visit trends varied with the daily visitor trends for the Fall 2008 semester. A similar variation appeared, comparing the

metrics pageviews with visitors (as seen in the cropped view in Figure 8 -a, b). The daily

pageviews varied accordingly with the daily visitors' trend.
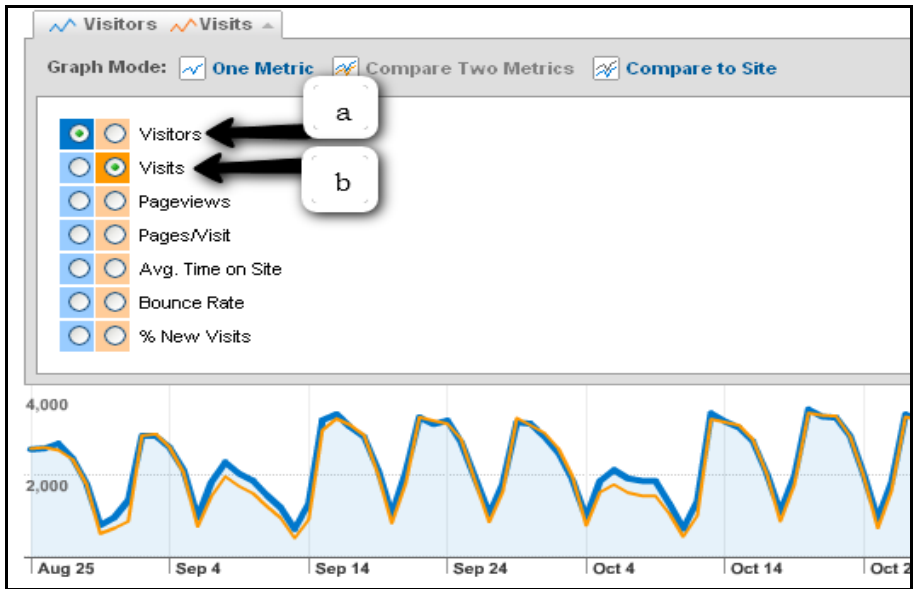
Figure 7 . Visitors Trend Variation Compared with Visits for the Fall 2008 Semester
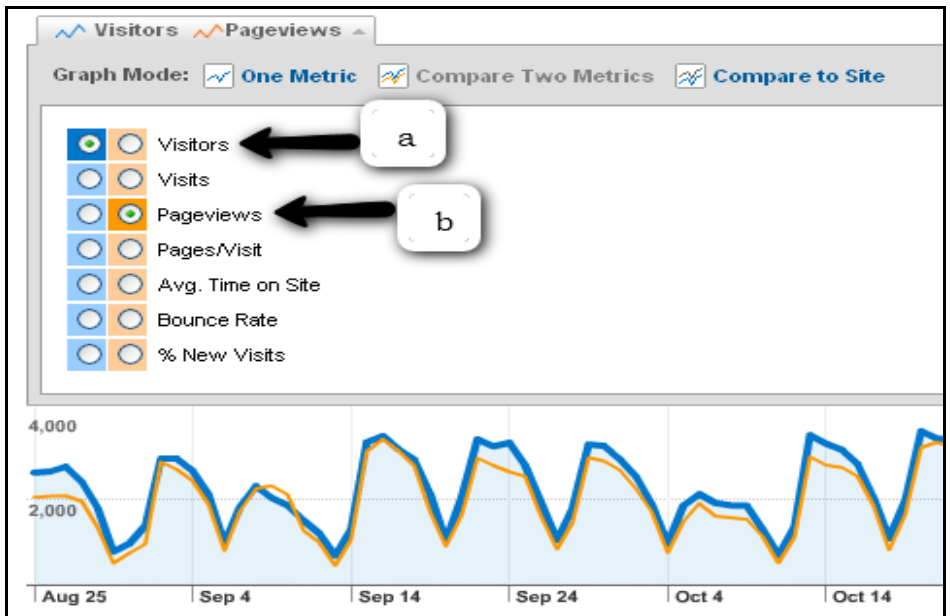
Figure 8. Visitors Trend Variation with Pageviews for Fall 2008 Semester

77

4.5.2 Browser capabilities

Analytics provided information on the machine configuration of the users of the Library

website under the *browser capabilities* section under *visitors*. A purpose of this section was to see

how different aspects of the site design and structure could affect how the visitors see, and use,

the website (Ledford and Tyler, 2007). The report provides a detailed overview of user machine

configuration in terms of the following mentioned metrics. Appendix A provides a summary of

definition for these metrics. This section covers an explanation of findings from the metrics

mentioned below.

- Operating Systems

- Browsers

- Browsers and OS Combination

- Screen Colors

- Screen Resolutions

- Flash Versions

- Java Support

Figure 9 -a shows that the users visited the MU library website using 24 browsers. Most

of the visitors used Internet Explorer (IE). More than 200,000 visits were recorded from visitors

who used IE (Figure 9 -b). Figure 10 provides a snapshot of the original analytics report that

provides the usage of the different Internet Explorer versions. Out of the Library visitors 81%

used IE 7.0 (Figure 10 -a), immediately following which were the Firefox users close to 170,000

visits (Figure 9 -c); the third most browsers used was Safari (Figure 9 -d).

Figure 9. Browsers that Visitors Used for the Fall 2008 Semester

	Browser Version		Visits	Visits
1.	■ 7.0	a	209,964	80.96%
2.	■ 6.0		48,363	18.65%
3.	■ 8.0		929	0.36%
4.	■ 5.5		38	0.01%
5.	■ 5.23		21	0.01%
6.	■ 5.01		11	> 0.00%
7.	■ 5.0		6	> 0.00%
8.	■ 5.14		4	> 0.00%
9.	■ 5.22		4	> 0.00%
10.	■ 5.16		3	> 0.00%

Figure 10. Different Internet Explorer Browser Versions used by the Library Visitors for Fall 2008 Semester

The percentage view, provided by analytics, presents a percentage-wise distribution of the figures reported. Chapter three detailed the different views available in analytics. Figure 11 provides the report of the different Operating Systems (OS) the visitors use to access the website. The percentage view of the OS show that the most used OS was Windows (Figure 11 -a), followed by Mac (Figure 11 -b) and the third most used OS was Linux (Figure 11 -c).

Overview »

Operating Systems

Aug 25, 2008 - Dec 19, 2008

Visits

Graph by:

August 25, 2008 | September 4, 20 | September 14, 2 | September 24, 2 | October 4, 2008 | October 14, 200 | October 24, 200 | November 3, 200 | November 13, 20 | November 23, 20 | December 3, 20(

508,606 visits used 13 operating systems

Site Usage | Goal Set 1

	Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
	508,606	3.33	00:03:02	29.30%	46.99%
	% of Site Total: 100.00%	Site Avg: 3.33 (0.00%)	Site Avg: 00:03:02 (0.00%)	Site Avg: 29.31% (-0.03%)	Site Avg: 46.99% (0.00%)

	Operating System	Visits	Visits
1.	Windows	389,387	76.56%
2.	Macintosh	116,370	22.88%
3.	Linux	2,124	0.42%
4.	iPhone	328	0.06%
5.	iPod	176	0.03%
6.	FreeBSD	128	0.03%
7.	(not set)	42	0.01%
8.	PalmOS	26	0.01%
9.	Android	11	> 0.00%
10.	SunOS	7	> 0.00%

Operating System contribution to total: Visits

a 76.56%

b 22.88%

c

Figure 11. Operating Systems of Library Users during Fall 2008 Semester

Figures 12, 13 and 14 provide cropped views of the machine settings of the visitors as described. Amongst the ones who used Windows more than 80% were XP users (Figure 12). The percentage view for the combination of browsers and OS shows that the most used combination was IE and Windows (Figure 13 -a), following which was Firefox and Windows (Figure 13 -b). Safari and Mac was the third most used combination (Figure 13 -c) followed by Firefox and Mac (Figure 13 -d).). The most used *screen resolutions* are shown in Figure 14 that were 1680 X1050 (Figure 14 -a) and 1024 X 768 (Figure 14 -b). Other than this the analytics reports showed that the most used screen color of the visitors machines was 32-bit, visitors were using Flash Version 9.0 r124 and 98% of users had Java support available to them.

	Visits 389,387 % of Site Total: 76.56%	Pages/Visit 3.41 Site Avg: 3.33 (2.58%)	Avg. Time on Site 00:02:43 Site Avg: 00:03:02 (-10.56%)	
	Operating System Version	None	Visits	Visits
1.	XP		338,719	86.99%
2.	Vista		48,386	12.43%
3.	2000		1,446	0.37%
4.	Server 2003		490	0.13%
5.	98		289	0.07%
6.	CE		24	0.01%
7.	ME		12	> 0.00%
8.	NT		12	> 0.00%
9.	95		8	> 0.00%
10.	(not set)		1	> 0.00%

Figure 12. Versions of Windows Operating System Used by Visitors for Fall 2008 Semester

Figure 13. Browsers and OS Combinations Used by Visitors

Figure 14. Screen resolutions of the Visitors Machines Accessing the Library Website

4.5.3   Network properties

Network properties can be further used to understand the characteristics of the visitor's network. It provides information that can help the web administrators design the site so that it is accessible to as many visitors possible (Ledford and Tyler, 2007). The *network properties* section provides information on visitors' *network location*, *hostname*, and their *connection speed*. Appendix A provides a definition of each of these metrics. The metric discussed here is the *network location* in order to find whether the visitors are accessing from on-campus or off-campus.

According to the Fall 2008 data for users' network locations, there were 356,241 users who accessed the website from on-campus using the university network cited as University of Missouri-Columbia. The percentage of visitors from the university network formed 70% of the visitors of the website (Figure 15 -a). A lesser number of users, about 43,600 who connected using Mediacom Communications Corporation formed almost 8% of the visitor population. Information on whether visitors came from staff machines could not be found since the *filter* option was not set for the Library. The filter option enables the excluding of some specific IPs or IP range from the main pool of visitor IPs. The *user defined* feature in analytics also provides a way to segregate users by their profile. Once set, the user defined feature under the *visitors* section could provide a way for the Library to know the visitors' profile. However, at this point since the user defined or the filter option was not set for the Library, investigation was only done for the analytics data that were collected and presented for the Fall 2008 semester.

	Visits			Pages/Visit			Avg. Time on Site	
	508,606			**3.33**			**00:03:02**	
	% of Site Total: 100.00%			Site Avg: 3.33 (0.00%)			Site Avg: 00:03:02 (0.00%)	

	Network Location	None ⌄		Visits ⌄ ↓	Visits
1.	■ university of missouri-columbia		a	356,241	70.04%
2.	■ mediacom communications corp		b	43,632	8.58%
3.	■ centurytel internet holdings inc.			22,434	4.41%
4.	■ at&t worldnet services			6,026	1.18%
5.	■ charter communications			5,659	1.11%
6.	■ mediacom communications corp.			4,576	0.90%
7.	■ morenet			4,340	0.85%
8.	■ road runner holdco llc			3,090	0.61%
9.	■ socket internet services corporation			2,606	0.51%
10.	■ level 3 communications inc.			2,494	0.49%

Figure 15. Network Location of Visitors of the MU Library Website during Fall 2008

4.6   *Traffic Sources*

The *traffic sources* section provides information on where the visitors are coming from on the web in terms of the access paths they take to the website. The options under *traffic sources* are *direct traffic, referring sites*, *search engines*, *all traffic sources*, *keywords*, *campaigns*. Appendix A provides a detailed overview of these metrics. Through the interview with the MU library web committee, its needs related to the visitors' access behaviors were apparent. They would have liked to know what access terms users access to come to their website, if they were entering the right spellings for the keywords they used, did users recall the location of the resources they visit, whether they accessed resources by bookmarking the links, and whether they searched every time they visit. They also wanted to know whether users accessed the website directly or Google searched the website, did users access using referring sites, and if they did then how many users accessed through which referring sites?

The metrics are discussed under the following three categories:

- *All traffic sources* provide an account of the percentage of user distribution through the different paths of access, all in one view.

- *Direct traffic* and *referring sites* provide information on user visits that came directly to the website through the *direct traffic* metric, and those who came from other referring sites through the *referring site* metric.

- *Search engines* and *keywords* show information on the search engine used to access the website and the keywords that were used respectively.

4.6.1 All traffic sources

In the consolidated information provided by *all traffic sources,* the different mediums of user access are presented. They were direct, through referring sites or using search engines. Figure 16 provides a percentage of the traffic sources. About 60% of users accessed the Library website directly (Figure 16 -a), a lesser percentage of users, about 18%, visited through the university home page (Figure 16 -b); and about 12% of visitors used Google to access the website (Figure 16 -c).

All traffic sources sent 508,606 visits via 704 sources and mediums

Show: Source Medium »

Site Usage | Goal Conversion

Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
508,606	**3.33**	**00:03:02**	**29.30%**	**46.99%**
% of Site Total: 100.00%	S... .3 (0.00%)	Site Avg: 00:03:02 (0.00%)	Site Avg: 29.31% (-0.03%)	Site Avg: 46.99% (0.00%)

Views:

Source/Medium contribution to total: Visits ›

12.19%

18.10%

60.24%

c

b

a

	Source/Medium	Visits	Visits
1.	(direct) / (none)	306,372	60.24%
2.	missouri.edu / referral	92,036	18.10%
3.	google / organic	62,018	12.19%
4.	laurel.lso.missouri.edu / re...	8,542	1.68%
5.	library.muhealth.org / referral	7,063	1.39%
6.	yahoo / organic	3,902	0.77%
7.	mulibraries.1cate.com / referral	2,671	0.53%
8.	healthlibrary.missouri.edu / referral	1,923	0.38%
9.	live / organic	1,730	0.34%
10.	runonce.msn.com / referral	1,693	0.33%

Filter Source/Medium: containing › [Go]

Go to: 1   Show rows: 10 ›   1 - 10 of 704

a   b   c

Figure 16. All Traffic Sources to the Library Website for Fall 2008 Semester

87

4.6.2 Direct traffic and referring sites

The *direct traffic* and *referring site* reveals the users' access behavior when they were navigating to the Library website. The *direct traffic* metric provides information on whether the visitors accessed the webpages of the website directly. The *referring sites* metric provides information on the different websites that the visitors came from. Figure 17 provides a cropped view of the original analytics *direct traffic* report based on the dimension, *visitor type.* Out of the direct visits on the website, 70% of the visitors were returning users (Figure 17 -a). Also, as depicted in Figure 17 -b, the direct traffic accessed 2.63 pages/visit. The average time spent on-site for the direct traffic was 02:37 (Figure 17 -c), and the bounce rate was at a high of 61.95% (Figure 17 -d). Figure18 provides a cropped view of the analytics report for visits through referring site. Traffic through referring sites showed 78% of the visitors as returning (Figure18 -a). The referring site visitors used an average of 4.67 pages/visit (Figure18 -b). The average time spent on site was 03:54 (Figure 18 -c), and the bounce rate was 18.42% (Figure18 -d).

Looking at the cropped view of the referring sites report in Figure 19, it shows that 69% of users' accessed through the university website (Figure19 -a). Some other referring sites that send, if not a huge but comparatively higher percentage traffic, were the Health Sciences library, about 5% (Figure19 -b) and that of the Journalism school website of 1.15% (Figure19 -c).
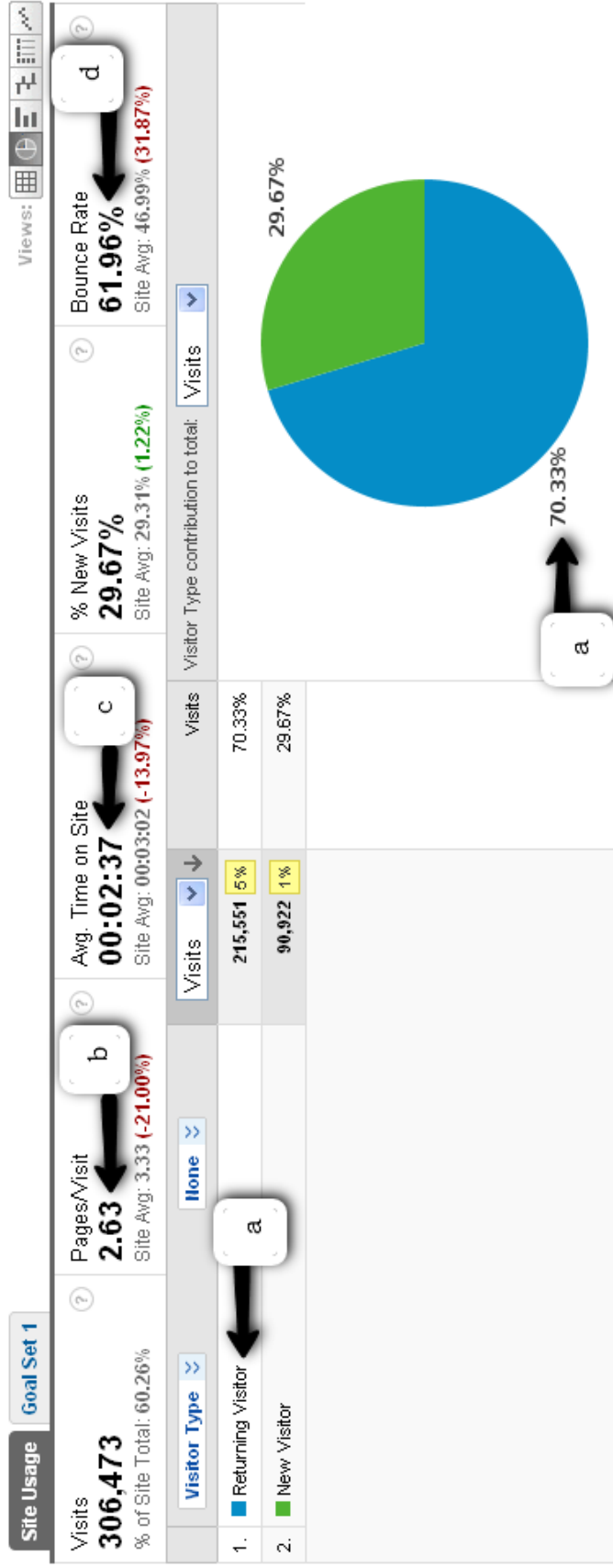
Figure 17. Direct Traffic Details by Visitor Type for Fall 2008 Semester
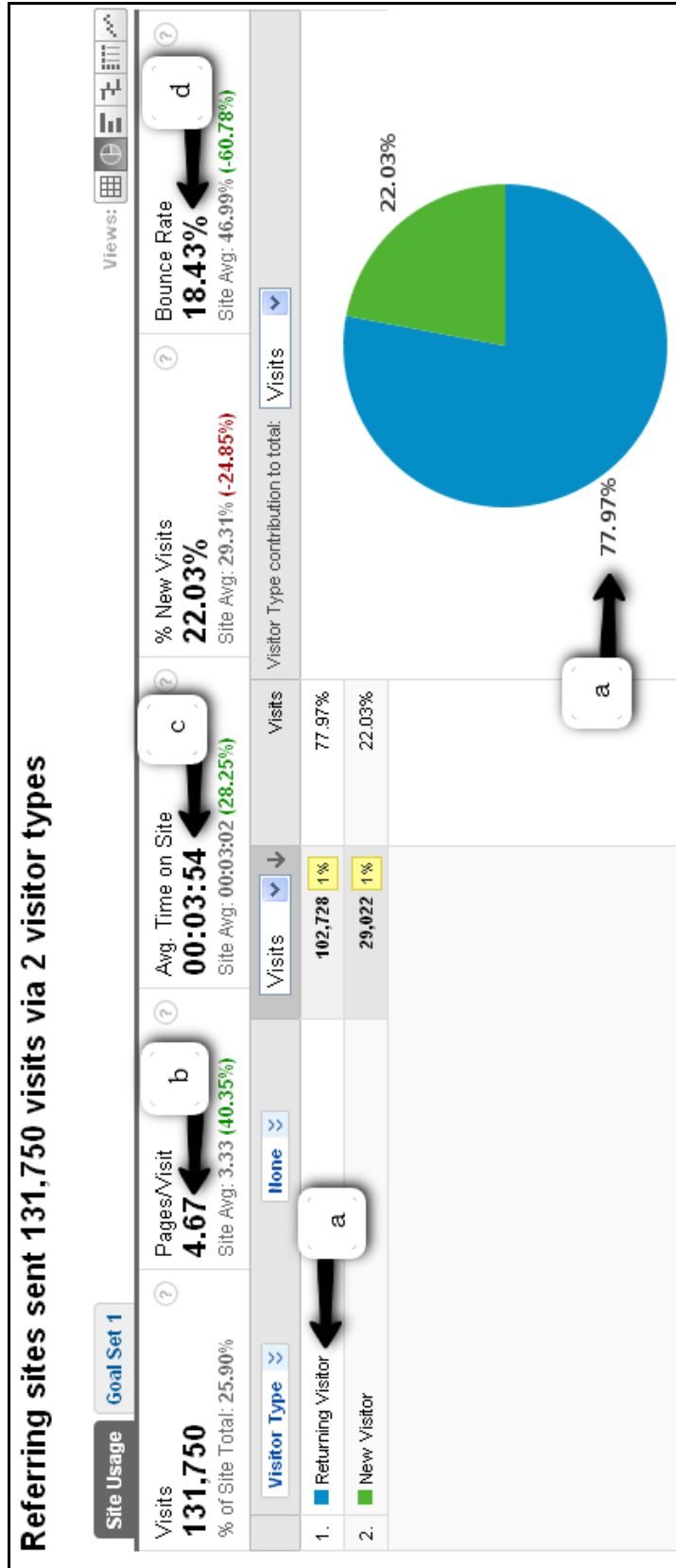
Figure 18. Referring Site Traffic Shown With the Visitor Type Dimension for Fall 2008
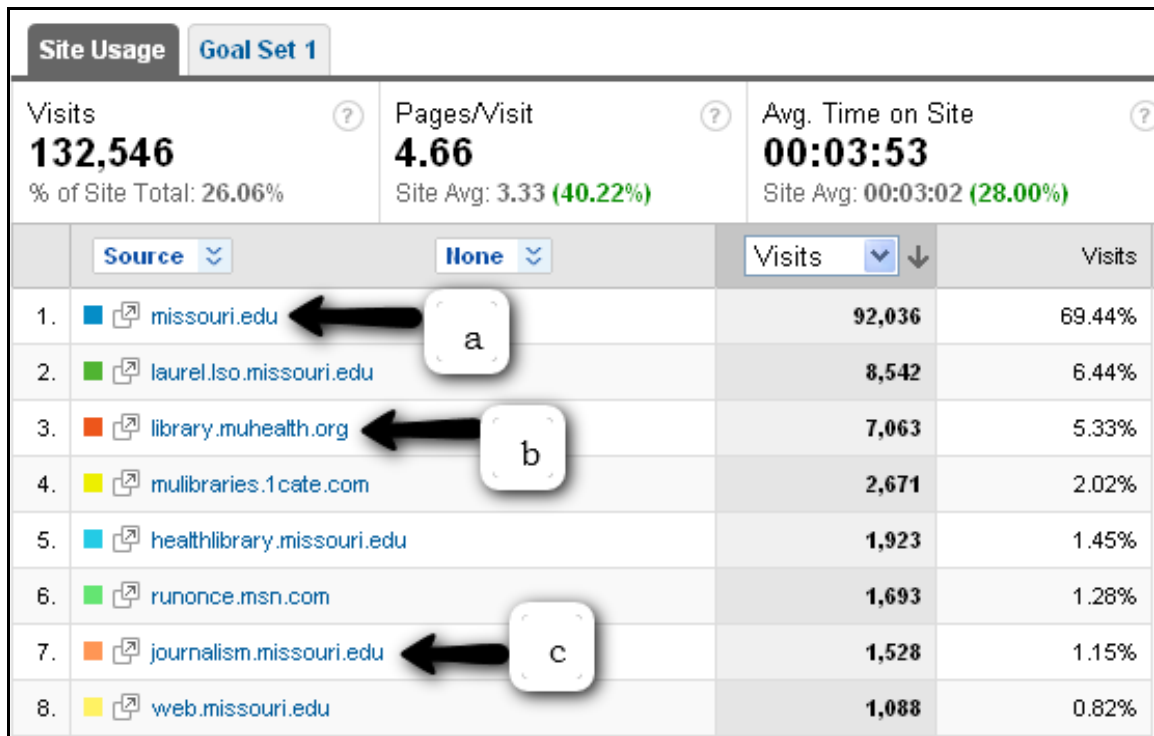
Figure 19. Referring Sites that Directed Traffic to the Library Website during Fall 2008 Semester

4.6.3 Search engines and keywords

The analytics report on *search engines and keywords* provides information on the search engines the visitors used and the keywords that they entered to access the website respectively. Figures 20 shows a full view of an analytics report that provides information on the visitors search preferences. Figures 21 and 22 provide a cropped view of the original analytics reports on the landing pages the visitors come to when conducting a search and the terms that they use in their searches respectively. As mentioned before, 12 % of the visitors who used the Library website Googled to the website (Figure 16 -c). Figure 20 -a shows 89% of the website audience Googled to the MU library website. These users visit 3.77 pages per visit (Figure 20 -b) and spent an average time on site of 03:13 minutes (Figure 20 -c).

91

A majority of the visitor types were returning users to the website. Almost 50% of the searches ended up on the Library's main gateway (Figure 21 -a). Also, as shown in Figure 22, the top three keywords used to access the Library website were "ellis library," "mu libraries," and "mizzou library". Other combinations of the term mu, mizzou, university of missouri along with other terms such as library or libraries, were used. These words were sometimes also used as one word such as "mulibraries." Another key term that users employed to come to the Library website was of the database "web of science" that got 891 visits in the Fall 2008 semester.
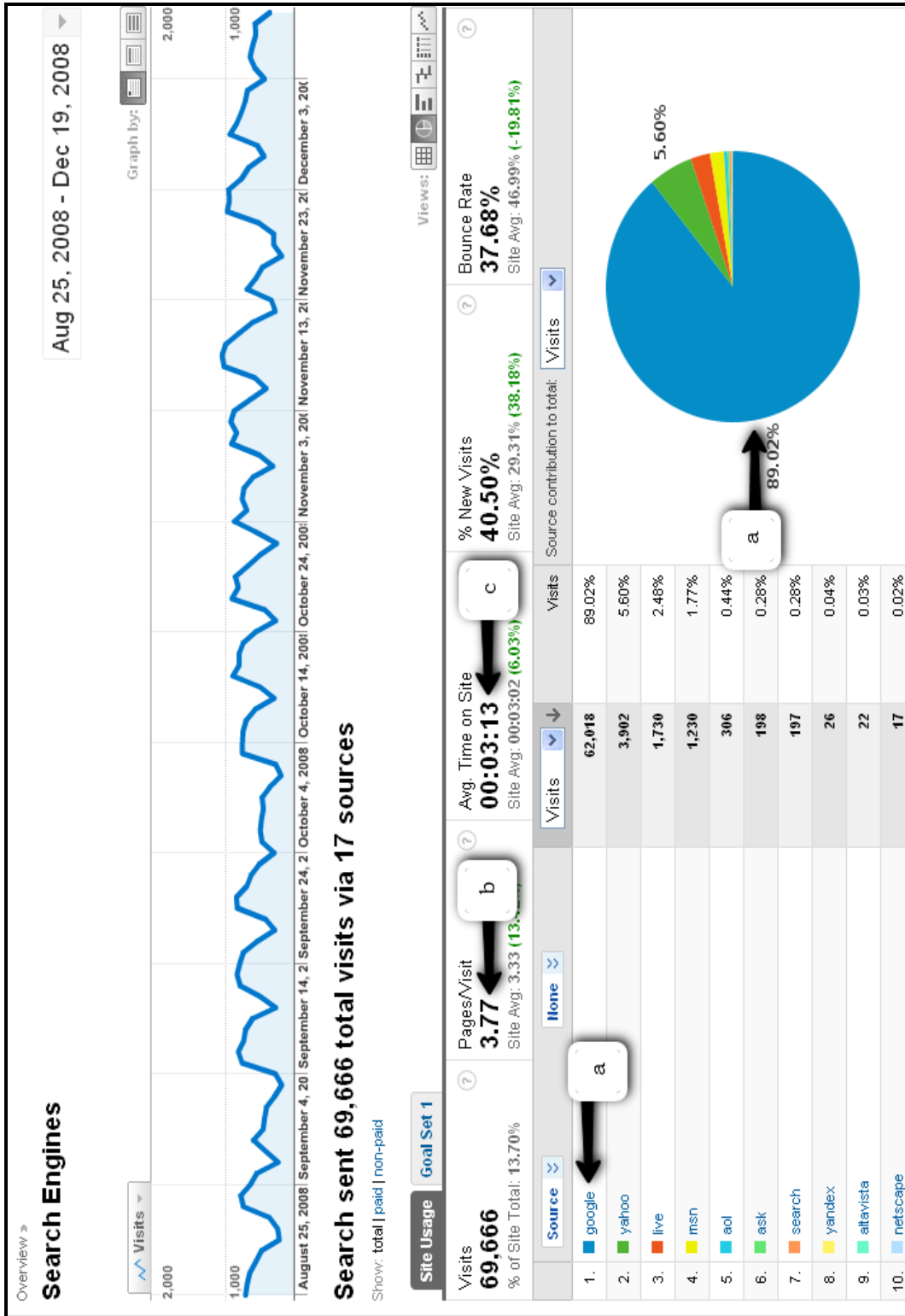
Figure 20. Percentage of Search Engines Used by the Visitors

		Visits 68,348 % of Site Total: 13.44%	Pages/Visit 3.79 Site Avg: 3.33 (13.85%)	Avg. Time on Site 00:03:15 Site Avg: 00:03:02 (6.78%)	
	Landing Page	None	Visits		Visits
1.	/default.aspx		34,800	2%	50.92%
2.	/search/databases/default.aspx		2,909	8%	4.26%
3.	/journalism/default.aspx		2,279	9%	3.33%
4.	/engineering/default.aspx		1,435	20%	2.10%
5.	/collections/documents/mo/stat_az.htm		1,322	6%	1.93%
6.	/guides/reserves/reserves-student.htm		553	10%	0.81%
7.	/guides/reserves/default.htm		546	10%	0.80%
8.	/search/databases/main.asp		518	19%	0.76%
9.	/search/books/findbooks.htm		455	10%	0.67%
10.	/guides/subjects/chemistry/words.htm		438	10%	0.64%

Figure 21. Landing Pages the Users got to Using Search during the Fall 2008 Semester

		Visits 69,666 % of Site Total: 13.70%	Pages/Visit 3.77 Site Avg: 3.33 (13.42%)	Avg. Time on Site 00:03:13 Site Avg: 00:03:02 (6.03%)	
	Keyword	None	Visits		Visits
1.	ellis library			4,615	6.62%
2.	mu libraries			4,440	6.37%
3.	mizzou library			3,070	4.41%
4.	mu library			2,350	3.37%
5.	university of missouri library			1,710	2.45%
6.	mizzou libraries			1,133	1.63%
7.	mulibraries			1,012	1.45%
8.	web of science			891	1.28%
9.	library			867	1.24%
10.	"anne barker"			835	1.20%

Figure 22. Keywords Used to Access the Library during the Fall 2008 Semester

4.7   *Content*

The *content* section provides information on how visitors use the different pages of the website. This section is further explained by the metrics on *top content, content by title, content drilldown, top landing pages, top exit pages, site overlay, and site search.* Appendix A provides a summary of definition for each of these metrics.

The interview with the MU library web committee provided more information on which of the features and metrics provided by Google analytics could be used to explain the users' behavior under the *content* section. The respondents mentioned they would like to know if resources that the Library provide were being used by the visitors of the website, and whether these visitors could access these resources without any problems. The MU library web committee also wanted to know the users' specific navigation behavior on the website, and the reasons for it. For example, if a user is on a webpage for a long time, what keeps them for a long time on that page, are the visitors lost on a page, are they leaving from a page or going back from that page? The Library web committee wanted to know the continuity of actions between the layers of the page. Information on whether users were directly accessing a page by bookmarking it, or if they were browsing through certain expected paths, would have been useful for the Library's decision-making needs. They would like to know if any links on the website were redundant, how users searched the website, using which terms and the selections from the drop down menu. They were also curious about the usage of the catalog on the Library website, and if the visitors shifted to searching information within the website.

The metrics discussed in this section are further grouped under the following categories:

- *Top content* provides information on the most accessed pages of the website in descending order of *pageviews* by default. Other metrics presented for each record in the table view are *unique pageviews, average time on page, bounce rate* and *% exit*.

- *Content by title* provides information on the groups of pages that have the same title as supplied by the metrics mentioned above for top content.

- *Top landing pages* and *top exit pages,* as the names suggest, present information on the pages where visitors arrive on the website and the pages from where the visitor leaves the website.

- *Site overlay* directly shows navigational information on the website.

- *Navigation summary* provides information on the navigational path of the users. This feature shows how often were webpages an entrance page, what webpages the users visited before were, and where they went eventually.

4.7.1   Top content

Top content provides information on the topmost visited URLs of the website. Figure 23 presents a view of the analytics report for *top content*. Figure 24 and 25 provide a cropped view of the analytics report that shows the top content page contribution to *exits* and *time on page* respectively. According to Figure 23, the percentage view of analytics' *top content* report provides unique pageviews. The report presents the unique pageview contribution to the total unique pageviews of the website as shown in Figure 23 -e. A *unique pageview* represents the number of sessions during which that page was viewed one or more times (Google analytics, 2009). Figure 23 -a, shows that the user prefers to go through the home page of the Library website when accessing the Library resources. Almost 32% of the unique pageviews were reported on the home page. As indicated in Figure 23 -b, a second page that got a 7.72% unique pageview contribution in the Library website was the database page. The database page lists the online databases recommended by librarians. A low percentage of unique pageview contribution of 2.92% was recorded for the Merlin catalog page. The Journalism library website recorded a 2.74% unique pageview contribution (Figure 23 -c, d).

Figure 24 presents the *exit rate* (Figure 24 -a) contribution to the total *exits* for the Library webpages. For the most viewed page, which was the Library home page, the exit rates were the highest and contributed to almost 40% of the total exits (Figure 24 -b). The *exit rate* contribution for the comparatively less-viewed Journalism page was 5.05% (Figure 24 -c). Also the figure shows that the exit rate for the second most viewed database page and the Merlin catalog page was less than 5%.

Figure 25 -a shows that the most time spent was on the Library's home page, and the contribution to total was of 28.27% (Figure 25 -b). Figure 25 -c,d,e show the percentage contribution on time-on-site for the rest of the three most viewed pages: the database page that had a time-on-page contribution of 6.94%, the Merlin page and the Journalism page that had a time-on-page contribution of 6.64% and 5.14% respectively.
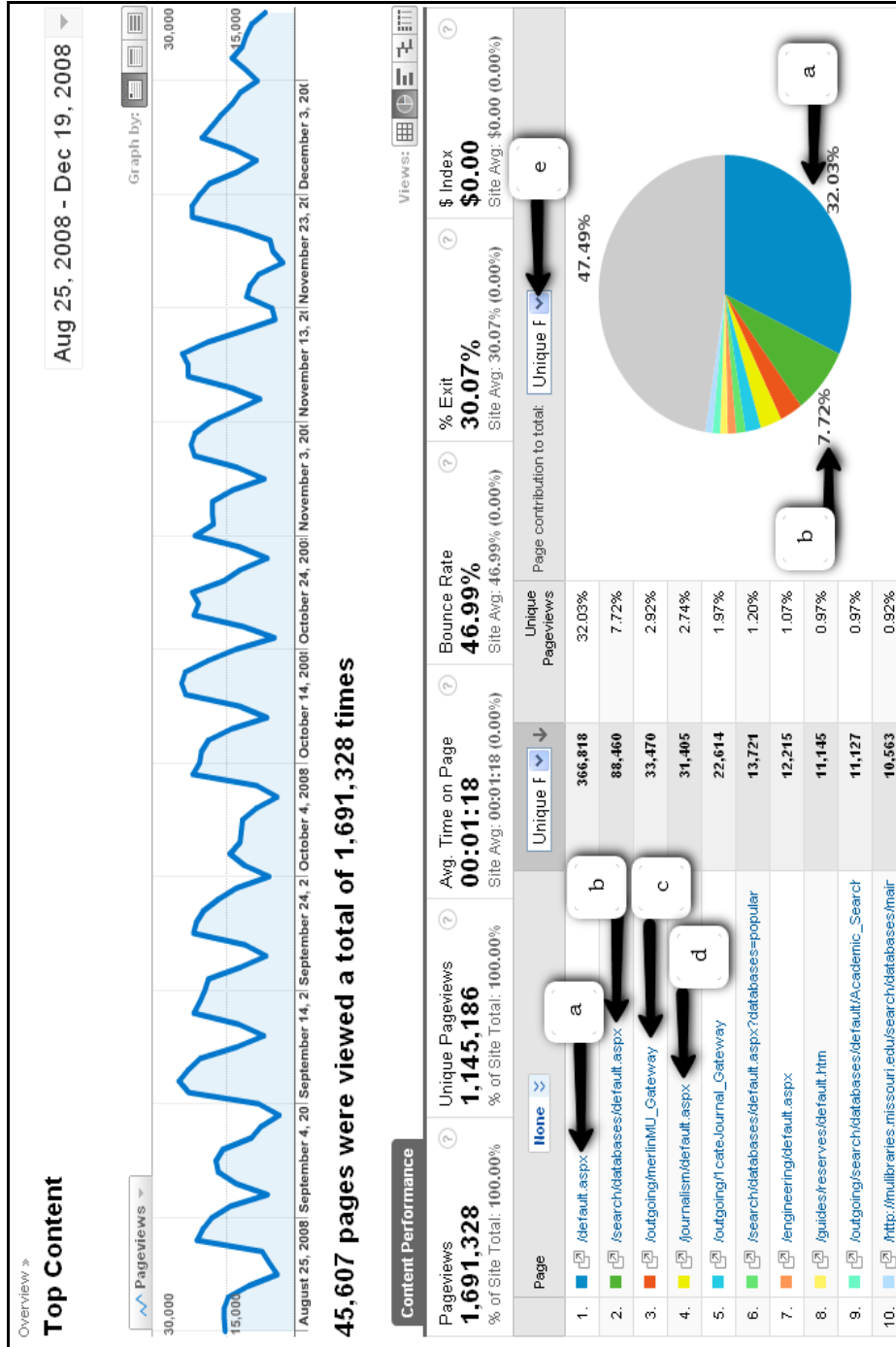
Figure 23. Unique Pageviews Contribution to the Total Unique Pageviews for the Top Content for Fall 2008 Semester
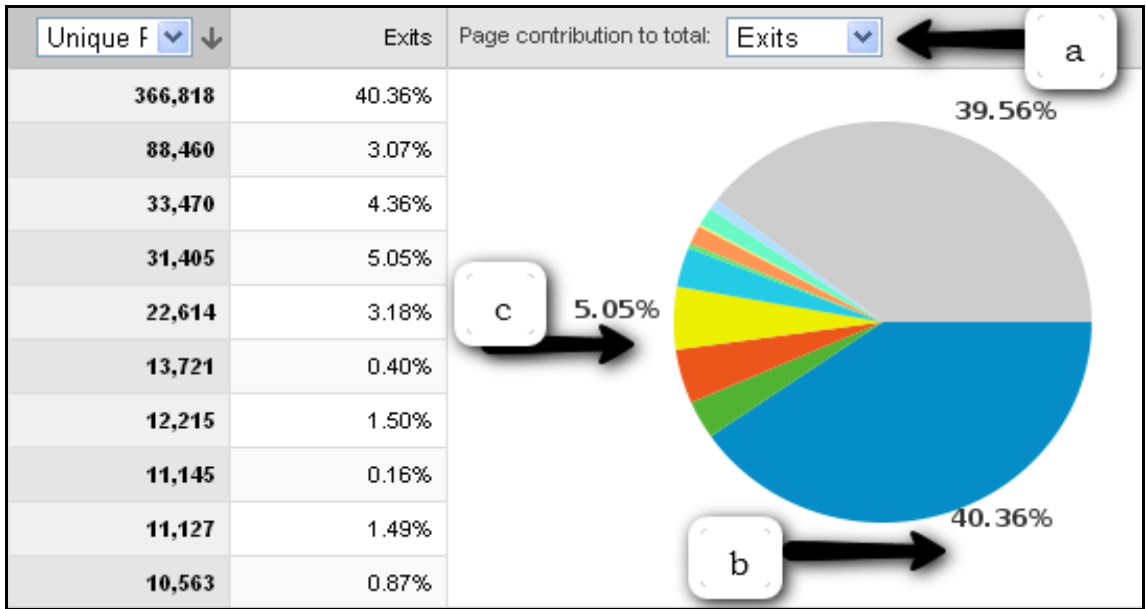
Figure 24. Exit Rate Contribution of the Most Viewed Page for Fall 2008 Semester
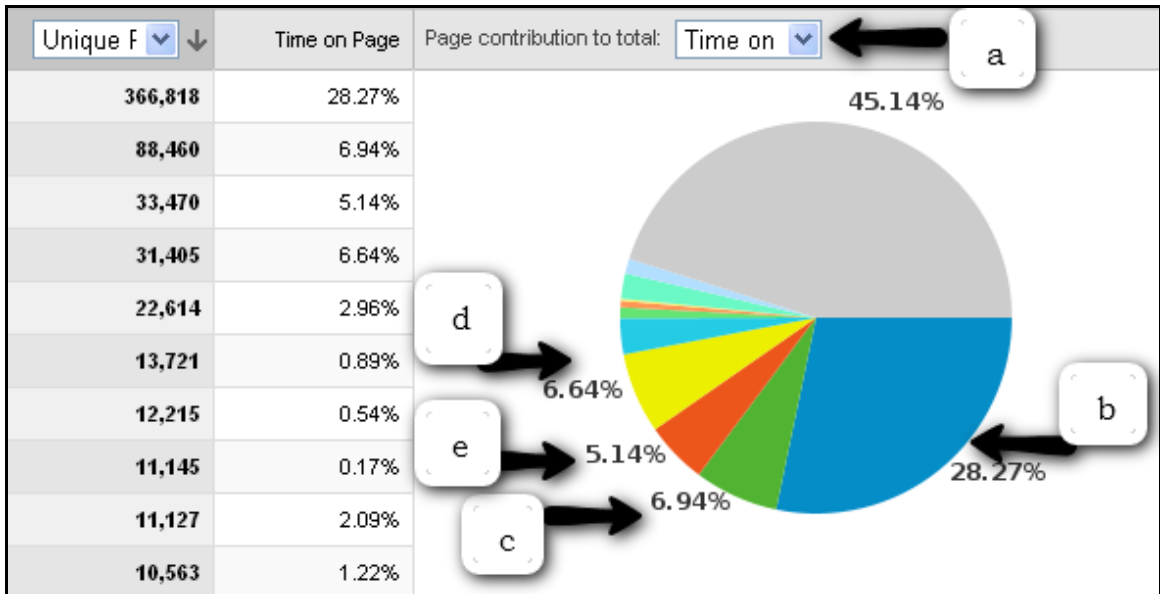
Figure 25. Time on Page Contribution of the Most Viewed Pages for Fall 2008 Semester

4.7.2 Content by title

The *content by title* option provides information on groups of pages that have the same title. Figure 26 provides a snapshot of the unique pageviews for the *content by title* report. Figure 27 and 28 are a cropped view of the original analytics report that provides information on the unique pageviews contribution for the *University of Missouri* groups of pages and *find articles and databases* groups of pages respectively. Looking at the performance of groups of pages, the two top performance groups of pages were those of "University of Missouri Libraries" (Figure 26 –a) that had a 38% *unique pageviews* contribution (Figure 26 -b), and "find articles and databases" that had an 11% *unique pageviews* contribution (Figure 26 -c). Some other top performance groups of pages were the pages that belonged to the titles "search," "journalism library," and "eRes and print reserves" that had less than 5% of *unique pageviews* contribution.

Further, looking at the groups of pages for University of Missouri Libraries, the default page had 80% of the unique pageview contribution (Figure 27 -a), the Merlin catalog had 7.66% contribution (Figure 27 -b), and the link to journals had around 5% unique pageview contributions (Figure 27 -c).

Figure 28 -a shows the default database page that users visited most had almost 50% of unique pageview contribution (Figure 28 -b). Users also viewed the popular databases with unique pageview contributions of almost 8% (Figure 28 -c). Some other databases mentioned in the report in Figure 28 that had unique pageview contributions between 2-7% are Academic Search Premier, JSTOR, Lexis-Nexis Academic, and Psycinfo.
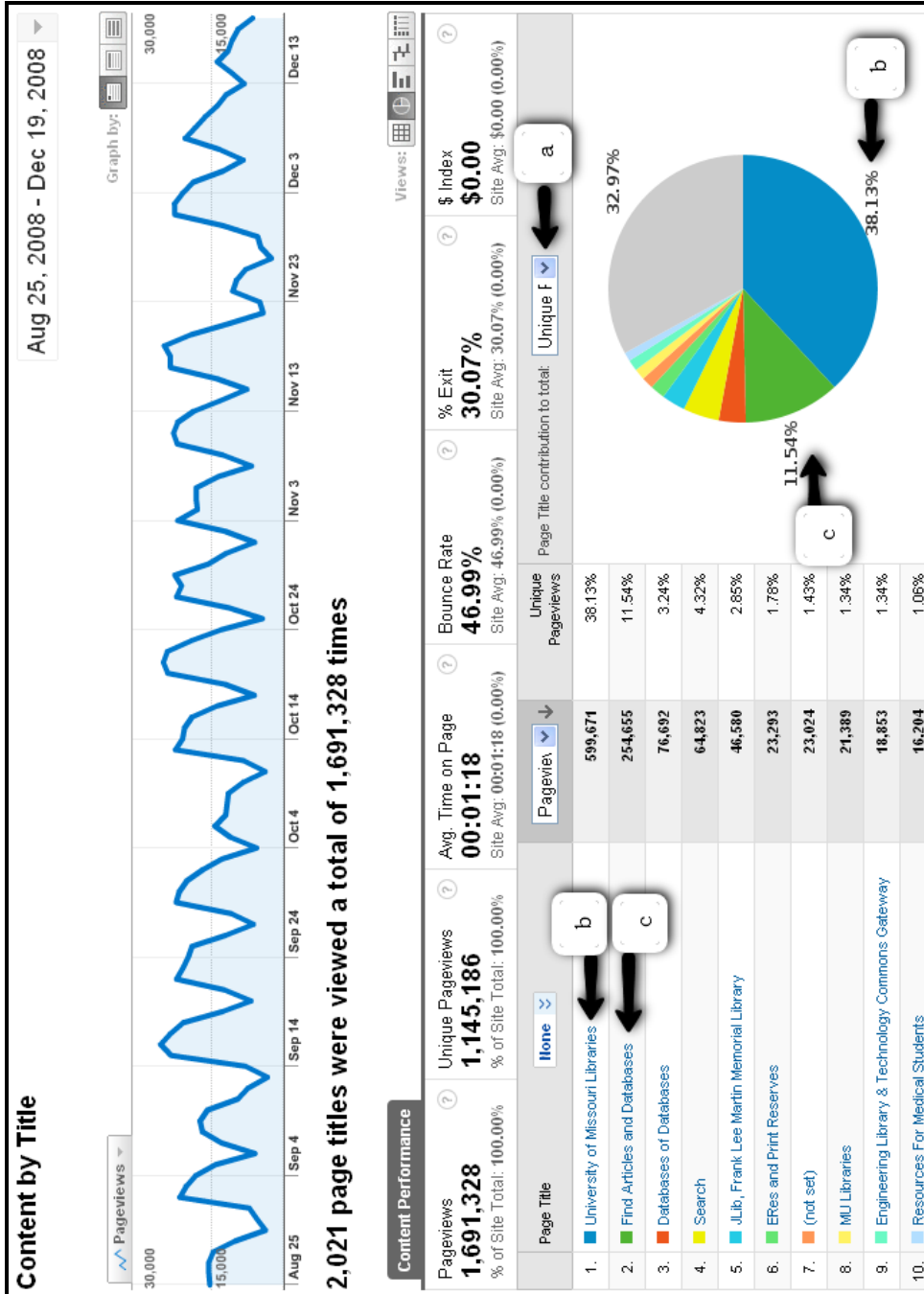
Figure 26. Content by Title Unique Pageview Contribution for Fall 2008 Semester
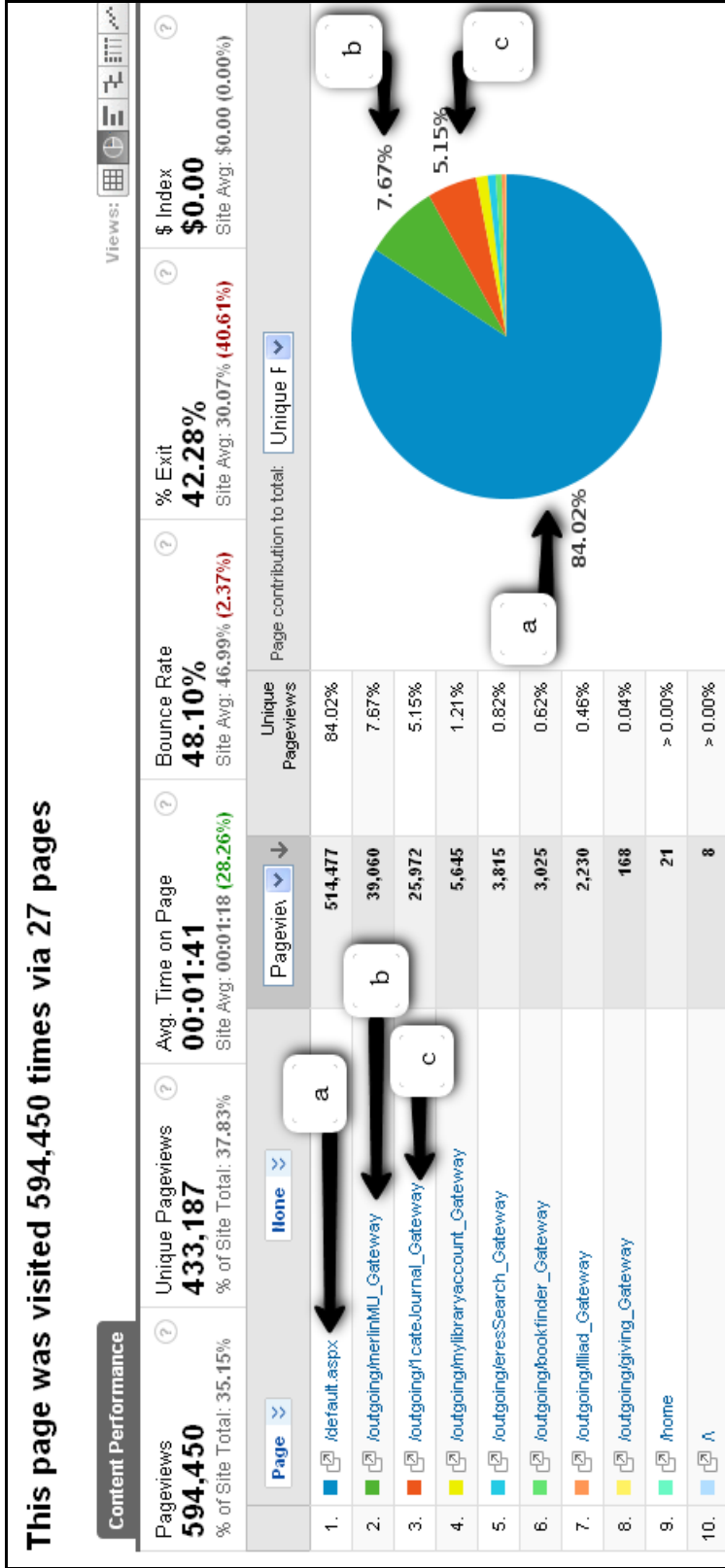
Figure 27. Unique Pageviews Contribution for the "University of Missouri Libraries" Groups of Pages for Fall 2008 Semester
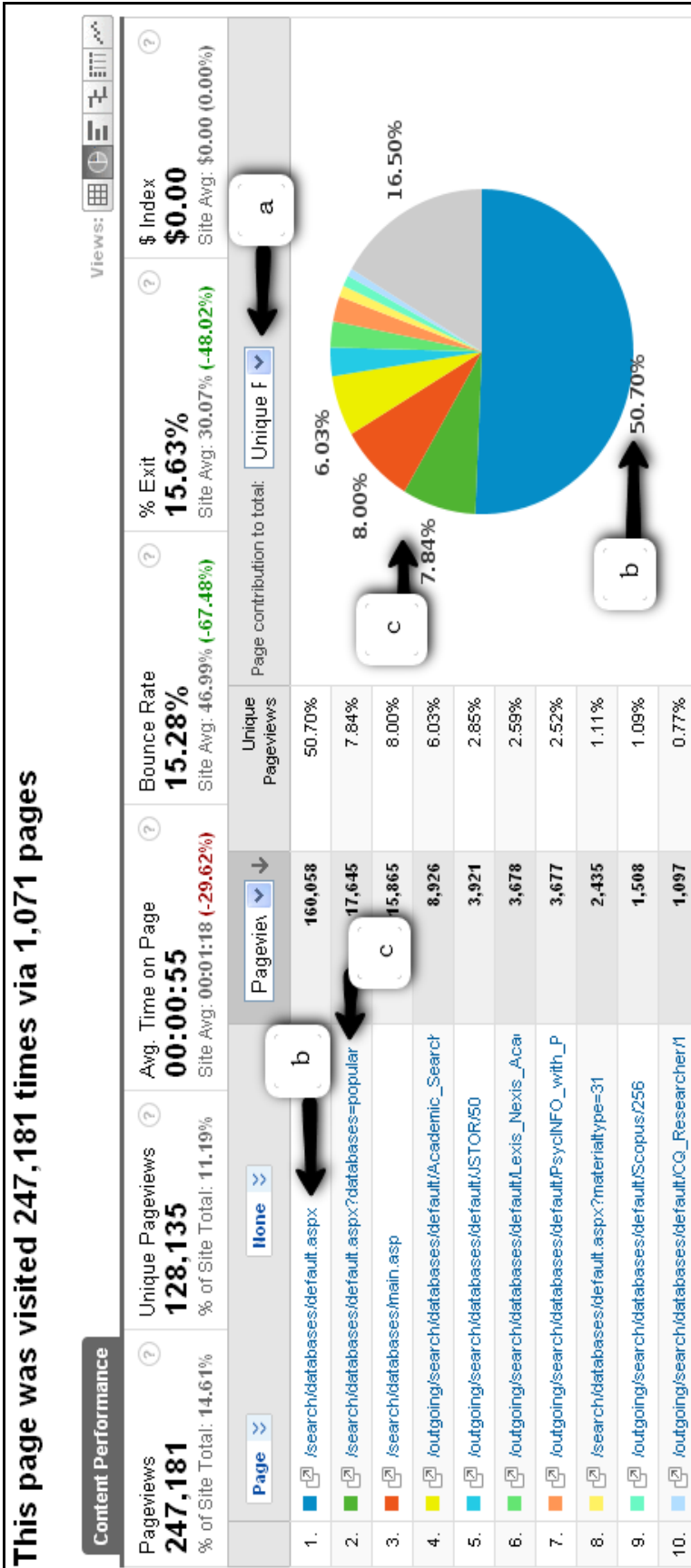
Figure 28. Unique Pageviews contribution for the "Find Articles and Databases" Group of Pages for Fall 2008 Semester

4.7.3   Top landing pages and top exit pages

The *top landing pages* category provides information on the pages that the visitors arrive at on the website. *Top exit pages*, on the other hand, provide information on the pages from which the visitors leave the website. Figure 29 provides the analytics report for the *top landing page* (Figure 29 -a) contribution to total entrances. The report presents information on the percentage each page contributes to the total website entrance page. According to the report 70% of the total users visiting the Library website entered the website through the home page (Figure 29 -b). Other figures suggest that less than 6% of the website visitors landed on the pages which were Journalism, database, engineering or a page in the health sciences library.

A cropped view of the analytics report for the *top exit pages* are reported by the contribution to total exits (Figure 30 -a) in Figure 30. Based on the report, the home page had a percentage exit contribution of about 40% (Figure 30 -b). Some other pages such as the Journalism, catalog and eJournals recorded a far lower exit rate contribution of less than 5%.
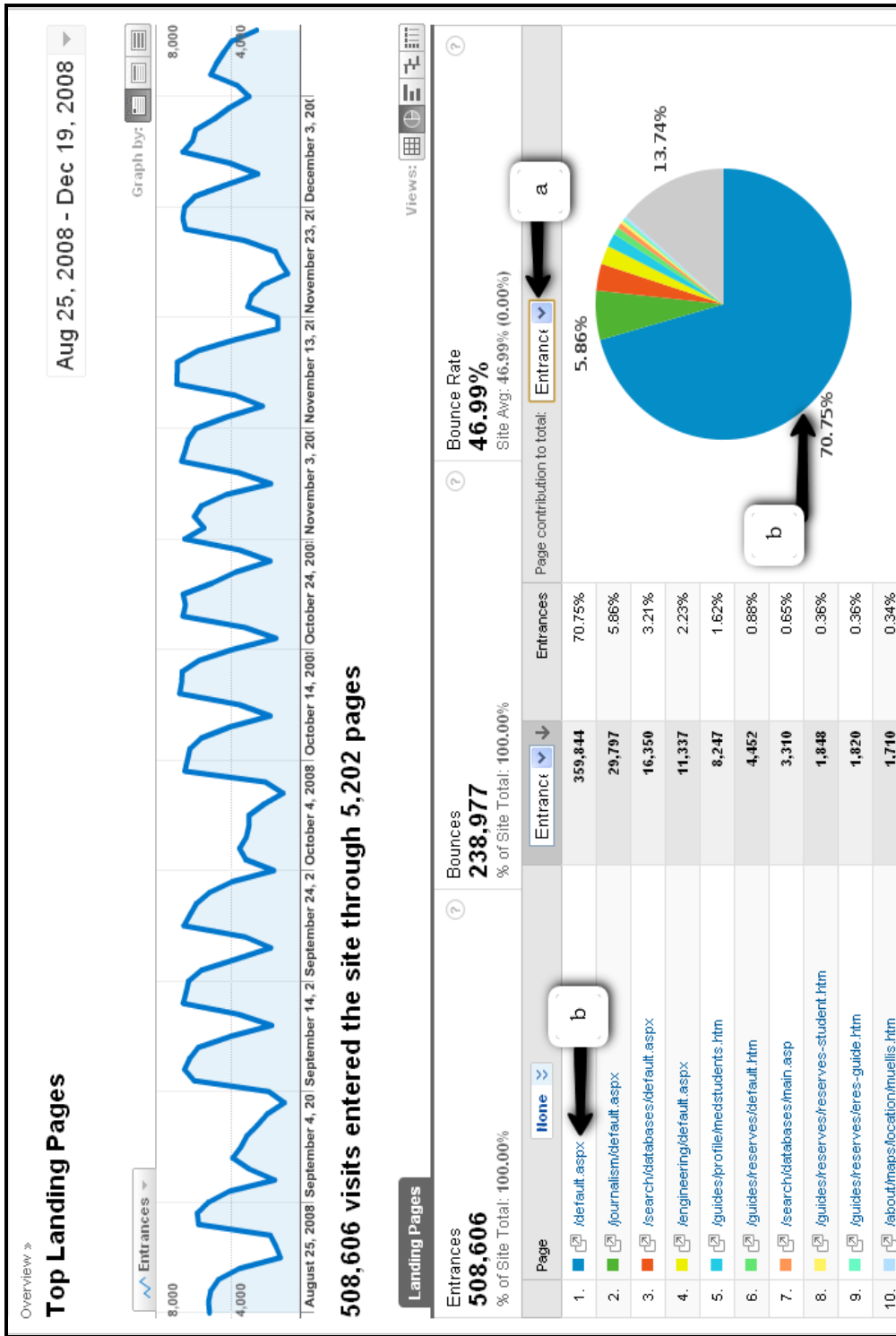
Figure 29. Top Landing Pages of the Library Website by Contribution to Total Entrances for Fall 2008 Semester
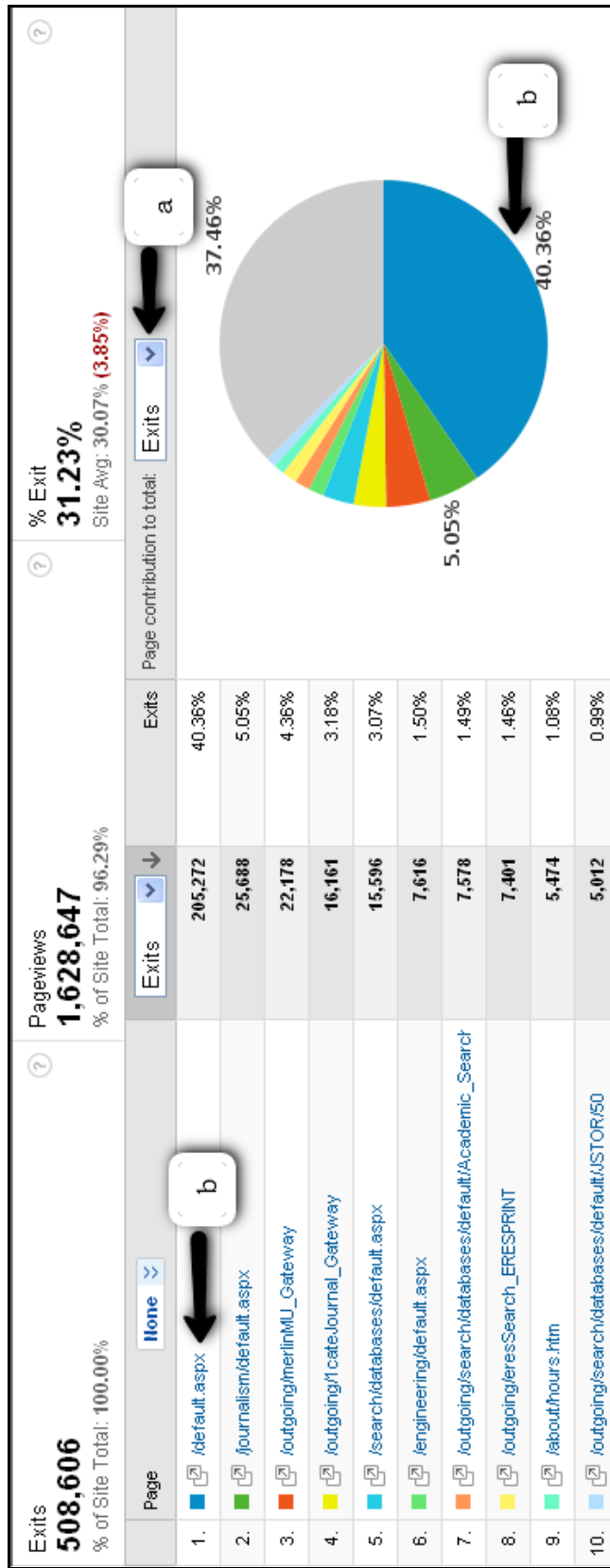
Figure 30. Top Exit Pages of the Library Website by Contribution to Total Exits for Fall 2008 Semester

4.7.4   Site Overlay

The *site overlay* information provides percentages on the clicks on the different hyperlinks on the Library website (Figure 31). Looking at the site overlay report of the Library website, most of the user clicks were made on the "find database" and "search for articles on a topic." Both the links got a percentage click of 26% (Figure 31 -a, b). The other links on the page showed a percent click of less than 3%, with some of the links showing percent clicks as low as 0.2%.
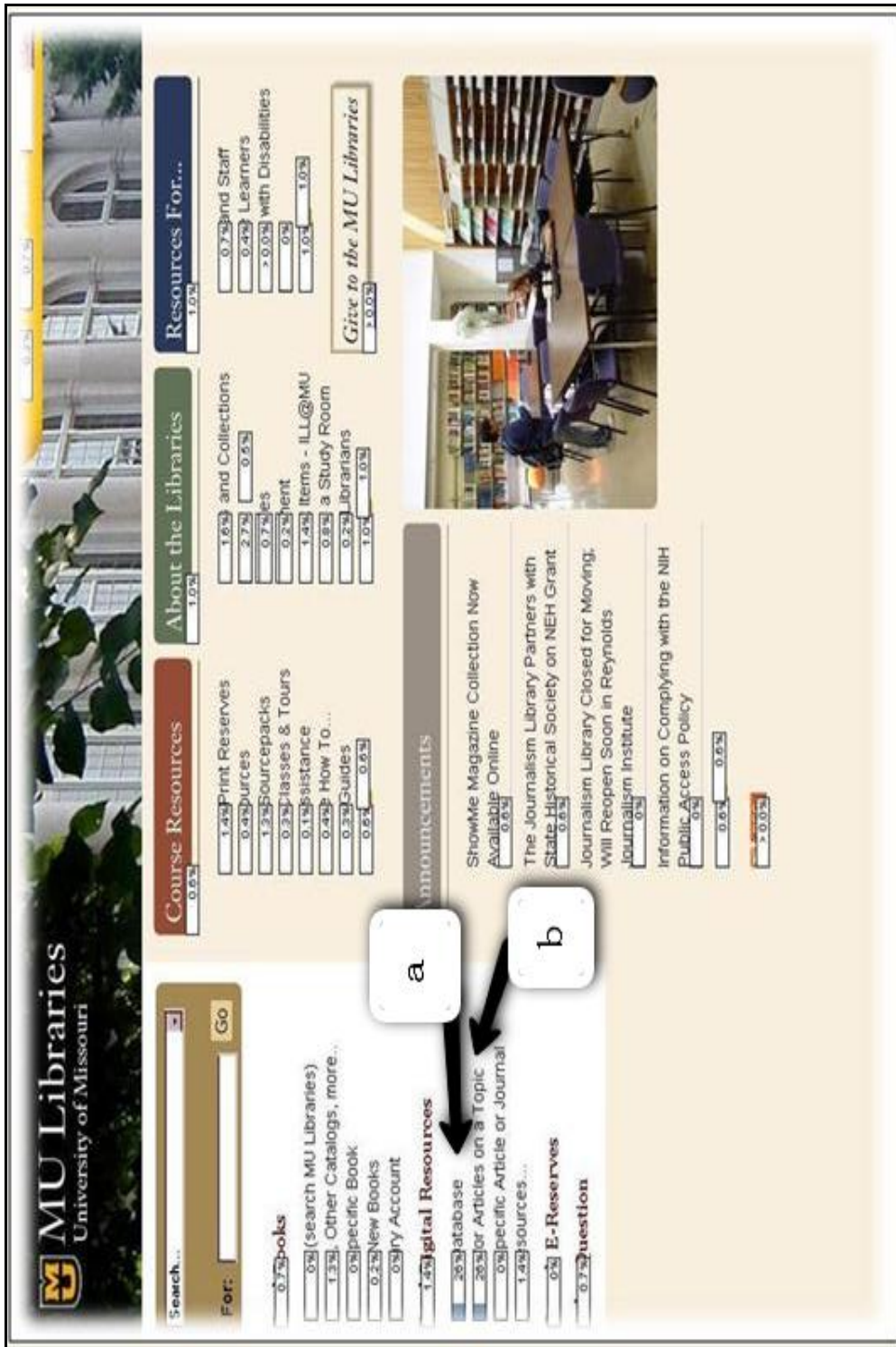
Figure 31. Site Overlay of the Library Website

4.7.5 Navigation Summary

The *navigation summary* option in analytics provides a visual graph of the users browsing patterns through the Library website. It shows a page by page sequence of users actions and provides information on whether the user exited from the page. Figure 32 provides a navigation summary report. Figures 33, 34 and 35 provide cropped view of the original navigation summary report for popular databases page, academic search premier page and the Merlin catalog page.

Most of the users went to the database page from the Library's home page (Figure 32 -a). The database page recorded the second most unique pageviews of 7.72% after the home page (Figure 23 -b). Looking at the *navigation summary* of the database page about 11% of the users exited from the page (Figure 32 -a) and 88% went on to a next page (Figure 32 -b). Out of the ones who moved to the subsequent page about 11% of the visitors went to "popular databases" (Figure 32 -c) from where 43% ended up in Academic Search Premier (Figure 33 -a), and 11% ended up in Lexis Nexis Academic (Figure 33 -b). About 60% of the visitors who went to the Academic Search Premier exited the Library website (Figure 34 -a), another 16% approximately went to the home page (Figure 34 -b), and about 15% of the visitors went back to popular databases (Figure 34 -c).

Further looking up the navigation summary for the Merlin Catalog (Figure 35), more than 50% of the users exited after they clicked on the catalog (Figure 35 -a) and about 38% ended back to the home page (Figure 35 -b).

Overview » Top Content »

Navigation Summary:
/search/databases/default.aspx

Aug 25, 2008 - Dec 19, 2008

Pageviews

4,000

2,000

Graph by:

4,000

2,000

August 25, 2008| September 4, 20| September 14, 2| September 24, 2| October 4, 2008| October 14, 2008| October 24, 200| November 3, 20| November 13, 2| November 23, 2| December 3, 20|

This page was viewed 221,986 times

Visit this page   Analyze: **Navigation Summary** »   Content /search/databases/default.... »

12.34% Entrances

87.66% Previous Pages

Content	% Clicks
/default.aspx	64.57%
/search/resources/default.htm	1.60%
/search/databases/default.aspx?dat...	1.40%
http://mulibraries.missouri.edu/searc...	0.87%
/search/articles/findarticles.htm	0.60%
/outgoing/search/databases/default/...	0.59%
/outgoing/search/databases/default/...	0.57%
/guides/subjects/business.htm	0.43%

11.77% Exits

88.23% Next Pages

a

b

c

/search/databases/default.aspx?databases=popular

Content	% Clicks
/search/databases/default.aspx?da...	10.96%
/default.aspx	9.62%
/outgoing/search/databases/default/...	3.70%
/outgoing/search/databases/default/...	3.26%
/guides/subjects/business.htm	3.11%
/guides/subjects/psychology.htm	2.87%
/guides/subjects/education.htm	2.34%
/outgoing/search/databases/default...	2.18%

Figure 32. Navigation Summary for the Database Page for Fall 2008 Semester

This page was viewed 24,128 times

Visit this page   Analyze: **Navigation Summary** »   Content   /search/databases/default.... »

2.33% Entrances

9.82% Exits

97.67% Previous Pages

90.18% Next Pages

/outgoing/search/database/default/Academic_Search_Premier/99

/outgoing/search/database/default/Lexis_Nexis_Academic/295

a

b

Content	% Clicks
/search/databases/default.aspx	70.94%
/outgoing/search/databases/default/...	9.11%
/outgoing/search/databases/default/C...	3.54%
/outgoing/search/databases/default/L...	3.36%
/outgoing/search/databases/default/...	2.05%
/search/databases/default.aspx?subj...	0.84%
/search/databases/default.aspx?mat...	0.50%
/default.aspx	0.45%
/search/databases/default.aspx?mat...	0.40%
/search/databases/default.aspx?mat...	0.23%

Content	% Clicks
/outgoing/search/databases/default/...	43.21%
/outgoing/search/databases/default/...	11.11%
/search/databases/default.aspx	9.09%
/outgoing/search/databases/default/	6.44%
/outgoing/search/databases/default/...	4.06%
/default.aspx	3.24%
/outgoing/search/databases/default/...	0.83%
/outgoing/search/databases/default/...	0.44%
/guides/courses/sourcepacks.htm	0.43%
/guides/subjects/psychology.htm	0.40%

Figure 33. Navigation Summary for the Popular Database Page for Fall 2008 Semester

This page was viewed 12,788 times

Visit this page Analyze: [Navigation Summary »] Content [/outgoing/search/database... »]

0.06% Entrances

99.94% Previous Pages

Content	% Clicks
/search/databases/default.aspx?dat...	70.08%
/search/databases/default.aspx	22.90%
/search/website/search.aspx?cx=0...	1.00%
/search/databases/default.aspx?mat...	0.88%
/search/website/search.aspx?cx=0...	0.83%
/search/databases/moreinfo.aspx?b...	0.76%
/search/databases/moreinfo.aspx?id...	0.67%
/search/website/search.aspx?cx=0...	0.51%
/search/website/search.aspx?cx=0...	0.38%
/search/website/search.aspx?cx=0...	0.31%
/search/databases/default.aspx?mat...	0.21%

60.04% Exits

39.96% Next Pages

a

b

c

Content	% Clicks
/default.aspx	15.87%
/search/databases/default.aspx?da...	14.77%
/search/databases/default.aspx	6.02%
/journalism/default.aspx	0.31%
/search/databases/default.aspx?m...	0.19%
/search/databases/default.aspx?su...	0.17%
/search/databases/moreinfo.aspx?i...	0.17%
/search/databases/moreinfo.aspx?...	0.16%
/about/maps/location/muellis.htm	0.14%
/search/website/search.aspx?cx=0...	0.13%

Figure 34. Navigation Summary for the Academic Search Premier Page for Fall 2008 Semester

This page was viewed 39,393 times

Visit this page | Analyze: **Navigation Summary** >> | Content | **/outgoing/merlinMU_Gateway** >>

0.56% Entrances

99.44% Previous Pages

56.68% Exits (a)

43.32% Next Pages (b)

Content	% Clicks
/default.aspx	97.94%
/search/books/catalogs.htm	0.36%
/search/databases/default.aspx	0.25%
/outgoing/eresSearch_Gateway	0.11%
/staff/default.htm	0.06%
/search/resources/default.htm	0.04%
/guides/reserves/default.htm	0.04%
/outgoing/1cate/Journal_Gateway	0.03%
/collections/default.htm	0.03%
/about/hours.htm	0.03%

Content	% Clicks
/default.aspx	37.87%
/about/maps/location/muellis.htm	2.12%
/collections/annex-depository.htm	0.53%
/search/databases/default.aspx	0.28%
/search/website/search.asp?page=...	0.16%
/guides/virtour/selfguided/tour.asp?ti...	0.15%
/journalism/default.aspx	0.15%
/guides/virtour/childrensbooks.htm	0.14%
/showmehow/findemplid.htm	0.13%
/about/maps/location/annex.htm	0.11%

Figure 35. Navigation Summary for the Merlin Catalog for Fall 2008 Semester

113

A further investigation of content accessed regarding the visitors would have been made possible if the *site search* feature in analytics was set. The *site search* feature provides the keywords entered by the user using the website's search function. However, the *site search* feature was not set for the Library website.

The findings for research question 2 suggest that metrics reported by Google analytics can provide information on user activities when on the MU library website. The metrics are reported here by the three major sections Google analytics provides – *visitors, traffic sources*, and *content*. Not all features were used to report findings because some of these features were not applied for the MU library Google analytics implementation. A second reason is because metrics and features were not mentioned in the interactive group interview with the MU library web committee.

Chapter 5

Discussion

This study investigated the use of web analytics to understand academic library users' online behavior. Web analytics has been used extensively in commercial settings; however it has not been used in an academic library setting as extensively. The purpose of a library differs from that of a commercial entity in terms of its profit-making motive. Since analytics was developed in the commercial setting, most of its metrics are developed to suit commercial motives. This study sought to understand how analytics metrics and features can be used to interpret users' behavior within the academic library website. The study was informed by the MU library web committee interview and the Google analytics implementation of the MU library for the Fall 2008 semester. To interpret Google analytics reports, the MU library web committee's feedback on the library's web analytics implementation was a key. The interpretation of metrics featured in the Google analytics reports was important in interpreting the MU library users' online behavior.

This chapter presents discussions on major topics derived from findings. The discussions are structured around the two research questions. The first research question about the use of web analytics in library decision-making discusses the MU library web committee's use and impressions of their Google analytics implementation. The second research question about the use of analytics to interpret the website visitors' information behavior discusses the visitors' online information activities through analytics and interprets their information behavior. In the following discussion the MU library is also referred to as "Library," or the "university library." Other academic libraries are referred to as "library." Further, the discussions are supported by review of literature.

RQ1- How can web analytics data inform Library decision-making?

The information environment has posed technology challenges to the academic library with the fast developing web environment. As mentioned earlier in the literature review, academic libraries are urged to develop better and more efficient ways of collecting and analyzing library online traffic. The challenge for the library is not only to respond to the technology challenges, but also to adapt to the changing information behavior practices of new users. The quality of the library services provided needs support amidst the challenges of quick and easy access to information available through the web. The decisions made within the library at this time have an immediate need for support by a user-centric approach that will help keep the libraries in sync with the latest adaptations of information behavior for its users.

The response obtained from the web committee revealed aspects of the MU library decision-making that relate to web analytics implementation of the Library. The themes that emerged from the discussions are situated under the findings section in chapter four. The discussions in this section cover the major issues from the findings with the following headings-

- Decisions about services

- Decisions about web design

- The Librarys' interest in user behavior

- The Librarys' management and function

5.1   Decisions about Services

The MU library web committee would like to ensure proper services to users. The web committee would like to understand how users are behaving when they are on the Library's website. They expect the Library's web analytics implementation to provide a useful solution for understanding the usage of the Library services. Then they can identify any problem in the

Library's online services offered, and therefore take appropriate decisions to provide better services. The topics under which the *decisions about services* that elaborated are:

- Use of online reference help by Library visitors

- Appropriate user machine settings

- Electronic journals subscription

- Users' search habits

- User-specific information needs

5.1.1 Use of online reference help by Library visitors

Providing support to their patrons has been a crucial element for academic libraries. Information on web traffic trends can prepare the Library for high traffic load situations. In a physical library setting, the library provides services to help the patrons with any information need that they might have through the reference desk. A reference librarian usually acts as the pivotal person when the patrons need any kind of information or service from the library.

The advent of easy and quick access to online information has made the online user independent. The information seeker relies on a one-stop search option that provides her a number of links to choose from just by conducting a search. Head (2007) found that students rely on the library's resources for quality information. However, when accessing the library's online resources the users hope to be able to satisfy their information need on their own.

The Library, nevertheless, would like to be prepared with appropriate staffing during uptakes of users access to the website. The MU library provides quality resources that its users can avail themselves of. Getting to these resources might not be as easy as getting to information available with quick searches on the web.

The MU library seems to acknowledge the need for support of its online patrons by staffing the online reference desk appropriately. By looking at the analytics data for online

117

reference usage of the MU library website, the Library could gain a data-based knowledge of how many users actually make use of the online help desk on the Library's website. Information such as if the reference help is situated at appropriate places where visitors need it the most, and if it is easy to locate, can be ascertained. Crowley et al. (2002) pointed out how users expected instant help. The users in their study preferred to see more responsive staff that answered their email queries and they received help when needed.

Therefore, the Library saw analytics making useful contribution to their staffing decisions. In short, the following points came up regarding the Library's decision-making needs:

- The MU library would like to provide adequate online support to its web visitors.

- Though the online users are increasingly getting independent in their search behavior, they access the academic library for quality resources.

- The Library staff would like to know the variations in online traffic in order to be able to appropriately add librarians to their online reference help.

Analytics can provide the Library adequate evidence for supporting their staffing decisions.

5.1.2 Appropriate user machine settings

Recently change and adoption of new web technology is happening fast. While some users of this technology are early adaptors who are quick to upgrade their machines and identify the advantages of these technological enhancements, some others are slower. The slower users may wait for a longer time to adapt to the changes. The Library needs to keep in mind the wide variety of user base that include people from different profiles and demographics.

The smooth operation of online services is reliant on certain aspects of the users' machines. Users' machines need to be well-equipped to receive appropriate services when it comes to online access. Certain issues that can affect the online reception of a service are the incompatibility of the machine and the bandwidth of the network used to access the service. A machine can be incompatible when it does not have the appropriate browser settings such as the

browser being used, version of the browser, appropriate add-ons to run different applications or images available on a page. The network bandwidth further sets the size-limit of the downloadable objects sent over the internet.

The members of the web committee mentioned the importance of knowing the end-users' machine configurations to be able to provide them the appropriate services. In one instance Respondent three mentioned that the Library has used images in the past that were not compatible with visitors' machines that used larger monitors. By implementing analytics they can further enhance their services provided by using the latest available technology and ensuring that the visitors are equipped to receive the new services.

User machine settings can be crucial to provide efficient services to the end-user. Based on the above discussion, the following issues concern the Library decision-making process related to user machine settings:

- Technology changes are creating wide gaps between different users' machine settings.
- The Library has to decide what format of services to provide based on whether the technology used by the visitors supports that format.
- The Library is aware of the importance of the visitors' machine settings and would like to comply with them.

Therefore the Library sees the use of analytics to help it support certain service decisions by providing the technology configurations of users' machines.

5.1.3 Electronic journals subscriptions

Among the various MU library services provided, the Library needs to decide about the eJournals that it needs to subscribe to. The members stated their concern on how, until recently, the Library has continued to subscribe to eJournals that they were already subscribing to. In order

119

to assess journal usage, the Library had to commit time and resources. As mentioned in the literature, *Evidence Based Librarianship* (EBL) has its own drawbacks regarding the time it takes to conduct a study and the lack of appropriate research articles that are applicable to the Library's situation (Nicholson, 2006a). With the Library's online system it can monitor the usage of the visitors and their information seeking activities. The reports can then be used to make decisions regarding to which eJournals they should subscribe.

By subscribing to eJournals libraries have made online access to quality information easier. King and Montgomery (2002) conducted a study to learn if migration to electronic journals had any effect on users' reading, outcomes from reading, or information seeking and reading patterns. They found that most of the library provided readings used by faculty and doctoral students are used from electronic articles. The respondents of their study reported that they spend much less time locating and obtaining library-provided articles when they are available online. Regarding outcomes of the readings, they found that outcomes from reading library-provided articles tend to be more favorable, than readings from other sources.

Since eJournals are external to the Library system, the Library cannot get information on how useful these are to the online users. Also, a thorough report on visitors' eJournal usage is difficult to get from the vendors. Analytics can only provide an estimate of the journal usage by providing the information paths that the users take to get to these external resources.

The Library's decisions regarding eJournal subscriptions can be useful for its online visitors. Consequent to the above discussion the following points emerge:

- The Library has been sticking to previous subscriptions of eJournal without assessing their use.

- The Library's analytics implementation can provide data on the usage of the various resources within the Library website.

- eJournals are external to the system, and analytics cannot capture any usage information.

- Some estimate of eJournal usage can be gotten by looking at the paths the users take that lead them to the external eJournal resources.

Analytics can be restrictive when it comes to interpreting external resources that the library provides.

5.1.4   Users' search habits

The web provides an abundance of information that is available quickly and freely. Information seekers find searching as a way to deal with excess information. The services and resources that the library provides are better in quality as compared to the web as a whole. However, literature suggests that the online user finds it difficult to browse through the library resources (Cockrell and Jayne, 2002; Crowley et al., 2002).

The members of the Library web committee would like to improve the visitors' online search experiences by constantly monitoring the terms that they use to search, and if these searches need to be set as a keyword to get to the resources, then they should do so. The Library's menu-based search options allow the user to do a more controlled search within a specific information item such as a book or article. The Library has found it important to know if the users are pulling up the accurate selections from the menu.

Kim (2006a) looked at the users' characteristics for subscription database use. He found that undergraduate students would prefer open-internet searches as more convenient than searching subscription databases. The major determinant of online use was convenience of access.

However, Head (2007) in her study on 'higher level' students, found that they do not know what quality research resources are and how to locate them on the internet, and as a result

they would try to balance between both online and offline resources and resort to hybrid searching. According to Head, students "[1] accessed convenient, vetted, and aggregated online resources from course readings and the campus library website, [2] to a lesser extent, used internet sites, such as Yahoo!, Google, and Wikipedia, and [3] worked with professors or librarians one-on-one to narrow down searches and clarify expectations for assignments. (p.8)"

The MU library website has different types of searching. They are website search, basic search with a drop down menu, the database search, and the eJournal search. The website search indexes the MU library's website, the basic search works towards helping the users do a more controlled search, and the third search option is for finding a specific article or journal.

Basic searches where users can select the resources they would like to search are often considered useful by libraries for efficient search results (Caswell, 2006). 'Guided searches,' as Caswell calls it, are a step above what was earlier a dropdown menu. According to him, guided searches are search forms in which the user can enter information in one or more boxes and select qualifying parameters to focus the search. Caswell discusses how simple guided searches are, highly usable as opposed to searches with more than two dropdown boxes. An efficient guided search is meant to be tailored to the context of the page, which would not require users to put in more details.

Additionally, it is important to consider the search traits of the online user when looking at effectiveness of different search options. Nicholas et al. (2006) point out a typical characteristic of information behavior of user is their shallow depth of searching. The depth to which a user goes when searching can be an indication of the involvement and engagement of the user. The respondents of the MU library web committee were interested to know about the terms the users were inputting on the Library website.

Analytics has a site search feature that needs to be set up in order to capture the users' terms entered. Usually non-library websites have one search option feature that indexes the

122

website for any information. However, the Library website has a different approach to searching. Librarians see a visitor to the website conducting more than one type of search – searching the website for information regarding the Library services, searching for information in the Library's collections, and searching the information sources such as database and journals to which the Library subscribes. The Library's primary aim is to help users find information, in other words, to help them conduct efficient searches.

Analytics can be used to capture terms entered for one site search option, whereas a library website may have multiple search options under one profile. Users' behavior on the Library website, therefore, cannot be captured for all types of searches that the visitor uses such as Journals, Databases, Library Collection, or Website Search. If a visitor uses the search option to look for appropriate database, and the site search feature of analytics is set for the website search, then analytics cannot capture information on the databases that the visitors looked for. The Library is still going to miss out a lot of information about the visitors search behavior.

In summary, most of the online users' information-seeking habits on the web start with searching. The Library has offered multiple search options for their visitors. The discussion above reveals the following:

- The members of the web committee would like to know the visitors' information search habits.

- The committee sees the terms that the visitors use to access information on the Library's website as significant to improve the visitors' information searching process.

- Analytics provides a site search feature that enables the Library to capture visitor information on the website which is not set for the MU library.

- The Library has multiple search options, but analytics can only capture information on one search feature of the website.

For decisions on search the MU library website cannot entirely rely on their web analytics implementation unless they would like to create multiple profiles.

5.1.5   User-specific information needs

The MU library's user base consists of different user profiles, including students, faculty, and staff from all ranges of experiences, motivation, and information needs. The members of the web committee expressed the need to know which visitors are students, faculty, and staff. Knowing the visitor profiles can help the Library provide targeted services that will improve their information seeking experience when on the Library website.

However, the profile of the Library user can become unclear when multiple users access from the same computer when collaborating. In other cases when multiple users access through adjacent computers, their information seeking is influenced by the group work. Also, visitors who access from a campus come from different computers, either on-campus or off-campus. As a result, it might be difficult to separate out the visitor profiles on the Library website. Though analytics does not provide the IP address for each visitor, the reports are IP-based. When multiple users access through the same computer, the different behaviors through the same IP might be interpreted as one. At other times the same user might access the Library website from different machines that might be interpreted as different visitor access and might affect *absolute unique visitors*.

Absolute unique visitors provide information on the number of visitors who access the Library website not taking into account the repeat visits they make to come to the website. When the visitors use different machines to access the website, a different IP is recorded for the same visitor. Such information might bias interpretations when applied to the Library's user-base.

Analytics provides traffic source options that indicate where the visitors accessed the Library website from directly, through referring sites or by conducting a search. However without profile information, user-specific information about access behavior cannot be made. In order to

124

understand the specific profile of the visitor, analytics provides an option called user-defined that

the Library could benefit from. The *user-defined* feature needs to be set up in order to be

operative. Users' access from multiple computers, or multiple users through same computer, can

still inhibit the user's profile.

Identifying the visitor-specific behavior has been difficult when it comes to online

applications. Thelwall et al. (2005) discuss how issues of interpretation using log files may arise

when different people access the same computer, when the same person accesses different

computers from work, home, or even when multiple people access the same browser. Tracking by

IP address may create problems in determining appropriate user actions on the web.

Other studies have found how the different user profiles have different information-

seeking habits. Cockerell and Jayne (2002) found that less experienced users such as

undergraduates, who would give up too soon are less persistent with the information tasks in

hand. Their lack of search knowledge affected their information seeking. They also found that

student's search activities depended on their knowledge of the information source such as

distinction between a journal and a magazine title. Griffiths and Brophy (2005) pointed out that

students prefer a simpler navigation structure, and their low use of academic resources makes it

difficult for them to locate information. Such varying needs of the visitors are an important

consideration that the Library needs to make while providing targeted services to its users.

Therefore, getting the visitors' profile information could prove to be valuable when

interpreting visitors' information. The following summarizes the points related to the Library's

visitor specific information needs:

- The Library has multiple groups of visitors who access their website.

- It is difficult to separate out the profiles of the visitors from the data captured through

   analytics.

- The same visitor or a group of visitors accessing from different machines can affect the interpretation of reports provided by analytics.

- The user-defined feature in analytics can be set to capture user specific profile information.

Any information on a user-specific profile can be very useful for interpreting users' behavior, but it cannot be perfectly attained.

5.2 Decisions about Web Design

The MU library would like to assess and make improvements in their web architecture. The willingness of the Library to improve their services has prompted them to look at some of the analytics figures and make changes on the website, such as moving links, replacing images and providing more usable website. The discussions in this section expand on how analytics has been accepted for use by the Library to make changes to its website design, and how navigation on the Library website poses challenges in interpreting visitors' information behavior. The two topics that will be discussed here are:

- Use of analytics to make changes on the Library website

- Difficulties in understanding visitors' navigation behavior

5.2.1 Use of analytics to make changes on the Library website

The Library's analytics implementation has been in place since March 2007. Though the implementation was done acknowledging the usefulness of capturing online usage data, not many data were used except for applying them to move some links that seemed redundant. The MU library web committee hopes that it could use the analytics implementation more. However, it seems to have concerns about solely relying on analytics for any usage information.

The respondents expect more details in the reports. Respondent three mentioned that "Seeing one page at a time is useful, but seeing one thing after the other is useful some times."

Members would be more comfortable inferring from a detailed report than pieces of disjointed figures. The Library committee is used to a more context-based approach and expects more continuity among the visitors' actions. The Library has tried to use navigation summaries before but they are confusing to them such as when the previous and the subsequent pages report where a user comes from and goes to are the same URL.

A single implementation for all the branch libraries embeds a lot of valuable information in the same report. In order to get information on the usage of URLs, the particular URL needs to be searched from a list of other URLs that analytics is capturing. Having to search for a specific URL further complicates analysis.

Nevertheless, the respondents acknowledged how analytics could be useful to compare the usage of resources. For example, the use of catalog compared to online database over time could show how visitors' information seeking habit is changing if they are shifting to database use from catalog use. The Library also sees analytics as useful in increasing visibility of some links and sections. Respondent six mentions noticing the *site overlay* feature of analytics that the 'Announcement' section on the home page is not getting as many hits, and so they might decide to make the section scroll.

Thus, analytics provides the Library a way within their infrastructure to capture online information on a continuous basis that can be used for any design related decisions. Hiller and Self (2004) discuss the importance to libraries of having their own data that can be collected and used for decision-making. They mentioned the importance of outcome assessment that determines the impact on library services brought out by service related changes. They stressed the importance of building a robust infrastructure to develop and sustain data-based decision-making. However, Hiller and Self also caution concerning total dependence on figures that might eliminate the context and any other meaningful interpretation.

127

Nicholson (2003) also states the importance of the library using advanced technology that can provide useful insights into the library's electronic data. Collecting aggregate data, though, maintains privacy of the library's online user; however, it restricts testing statistical assumptions. He calls for libraries to plan ahead for the new web technologies with which the web is inundated. Being prepared for new technologies will help the Library suit their services to the developing behaviors of their online users on the website.

The discussion above points to the Library's use of analytics for web design related changes. The following can be surmised:

- The Library has used analytics for making minor changes to the links in the website.

- The Library committee would like to see more contexts in the analytics reports.

- Sometimes there is a lot of data to be interpreted that staff do not know how to interpret.

- The Library committee has learned that analytics will be useful to compare how visitors adopt one service over another over time.

There has been a mixed reaction to the use of analytics for the Library.

5.2.2 Difficulties in understanding visitors' navigation behavior

The Library provides most of the information through categories and links. The Library website is a gateway to a number of resources. In order to ensure that the Library users get to the resources and services provided by the Library, the Library needs to make certain that the visitors are able to navigate the intended resources.

However, the Library has accumulated many redundant links over time that might complicate interpretation of the users' behavior when they use these links. Though multiple links can provide multiple paths of access, these links need to be present at appropriate places. Respondent three mentioned that most of the users get lost because of the presence of multiple links and eventually found themselves on the same page on which they started. Some other

resources have multiple URLs. The Merlin catalog is denoted by

http://mulibraries.missouri.edu/outgoing/merlinmu_gateway and also by the link

http://laurel.lso.missouri.edu/search~S1/. Respondent five was worried about interpretation of the

Library's resources through multiple links.

The MU library has used analytics to understand visitors' navigation behavior. In order to

differentiate between different links that the users can use to go to the same destination pages, the

Library has tagged some links. Further, there was interest in understanding how visitors get to a

page, whether by bookmarking, or through a referring site. The committee was also interested in

knowing if users recall the paths they visited before.

However, interpretation of visitors' behavior can be confounded by issues related to the

website design. Wang et al. (2000) used a process-tracing method to see what happens when a

user accesses the library website. They found that many of the website affordances complicated

interpretation of information needs of users. According to them, heterogeneous objects, poor

interfaces, and diverse web organization were detrimental to understanding the user's information

need.

Sometimes interpretation of the user's behavior is related to how the user understands the

website design. Users sometimes might get confused by the website design and this can further

complicate interpretation. Crowley et al. (2002) also pointed out some website design issues that

disrupted users' expectations. Users preferred more efficient online access pages that have

workable, faster links. They also hoped to see more visible catalog and 'What's New' section.

The inconsistency of the website and lack of understanding of the terminology made it more

complicated to use.

The Library website navigation is important for efficient online access. Consequently, the

following can be said from the discussions above:

- The Library has multiple links that lead to the same resource.

- Online visitors to the Library website get confused due to the presence of multiple links.

- The Library is aware of the issues with the website and is interested in how the online visitors navigate through the website.

- The Library has used analytics to understand users' activities by setting up tags that help in differentiating two identical destination links on the same page.

Website affordances can be a prime factor that can obstruct any interpretations on online visitor information behavior.

5.3   *The Library's Interest in User Behavior*

The Library's primary intention in implementing analytics was to use it for improving the visitor's experience. However, the interview revealed the respondents' curiosity regarding certain user behavior aspects. The Library provides a variety of quality services to its mixed profile of users through a variety of distributions of menus, categories, and links made available through multiple outlets that include the branch libraries. Looking at the vastness of the Library's operation and the large user-base it provides for, it is a large service enterprise.

The respondents from the web committee expressed their interest in knowing certain aspects of the users' behavior that are further elaborated under the following topics:

- Users' activities on the branch Library website

- Relative usage behavior patterns of visitors

- Visitors' approach towards the information available through the Library website

5.3.1   Users' activities on the branch Library website

The Library provides its services through the main Library website as well as through different branches. The branches are established from an organizational perspective, depending on the different disciplines the branch is affiliated to such as Law, Health Science, Veterinary

Sciences and so on. It becomes easier to assess the significance of the different branches with the online interfaces of the branches in place. Analytics has made it possible to look at information access through the branches of the Library. The Library's branches have their separate teams that manage the website. Looking at the various branch-wise usages can provide the Library useful insights about interactivity between the main Library and the branches of the Library through the online interfaces.

The users accessing the Library website from different branch websites do not mind where they are led as long as they can access the resources they want. A respondent from the MU library's branch expressed her desire to know users' activities on her specific branch website. Information on how these users are coming to the MU library's branch gateway, what keywords they are using, and if they are using the resources that are specific to the branch Library are of interest to the respondents. Additionally, the respondent from the main Library was curious to know what branch resources the visitors are using.

However, finding branch specific usage data through an analytics implementation under one profile can be a challenge. Some of the branches of the Library have their own separate analytics implementations that enable them to collect their own data under a different profile, while others branches do not have a separate profile. Having a separate profile will enable the branches to explore users who visit the respective branches activities.

Visitors' access behaviors for branch Library websites can vary based on aspects such as their location, experience, knowledge of the branch Library URL and so on. Bracke (2004) used web usage data to interpret users' activity on a Health Library website and determined that a large part of the user base who visit the branch website from off campus come through the main library website rather than the ones who visit from on-campus. This was probably because of the greater familiarity of on-campus users with the library than the off-campus users'. Also, the off-campus users, as compared to on-campus users, used external tools to access the library website. Bracke

also found major differences in database use among other resources of the branch library which he attributed to direct links to the branch database resources from the main library page.

The respondents of the MU library web committee are interested in users' activities on their website. The above discussion that relates to usage of branch resources can be summarized as follows:

- A separate web analytics implementation for the branch Library will enable the branch Library to collect their individual separate data under a different profile.

- The Library's branch management is curious to know about visitors' access behavior of the branch Library website.

- There is implied curiosity for the branch Library to understand how visitors see the branch as a part of the main Library.

- The Library would also like to know if users use the branch resources.

A comparison of users' information related activities across the branches can provide interesting insights into its relation to the Library.

5.3.2 Relative usage behavior patterns of visitors

Analytics provides various features and options to view data captured visually. Interpreting log files sometimes requires quantitative and data mining skills that might be a constraint for the Library in terms of time and resources available. The visual presentation of analytics data provided the Library a quick way to notice trends and variations that helps to explore interesting user behavior elements.

Further, the comparison feature available in analytics allows analysis among elements such as two time periods, two metrics, or comparison to the site. Being able to look at the comparisons allows a good way to discover interesting aspects using a frame of reference. Figure 1 shows a comparison of database use trends to the website. The comparison surprised the

members of the web committee as they saw how database trends were similar to website use trends, indicating that the database use forms a considerable part of the website use.

One other thing the respondents expressed their interest in regarded the Library visitors' shift in usage from one resource to another. The dynamism of the information environment has lead to quick changes in the users' adoption to different technologies. The respondents were curious about such changes. An example of a change that the Library would like to note is the shift in use of Library services. The Library would like to be able to compare if the increase in use of one resource such as the database reduces the use of another resource such as the catalog. Such information on relative usage raises their curiosity.

The respondents could see that analytics provided them interesting insights about the Library website. To summarize:

- The visual presentation of analytics has enabled the Library to discover patterns.
- The comparison reports raised the curiosity of the respondents when they saw interesting patterns of database use compared with those of the overall website.
- The Library web committee saw the potential applications of the comparison reports provided by analytics to explore changes that advanced technology can bring to users' information behavior.

Analytics has enabled the Library to explore and discover their users more than before.

5.3.3 Visitors' approach towards the information available through the Library website

The respondents showed specific interest in knowing users' approach to the information obtained online through the Library website. They would like to be aware of not only what actions the users do, but also why they do it when they visit the Library website. It is understood that the visitors have an information need when they visit the website; however what are they

actually doing when they are trying to satisfy the need is intriguing for them. An understanding of the visitors' attitude to the Library's portal interests them.

The theories of information behavior could provide academic libraries a picture of the users behavior, however Nicholas et al. (2006) state that the theories on information behavior have been mostly developed based on the professional users' behavior on applications such as Dialog, Nexis, DataStar, and Reuters. These theories hardly hold any relevance for what they call the "new digital information consumer." For a library, bypassing the middle person in accessing information has opened the end-user to a wide array of information resources (Nicholas et al., 2006). The end-users' behavior in dealing with the online resources of the Library, hence, invigorates the inquiry of the Library.

Though analytics would not provide insights to the users mind, access to users' actions can provide insights to the different approaches the user takes when visiting the Library website. Further, the discussion above can be summarized as follows:

- The Library is interested to know what do the visitors think when they access the website, and how does that translate to their actions.

- There are theories that have been developed in information behavior; however these were developed on professional users' behavior only.

- The Library is curious regarding how users deal with the online web system without the help of a librarian.

Google analytics can go till the extent of reporting what users' do. Librarians can use this information to look at interesting aspects of users' online behavior.

5.4   *The Library's Management and Function*

"The Library's management and function" are important considerations for analytics implementation in a library setting. An analytics implementation cannot be completely successful unless the initiative is from the Library's top level management. Certain other elements that

134

inadvertently influence implementation of analytics and interpretation of the data are the functional aspects of the Library. The discussion with the web committee members provided insights to how the Library specific elements can influence adoption of analytics. These elements provide a bigger picture of how analytics can blend into the Library's work culture. Further, the topics are as follows:

- Use of new technology for Library purpose

- Reservations about the appropriateness of analytics for academic libraries

- Issues with measuring usage in the Library setting

- Evidence based decision-making in Library practice

5.4.1 Use of new technology for Library purpose

The Library has always safeguarded its patrons' privacy. The notion of privacy has been held foremost in the MU library. In an online library setting, however, intricate details of visitors' actions are recorded. The idea of storing the visitors' actions at every step cautions the Library about the privacy of such data. Though analytics provides aggregate data that do not have users' information at the Internet Protocol (IP) level, the respondents expressed concern about whether unauthorized agents can get hold of the user data and violate privacy issues.

Analytics has been used by the commercial sector extensively. The goal of the commercial sector has been to maximize their revenue. Having to adapt to analytics that has been time tested in the commercial sector only is of concern. Besides, Google has always promoted its analytics solution for commercial purposes targeted to commercial entities. The Library, therefore, is not comfortable with accepting analytics for its purpose.

However, the Library needs to realize the benefits of analytics for its purpose. Nicholson (2003) urges libraries to adapt to new technologies that enable creating a data warehouse. Though he would like to see data at a non aggregate level that would allow him to test for statistical

assumptions, he acknowledged that aggregate data would help libraries to get over privacy concerns.

Therefore, the Library needs to gain a better understanding of the new technology tools and suit it to use for their purpose. The following can be said from the discussion above:

- The Library is cautious about the privacy of their users' online data.

- The use of analytics in the commercial sector has made the Library more skeptical about its use for their purposes.

- Libraries can benefit from technology such as analytics that is readily available for use by suiting it for their purposes.

Using new technology for the Library purpose will benefit the Library in the long run.

5.4.2   Reservations about the appropriateness of analytics for academic libraries

As mentioned in the previous section, the Library seemed to be averse to adopting analytics for its use. Analytics has been used more for commercial websites for keeping the visitors on their website as long as possible. The Library's primary purpose is to get the user to the different resources they are seeking and not keep them to stay with their website for long.

Due to this basic difference in ideologies between the library and the commercial sector, the Library has certain reservations about using analytics for interpreting the data captured. Lakos (2007) called for libraries to adopt ways and means to improve their decision-making capabilities suiting to the "information economy" by adopting new products that are available in the market.

Studies such as this can help libraries identify how the metrics provided by web analytics can be customized for use by the academic library in terms of achieving library goals. In doing so, there needs to be a clear vision of the goals of the library, how it operates, and its differences with the commercial sector. Based on the fundamentals of the library setting, a guideline of use of analytics for achieving the library goals could be provided. Further use of analytics for academic

libraries can build this guideline to a prescribed list that can be used across the library community.

Consequently, the Library needs to realize the potential of using analytics. The discussion above can be summarized as follow:

- Analytics has been applied differently in the past to different contexts.
- A guideline for the use of analytics should be developed for the purposes of academic libraries.

The Library can garner many benefits of analytics for their purpose.

5.4.3 Issues with measuring usage in the Library setting

The Library setting has functional characteristics that make it different than other non-library entities. There are numerous users who access the Library website from many locations using many computers. There are users who are internal staff who also access the Library's website. Also, the Library provides access to its resources through on-campus machines, using the university's wireless access. These characteristics tie the Library together under one broad umbrella of the university. The providers of the Library services are employees of the Library, and the receivers are employees and enrolled students of the university.

Having a well-defined user base that can be differentiated under the broad user groups of staff or student is valuable. Analytics provides filtering the data that can enable looking at groups of users differentiated by the Internet Protocol (IP) address of the machines or by the domain name they access from, or the subdirectories they access to. However, filters cannot be applied post-data collection; they can be good for applying to data not yet collected. Additionally, multiple reports that require filtering different IPs cannot be applied to one profile. For example, to get two reports, one that filters out only the Library's staff machine IPs, and another report that includes only a particular range of IPs, cannot be obtained on the same profile.

Apart from that, academic users' behavior can have aspects such as collaborative information behavior on same or multiple computers, faculty and librarian influenced behavior of students, information satisfying temperament of undergraduate level information seekers that bring multiple variables in the information behavior of academic users. The Library feels using analytics in the Library might pose a challenge due to the varied user profiles and motivations.

Though analytics provides various metrics that can help determine the user's actions such as 'Visits', 'Absolute unique visits', 'Pageviews', 'Time on site', or 'Bounce Rate', these reports still have chances of being misinterpreted due to instances when a user leaves the computer without closing out of the site. Most of the interpretations of users behavior will be based on assumptions on analytics metrics, but what is going on in the user's mind still cannot be determined.

Also, interpreting users' actions when they are doing different things on the same page is difficult. Looking at analytics figures does not illuminate the users' exact actions on the page. Study findings suggested librarians would like to know their users' exact actions and would like to see a continuity of user actions in the charts provided by analytics. Huntington et al. (2008) describe the importance of differentiating activities of a user on a page who is viewing an item, say the abstract, another user on the same page reading the article and printing it, and yet a third user who might have viewed the article and decided not to proceed with reading it. Nicholas et al. (2003) acknowledged in their study that view time on a page might be anything from the abstract view time to the documented scanning time.

Librarians need to conduct supplementary studies to account for the context of user action. For example, as mentioned earlier, Penniman and Dominick (1980) talk about the need to consider the environment where the evaluation takes place. They specify a detailed list of relevant parameters for evaluation and talk about a minimal required set of parameters that are necessary for each evaluation consideration. The data measures that they found necessary were system

138

usage profile and database usage profile, user error and error recovery, and user success and user satisfaction. The parameters of each measure were data obtained through monitoring both the system and the user's interaction with the system, as well as direct measures of user rating and comments about the system.

Metrics that are used to describe user activities need interpretation that keeps in mind the various contexts of users' actions. The MU library has many pages that are spread across various parts of the website and there are multiple services that the Library provides across its many branches. Interpreting user activity through assigning meaning to the metrics becomes very complicated as the metrics would need to be interpreted differently for different web pages of the Library.

Getting to the discrete level of users' actions need to be interpreted to understand the users' motives. Nicholas et al. (2003) considered the actual downloads as units of analysis as that would mean the user has actually viewed or downloaded the article and she is really interested or motivated to make use of it. According to them, "…The [website] hits encompass all kinds of activity – arrivals by accident, people navigating through the site, viewing content, downloading a page, using a page as office 'wallpaper' " (p.89). Further, Davis and Solla (2003) identified a relationship between the number of downloads to the size of the user population. In their study they used logs of server data to develop a quadratic relationship between the number of journals consulted and number of articles downloaded.

The Library is a part of the larger university. Interpreting the Library's website activities with analytics alone can be complicated. To summarize:

- The Library has different category of users that fall broadly under the two categories, staff or student.

- Analytics filtering options cannot provide the best solution to segregate the visitors by their IP under one profile.

139

- Visitors' collaborative practices might complicate their users' behavior interpretation.

- Visitors' habits of leaving in the middle of a session might provide misinformation about the visit metrics.

- Capturing users' different actions on the same page is difficult.

It is important to conduct supplemental studies in order to get the context of the Library users' activities.

5.4.4 Evidence based decision-making in Library practice

Academic libraries have relied on *Evidence Based Librarianship* in decision-making. Most of the evidence in EBL has been obtained from research literature on libraries. However, Nicholson (2006a) points out the lack of research articles. Also, Eldredge (2006) notes the time taken to collect evidence reduce the power of EBL. Web analytics can be seen as a possible solution to provide readily available data at any time. Decisions can be based on figures reported on the use of database or eJournal links that were clicked, what machine settings the visitors to the Library website use, the times the visitors use the Library website and so on.

Some of the evidence that the Library has to rely on are provided by external systems. The external systems are the vendors to whose services the Library subscribes. The Library acts as a medium to provide information from the resources supplied by the vendors. The vendors provide the Library regular statistics on usage of their resources. There is no way that the Library can access these external systems to generate their own reports.

At other times, the Library management may overlook suggestions from evidence. Evidence on the usage of some links might not inform as to political reasons why the Library is going to keep them. Other evidence of visitors' use of external links to visit the Library website is also going to be disregarded. Respondent three emphasized that getting in touch with the external *referring site* for marketing the Library website might be detrimental for the Library as these

external sites might get promotional ideas when they get to know how they are diverting users from their site to the Library website.

Analytics can provide required strength to Library service evaluation and decision-making keeping up with the EBL approach. Additionally, it can address the lack of research articles and application of evidence to the Library context. Also, issues of time and dated evidence can be addressed by providing "on the go" figures from visitors' actions that can prove to be useful to the current context. However, certain limitations of analytics implementation are in its applicability to external proprietary systems and certain other marketing policies of the Library.

Therefore, analytics can provide libraries with a solution to address some of the drawbacks of EBL. The points discussed bring out the following:

- EBL has certain limitations when it comes to providing updated, contextual, and applicable evidence to support decision-making.

- The Library has to rely on evidence provided by vendors as they cannot generate their own reports.

- Some of the evidence that analytics provides might be disregarded by the Library for political reasons.

As beneficial analytics might be, there could be political restrictions to its use.

RQ2- How can web analytics data be used to interpret the information behavior of the website visitor?

Metrics reported by Google analytics provides information on user activities when they are on the MU library website. The discussion in this section details how the different features available through Google analytics could be used to interpret visitors' information behavior in a broader context. Google analytics provides website usage data presented in terms of the various metrics of user activity. The discussion attempts to connect the data interpretations of these

discrete elements of users' activity to suggest how the visitor behavior might look. Any indication to information behavior has not been intended in the richer context as is normally done in *Human Information Behavior* literature. The analytics metrics that have been used to interpret users' activities on the Library website cannot provide the richness of the users' context. The Library is an information service provider, and so the users' visits on the Library website are assumed to be for information access and seeking. The terms "information behavior," "information seeking," "information search," and "information access" are used in this section to further connect the intention of the user to the activities performed, as depicted in the analytics reports.

The discussion follows the same structure as the findings section for the research question 2 and it is reported under the three major sections in Google analytics– *Visitors, Traffic Sources,* and *Content*.

5.5   *Visitors*

The *visitors* section in analytics provides information on user visits to the website. The findings provided in chapter four elaborated the metrics for *visitors* under the three major categories of *visitor trending, browser capabilities,* and *network properties*. The discussion in this chapter looks at the information behavior interpretations of the visitors from the metrics reported under the *visitor trending* section such as *visits, pageviews, average pageviews, time on site, and bounce rates*; *browser capabilities* section that includes *operating systems (OS), browsers, browsers and OS, screen colors, screen resolutions, flash versions,* and *java support;* and the *network properties* section that includes *network location*. The various metrics under the three different sections mentioned above can provide a view of the visitors' activities at a basic level. Though the figures reported are aggregate, the metrics reported inform the discrete acts that the visitors performed. The aggregate data reported through analytics can be used to generalize the metric reported.

The discussion under *visitors* is grouped under:

- Learning about visual representations of behavior through visitor trending

- Understanding users' technology readiness through their browser settings

- Using network properties to understand the visitors' information

5.5.1   Learning about visual representations of behavior through visitor trending

The *visitor trending* report provides trend reports about visitors under one section. Looking at graphs allows interpreting users' behavior on the website across a time period, for a particular metric being investigated. The rise and fall in the graph for a metric would allow understanding the increase or decrease in the metric reported. Figure 2 provides the graphs for user visit trends on the Library website. The data reported suggest that the visits fell for the third, seventh, and fourteenth weeks. Users' information behavior seems to be affected by the different time of the semester.

The major findings indicated that:

- Visitors' information access tend to increase for the weekdays and falls over the weekend consistently.

- Though the visits to the MU library website rose in the first week of classes for the Fall 2008 semester, they fell immediately the week after.

- In the third week of the semester visits fell for no apparent reason.

- Visits also fell for the fourteenth week of the Fall 2008 semester, which can be attributed to the Thanksgiving recess.

- There is a fall in visitors' seeking information in the seventh week of the semester without any apparent breaks or vacations.

- Due to Thanksgiving recess, visits fell down for a week.

There is a possibility that visitors tend to take breaks in their academic information seeking behavior at intervals. With the start of the Fall session there is a sudden increase in academic information seeking for reasons of curiosity or class requirements such as accessing course reserves. However, once the visitors are aware of the information available to them for use through the Library, it decreases for a week.

Once the Fall session was in progress the information access trends stabilized to an increase over weekdays and decrease over weekends. The decrease, however, during the middle of the semester around the seventh week could be due to the mid-term pressures or because the visitors' academic information behavior might have slacked seven weeks into the semester.

Figure 6 provides total visit trends over a day by hour. To summarize the findings of the graph:

- Visitor access increases over a period of the day and falls during night time.

- There are visits on the Library website 24 hours a day.

- Some visitors access late at night while others access during early mornings.

- There is a time during the day between 9 AM to 6 PM when the visits to the MU library resources are the greatest in number.

Though it is not clear what the specific user activities are during the whole time, it is clear from the graphs that there is an information need for which the user accesses the Library website during all times of the day. However, it cannot be ruled out that some of the visits might be because users leave the website open on the machines at night. Further, the information users of the MU library website increase during the day time.

Figure 3 and Figure 4 provide the graph for pageviews and time on site trends, respectively. According to the graphs a noticeable characteristic is the steady average page views and time on site that the MU library website gets throughout the Fall 2008 semester, in spite of

the variations in visitors' use of the Library website. The time on site is steady at 3:02 minutes with an increase to 4:53 at just one example.

Visits might differ from time to time across a semester, but the average page views for each day and the time spent on the MU library website remains steady across the Fall 2008 semester. The one rise in time spent on site maybe because of a system glitch or in case some user had left the website open for a long time, forgetting to close out of the session. An interesting aspect is that such a rise in time happened just once during the whole semester.

Therefore, it can be said that visitors' satisfaction of the Library's resources is not a reason for the variation of user visits across the period. In other words, the information services provided by the Library cannot be held responsible for a dip in usage from time to time. Decreased users' academic information seeking activity can be attributed to the weekends or a time during the semester when there are no classes. Any other decrease in information seeking activity may need more investigation through future studies.

An important finding regarding the bounce rate was further found. Figure 5 provides the bounce rate graph. The graph indicates a high bounce rate on the MU library website across the Fall 2008 semester. The bounce rates on the website go higher than 30% for most of the days.

As indicated in the findings in chapter four, the MU library wants the visitors to go to the resources that are provided to them by the Library, even if that means that the visitors have to leave the Library website. Bounce rates indicate that visitors don't stick to the website. In terms of users' information behavior on a Library website, the high bounce rate indicates that the visitors come to a Library page to divert to an external resource. For them the Library website is a medium for information seeking.

The visual aspects of the graphs mentioned above provide interesting information on any variation in visitor trends. Chi (2002) points out the aspects of web usage analytics that enable web analysts to look at a visual presentation of data. He commends the work of web analysts in

145

realizing how users' behavior patterns are more important than understanding web structure. With increasing complexity of the web, though, the ability to generate appropriate visuals has become a challenge. Further he stresses how direct visualization of users' session can help developers identify specific bottlenecks.

Lin et al. (1991) also affirm the importance of graphical representation of users' search data. In their study the authors tried to use different scientific techniques to juxtapose users' search states, the data collected such as keystrokes, mapping the data collected to search states, and finally displaying sequences of search states. The authors acknowledge that graphic representations have "provided us with insights into the information seeking process and served as points of departure for analyses of verbal reports and other data" (p. 477).A presentation of the graphical reports of the trends for the metrics enables a quick interpretation of the users' information activity.

Consequently, the following can be summarized about the visitor trend reports:

- The visual reports provided by analytics present a quick and clear way to understand users' information seeking behavior.

- The visual reports indicate that visitors use the Library website on a regular basis with regular uptakes and dips in usage.

- Such regular rises and falls indicate typical academic information-seeking habits for the Library users who visit the website for apparent needs.

- Visits could fall during the middle of the term due to mid-term assessments or fatigue that students tend to suffer during mid-semester.

- Though the visits rise and fall during the semester, the average use of the Library website remain constant, indicating a disconnect between number of visits and use of the website.

- A high bounce rate across the semester indicates that the Library website is a medium for accessing external resources.

The visual reports provide a quick way to interpret information trends of the Library website as indicated by visitor access.

5.5.2 Understanding users' technology readiness through their browser settings

Browser capabilities provide information on the browser characteristics of visitors' machines that they use to access the Library website. Figure 9 and Figure 10 provide information on the browsers that visitors use. The figures indicate the following:

- The majority of visitors who accessed the Library website during the Fall 2008 semester used Internet Explorer (IE), and a lesser yet considerable percentage of visitors used Firefox.

- Out of the visitors who use IE, almost 80% of them were using the latest IE version of 7.0.

Visitors' browser capabilities indicate the level of adoption of technology by the user. In the Fall 2008 semester, the visitors were avid users of IE, which has been in the market for a longer time than Firefox. The visitors seemed to be more comfortable with the familiar IE. Avid users of IE still use the upgraded 7.0 version. It is not true that familiarity with older versions is a reason for the users' closed mindedness to other non-IE browsers. The MU Library visitors are still open to upgrading to newer technology available in the market. There is a possibility, though, that the visitors would like to stick to the reputation of Microsoft as the browser provider for IE. The visitors' adoption of new technologies would suggest how willing the visitors are to adapt their information behavior to the requirements of the new technology. Conversely, better adaptability of new technology could also mean that the technology has successfully been able to address a crucial aspect of visitors' information behavior.

The visitors' access from public computers on campus that use the latest IE version could be a possible reason for a major IE use. Access through public PCs does not allow visitors control over the browsers they would like to use. The visitor, if given a choice, would rather use an earlier version that they are more comfortable using. In some cases, the visitors are being forced to adopt latest technology.

Figure 11, Figure 12, and Figure 13 provide visitors' OS, OS versions for Windows, and the Browser and OS combination use. Looking at the graphs it can be said that:

- A considerable majority of visitors use the Windows OS.

- For those who use Windows OS, more than 80% use XP.

- Most of the visitors who used IE accessed from the Windows machine.

Information on users' browser and machine preferences provide a picture of the users and their technology profiles. It seems a majority of users would prefer to be part of the mainstream Windows and IE users. However, they still are upgrading in the respective realms of the Windows OS and the IE Browser.

Figure 14 provides information on the visitors' machine screen resolution. Also information on other settings on the screen colors, resolutions, Java support, and flash versions as mentioned in chapter four indicate that:

- A majority of visitors of more than 70% had 32-bit color settings

- Thirty-five percent of visitors had highest screen resolution of 1680 X 1050.

- Sixty percent had the latest flash version.

- Almost all the visitors had Java enabled in their machines.

The figures prove the visitors' ability to receive higher quality images that are supported by Java and are accepted by the latest flash versions. It might not mean that the visitors are ready to move from the technology with which they have been accustomed for some period. The visitors who access from public machines, though, are restricted to the settings of the machines.

User machine settings can provide information on what formats of information the users can access. The above discussions can be summarized as follows:

- Visitors seem to be adapting to the latest technologies as they have the latest browser settings on the machines they use.

- It is quite possible that using some public machines might not allow the visitors to use their preferred settings for the browsers.

- It is also apparent that visitors would like to use technology that has been in the marketplace for a longer time.

- Information on visitors' adaptability of new technology can provide useful insights onto how successful the technology has been to address important deficits to visitors' information seeking behavior.

Information through analytics can thus allow interpretation of the technology readiness of users.

5.5.3 Using network properties to understand the visitors' information

Network location provides information on the internet service providers that the visitors use to access the Library website. Knowing what internet service provider the visitors use provides information about if the visits are from on-campus or off-campus. Figure 15 shows the network location of the MU library visitors. Seventy-percent of the visitors that access the Library come from the network location "University of Missouri-Columbia." Visitors who access using the university's internet service are mostly from on-campus. Therefore most of the visits to the Library website are made on-campus.

Academic information seeking of visitors is more extensive from the campus probably because academic users would like to use the Library from the campus machines. An important aspect of the visitors is their profile information that still remains blurry. Though a majority of on-campus users are students, there are still others who are faculty and staff. Even within students

149

there is a large variation of demographics. Though the network location can provide an indication of the visits that were made from on-campus, it does not say much about the visitors' profiles.

The MU campus has public stations that can be accessed through computer labs, or from within the Library. Also, multiple users could access from the same computer, or the same person can access from different computers at different times. There could be groups of users who are working on the same computer, thus providing multiple profile data through the same IP.

Profile information can provide a better understanding of the differences in visitor's behavior. Griffiths and Brophy (2005) concluded from their study that students may vary from faculty in their information seeking. Students are confused about the quality of academic resources, and would prefer simpler navigation structure. The authors also mention the need for students to be "information satisfied," that is getting as much information as is necessary for the purpose.

Though network location of the visitor does not provide enough information on the visitor's profile, Nicholas and Huntington (2003) suggest using a micro approach which involves looking at parts or segments of the log data. This can provide useful information about user specific activity. According to them looking at academic users from a subgroup of users is still useful as compared to looking at a vast spread of users across a geographic location.

The network location feature of analytics can provide limited information regarding visitors' information behavior. This information can be summarized as follows:

- A majority of visitors access the Library website when they are on campus.
- The visitor profiles of the users can be anyone from student to faculty who has different information needs and access habits.

Looking at subgroups of academic users can be useful as compared to looking at a vast spread of varied users across the globe.

5.6   *Traffic Sources*

The *traffic sources* section provides information on where the visitors are coming from on the web in terms of the access paths they take to the website. The findings section provides data under the three sections of *all traffic sources, direct traffic and referring sites,* and *search engines and keywords*. The visitors' use of the different access paths can provide information on the visitor information access behavior online. The discussion here further attempts to understand the Library website visitor by looking at their information access preferences, and their subsequent information seeking activities through the Library website.

The discussions in this section are categorized under the following topics:

- Understanding visitors' information-seeking strategies by their information access preferences

- Learning about visitors information seeking by the sources they use to access – direct or through the referring site

- Learning about visitors' search intentions using their search activities

5.6.1   Understanding visitors information seeking strategy by their information access preferences

The academic Library is a part of a larger institution, the university system. Most of the visits that it gets are from users who are authorized to access the system. Even in situations where the users have access to the Library there are different paths they could take to the Library resources. Getting to know what paths of access the user prefers to take from available options can help in an understanding of the users' access behaviors. Providing traffic distribution across all traffic sources enables a cross-section of the visitor access paths to the Library website.

Figure 16 provides all traffic sources to the Library website. The figure indicates that:

- A high rate of 60% of the Library traffic accessed the Library website directly.

- The second most visits of 18% came from the referring site- http://www.missouri.edu/referral.

- Another 12 % of visitors used Google to access the Library website.

Visitors would often directly type in the URL of the website if it was intuitive for them, simple to remember, and sometimes automatically shows up using the browser's auto complete feature, so that the visitor can select the URL from a drop-down in the browser. Direct access to the Library website also indicates the visitors have found the need to visit the Library website regularly, and hence have either memorized the URL of the website or have bookmarked it.

Further, bookmarking the URL can also be seen as a quick way of information access through the website without having to spend time on typing in the URL or browsing through other referring websites. However, the autocomplete feature of the browser provides a support for the visitors who have accessed the Library website from their machines in the past. The visitors' direct access suggests that they extensively use the Library resources and have opinions of it. They would indeed prefer to use the Library website for their academic information seeking habits.

Head (2007) suggested that students would like to rely on library websites for their information needs, and would like to be directed by librarians and their professors rather than resorting to simple searches using search engines. She inferred that students value quality information when it comes to their course-related work.

The referring site traffic uses the MU home page as a medium to lead them to the Library website. There is an evident association that the user perceives between the referring site and the Library website. The university homepage has a link to the Library website. Users who access using the link have to click twice through two different layers of pages to get to the library website that does not seem to hinder the online users' information seeking process through the referring website.

A smaller percent of visitors who used Google to access the Library website could be accustomed Google searchers. They would like to use a Google organic search that allows the visitors to look at all search results including from sites that are unpaid. These visitors on one hand may not be accustomed to use the Library website or might have accidentally bumped into the MU library website. There could also be a possibility that most of the visits through Google searches were during the beginning of a semester. New students who enroll in the school might use Google searches initially when they have to access the Library website. With time they get used to better and quicker ways to access the Library website.

Thus the information on traffic sources to the Library website provides interesting insights to the visitors' information access strategies. The following data can be summarized:

- Most of the Library visitors like to directly access the Library website, whereas a lower percentage access through referring site and a lesser percentage yet use Google search.

- Repeated access to the Library website has helped the Library visitors form strategies of information access such as bookmarking the Library URL, using the browser's auto complete feature, or the visitors may have memorized the Library URL.

- The direct access visitors might find browsing through the university website a hindrance to their style of information browsing, as it requires surpassing two layers of pages to get to the Library website.

- Other visitors who browse the Library website might find it easier to access the Library website through the university's home page.

- Visitors who accessed through Google might have stumbled upon the Library website or they might be new to the school.

Traffic sources can be a crucial start to learning the information seeking behavior.

5.6.2 Learning about visitors information seeking by the sources they use to access – direct or through referring site

Information seeking of visitors can vary. Visitors who do not search for the Library website either come to the website directly or they use a referring site. The discussion on direct traffic in the previous section showed how majority of the visitors preferred accessing the Library website directly. Further, this section looks closely at the information activities of visitors when they use the different access routes as direct or through a referring site.

Figure 17, Figure 18 and Figure 19 provide data for direct access visitors and the referring site visitors. The figures indicate that:

- For direct traffic, visitors access 2.63 pages per visit (Figure 17 -a), and average time on site is 02:37 (Figure 17 -b).

- Looking at the referring site traffic, there were 4.67pages/visit (Figure 18 -a), average time spent on site is 03:54(Figure 18 -b).

- More than 70% of visitors who access directly and through referring sites are returning visitors.

- The bounce rate for the direct traffic is higher at 61.95% (Figure 17 -c) and that of the referring site is 18.42% (Figure 18 -c).

- Most of the referring site visitors access the Library website from the university home page.

Visitors who access the Library website using referring sites visit more pages and spend more time on the website, where as visitors who access directly are more directed in their information seeking, and visit fewer pages per visit in a smaller average time on site.

It could possibly mean that direct users were aware of the paths of the information they seek than visitors who come from referring sites. However, the visitors who access both directly and through referring sites are returning visitors. It can be inferred that some visitors, no matter

154

how long they have been accessing, would still be using the university home page to access the Library website.

Also, the figures might demonstrate that visitors who come to the website through the referring sites take more time when searching for information on the website than the ones who come directly. Visitors who come directly to the Library website may be in a hurry to leave the website once they get their information need fulfilled.

The bounce rate figures indicate that visitors who access the Library website directly have a specific information need for which they visit the Library website. These visitors click on one of the external links and leave immediately, thus contributing to a higher bounce rate. On the other hand, visitors who access the Library website through a referring site do not seem to have well-defined information needs or they may not be aware of what resources the Library can provide to fulfill their information needs. They would, instead, like to browse through the Library website before clicking on an external link or deciding to close out of the Library website. The visitors who access through referring websites do not add to the bounce rate. It is important that any definite interpretation of how the Library visitor's information need is being met can be made by looking at the context of a user's and the content of the pages she visits.

Zhang et al. (2004) studied users' information seeking habits in an academic library portal system. They classified users' information seeking into "undirected," "semi-directed," or "directed," with an indication that most users fell into the latter two categories. For directed information seeking, they further postulated that users tend to spend lesser seeking time per click, more average clicks per session, more session time on an average and mostly monitor and browse. Analytics does not provide information on the number of clicks the visitors make during a session, so it is hard to know the level of involvement in terms of the clicks made on a page during their visit. On the other hand, since there is a high bounce rate for the direct traffic to the MU library website, analytics cannot capture users' activities once they leave the Library website.

155

Therefore, it is difficult to infer the information seeking activity of a visitor who comes through a direct access to the Library website to go to an external resource.

Visitors to the Library website can vary based on the nature of their information need and their experience with the information source. The following can be summarized:

- The visitors' traffic that comes directly to the website seems to be more goal-directed about their information seeking as compared to the referral site visitors.

- The referral site visitors might have an undefined information need or they might not have a clear information strategy in their minds, therefore they take more time and visit more pages.

- The direct visitors use the Library website to take them to the Library's external resources.

Knowledge of information behaviors and motives of users can be acquired.

5.6.3 Learning about visitors search intentions using their search activities

Information searching has been considered as a subset of information seeking (Spink et al., 2006). Visitors' information need is going to drive them to the Library's website for information seeking. Further, the information search strategy that visitors are going to formulate is going to be translated from the visitors' information need (Lucas and Topi, 2006). Figure 20 provides details for visitors' search preferences. The numbers indicate that the pages/visits when visitors come to the website using search is 3.77 pages/visit and an average time on site of 3:13 minutes, which is more than what the direct traffic figures for pages/visit and average time on site is.

Visitors who would like to search for information might indicate that they are time constrained and do not want to put an effort to browse. Browsing through the Library website would require visitors to have knowledge of the menu items that they might have gained with experience. For example, a visitor looking for a database needs to have knowledge of the

Library's subscription of the database, branch or the discipline the database should be categorized, and the exact spelling of the database in order to scan through a sorted list of database should be evident. Such an instance will be further influenced by visitors' experience. The visitor might want to escape these decision elements and prefer to conduct a search to access the Library resources. "Zipf's Principle of Least Effort" provides an explanation for the behavior of visitors who prefer to search and then adopt a course of action that will expend the probable least average of their work—the least effort (Case, 2005). The figures from the graph indicate that though these visitors are searching the website, they are still visiting more pages and spending more time on site. Therefore, an earlier assumption about visitors stumbling on to the MU library website when using search may not hold true. The interpretations provide a narrow slice of what the visitors information behavior might encapsulate, however more detailed investigations need to be pursued.

Search behavioral studies have tried to interpret why users search. A study on student's access behavior by Zhang et al. (2004) implies that if students know exactly what they want then it is likely that they would visit the specific website, but if not, then they would probably do a search. Bracke (2004) pointed out the necessity of library websites to consider visitors' search behavior and to improve their website design accordingly. They also suggested that the library should consider factors such as age of the user and the influence of the web at large on users' behavior.

Users' information search can be further investigated by looking at the terms that they use to search. Figure 22 provides the keywords that the visitors used for the search. Among the keywords used there were more than 1000 terms used that were associated with the University of Missouri and the Ellis library such as Mizzou, MU or Ellis. Some other searches were conducted for databases, contact names, services that the Library provides and events. It is evident that most

of the searchers who visited the Library website did not stumble upon information as they intended to search the MU library.

Ingwersen and Jarvelin (2005, p. 332) stress that work tasks are an important part of the information seeking behavior of a user. They point to certain actor dimensions associated with information seeking. Some of the actor specific dimensions that they point to are their "domain knowledge," "experience on search task," "sources of difficulty," and "motivation and emotional state." Further, they point at *information need* types of actors that relate to *information searching* that is more specific, that is when the actor knows what she wants. The need types that the authors describe were, "known item searching, known data element searching, known topic or content searching, factual data searching" (Ingwersen and Jarvelin, 2005,p. 292).

The following can be said from the previous discussions:

- Most of the visitors who access the MU library search formed a very small proportion of the whole user base of the Library for the Fall 2008 semester.

- It is possible that for a small section of the Library's user base, searching is a habit. A possibility of these visitors stumbling upon the Library website is negated.

- Among this small section of visitors, most users access the MU library website employing search terms that seem to be well directed to MU library such as "mu libraries", "mizzou library".

- Visitors that searched the website remained on the site for some time and visited, on average, multiple pages.

Users' information searching behavior can clarify perceptions of the Library website.

5.7   Content

The content section in analytics provides information on the metrics for the different pages of the website. The website content can be evaluated based on specific metrics of performance such as pageviews, unique pageviews, average time on page and bounce rate. The

data provided for content in chapter four in Findings is categorized under *top content, content by title, top landing pages and top exit pages, site overlay, and navigation summary.*

Information on the content of the website can further help in an understanding of what pages a visitor would access more than others. Looking at detailed metrics for each of these pages, users' activities through the pages can be interpreted. What groups of pages top the visit metric could provide the importance of particular set of pages among the numerous pages the Library holds. Visitors' information access to different pages can help identify the importance of those pages for the visitors.

Further interpretations emerge from the content of the page the visitor most frequently lands on or exits from. The navigation summary can help see the visitors' sequence of actions along the different page and the content of the same. Since the Library website holds resources that take the visitors out of the Library website, a very small section of the visitors' information seeking that is on the Library's website can be captured before the visitor leaves for an external link. An analytics implementation on the Library website can capture only a part of the user information behavior.

The following section provides interpretations of the visitors' information access through the Library pages. The discussion also attempts to understand how visitors' interpretations of information through their access path lead to successive information selection. The categories discussed in this section are as follows:

- Understanding the visitors' information selection through the content visited

- Inferring visitors' preferences of Library services based on the content by title

- Understanding visitors information seeking path preferences through the pages visited and left

- Learning visitors' information-seeking preference based on the website layout

- Understanding visitors' sequence of actions using navigation summary

159

5.7.1 Understanding the visitors' information selection through the content visited

The Library website provides services to its online users through the various pages that are hosted in the server. The content on these pages presents to the users various information that relates to the services provided by the Library. The visitors who access the Library website go through the pages, read the content and further decide which links on these pages they would like to click on. Looking at the most visited page would help in an understanding of what information the users seek.

Figure 23, Figure 24 and Figure 25 presents the unique pageviews, exit rate, and the time on page metrics for the listed content by the percent of contribution to the total site. The Library website home page for the Fall 2008 semester has the most views, most time spent, and most exits when compared to all other pages on the website.

Visitors seem to view the Library home page as an important place to find information. The Library home page is perceived as an important pointer to the resources that the Library has. However, the fact that the visitors, apart from viewing the most, also spend the most time on the home page, about 30% of the whole site, could mean that the information organization on the website does not match with the users' mental model. As a result, the user is lost and spends a longer time to find the link she is looking for. It could also be that the user finds a lot of distracting but useful information on the website.

Brantley et al. (2006) studied users' actions on a library portal. According to the study the users did not see having control of the library website as making a difference to their information search process. They would prefer shortcuts on the library website that allows them quick access to the resources provided. The study also proposed that library websites need to conform to good web practices such as improving visual layout, shorter pages, fewer textual explanations, and more understandable terminologies.

Tsakonas and Papatheodoru (2006) concluded that users valued information relevance and flexibility, which will enable them to control the quantity of information received. The two most influential attributes according to them were ease of use and ease of navigation.

How the Library decides to organize the content in the pages will determine how the users seek information. A summary here concludes that:

- Users visit the Library website home page the most and spent the most time on it.

- The users might perceive the Library site as a pointer to the different information resources that can be found in one place.

- There are links that go to pages within the Library website such as the list of databases, Journalism, Engineering and so on; however a majority of visitors would still like to go to the external resources that the Library provides directly from its website.

- There is a possibility that the visitors find the content on the Library website useful, and they take time to read it.

Useful behavior interpretations could be made based on the content of the page.

5.7.2 Inferring visitors' preferences of Library services based on the content by title

Content by title provides information on the Library pages grouping by the title of the page. All the same title pages fall under the same group. Information access of the visitor through the pages as seen by the title of the page can provide meaningful insights, because titles are more comprehensible than URLs. Also, by looking at the page performance by title an understanding of the significance of the pages to the users' can be obtained.

Figure 26 provide information on the performance of the top most groups of pages and the specific pages under those pages. Visitors find the "University of Missouri Libraries" groups of pages more useful with a pageview contribution of almost 38 % (Figure 26 -b). The second most unique pageviews were that of "Find Articles and Databases" that received a 11% unique

pageviews contribution (Figure 26 -c).Most of the activities that visitors perform on the Library website are centered on the Library's services provided.

Looking at specific pages under these groups, among the "Find Articles and Databases" groups of pages, the visitors searched for the default databases that the Library provides the most. The default database section provided a unique pageview contribution of 50%. The visitors accessed the University of Missouri default page more than the default databases page of about 80%; however the Library default page further leads to the default database page. The visitors might be accessing the Library's home page to visit the default database page. The MU library needs to create titles of pages and organize them keeping in mind the information organization of the website.

The visitors' access of the different groups of pages can provide an overall impression of the Library's roles in the users' information resource preference:

- There should be the title on the Library pages to show the role of that page in the service portfolio offered to the user.

- The title assigned to the pages of the Library website was most accessed when it had the title as 'University of Missouri Libraries.'

- The visitors also visited the Library mostly to use the articles and databases provided by the Library.

There was not much organization in the page titles for a greater understanding of Library website usage.

5.7.3 Understanding visitors' information seeking path preferences through the pages visited and left

Top landing pages and *top exit pages* provide information on the pages from which visitors entered the website, and the pages they left the website respectively. Information on the

pages most visitors used to enter the website, including the ones they landed on using search, will help understand their information preferences that they have.

Figure 29 and Figure 30 show the top landing pages and the top exit pages respectively on the Library website. The figures indicate that:

- A very high percentage of visitors to the Library website come through the Library home page.

- Visitors to the Library website have been starting their information seeking process from the Library's homepage.

- The top most exit pages that the visitors use are the Library home page (40.36%).

Landing pages on the Library's default home page suggest that visitors start their information seeking from the Library's home page either through direct access, referring sites, or when searching. Most of the Library URLs are long and complicated for the visitor to memorize and directly access it. The Library home page has a simple URL – *mulibraries.missouri.edu* that the visitors can directly access. Also, the visitors might see the Library home page as a one-stop access point to all the resources they can choose from, and the visitors are habituated to go to the resources from the home page.

Exit pages for the Library can have multiple interpretations depending on the specific user actions to exit. A user clicks on a link on the page that goes to an external resource that is the exit that the Library would like to have. In case a user closes out of the Library website that is considered as an exit too, but it is detrimental to the purpose of the Library. Also, exit rates from the Library's home page suggest that the visitor came to the Library website, visited certain other pages, came back to the Library's home page and exited by clicking on an external link or closed the website. It is interesting to see what the visitors are looking for before coming back to the home page and exiting.

The *top landing pages* and *top exit pages* to the Library website provide useful information about the visitors information seeking pathways. It can be summarized that –

- Most of the time visitors come to the Library website through the Library's homepage probably because the home page URL is the simplest one to remember

- The visitors might see the Library home page as a one-stop access to all the resources they can access.

- The visitor could be habituated to access information from the Library's home page.

- Most of the exits from the Library website happen from the Library's home page

- The visitors come to the website but visit other pages before coming back to the home page to leave the website.

- It is likely that visitors are unable to find the required information when they follow the links from the homepage thus they come back to the home page to exit.

Such data can provide various insights to visitors' information behavior.

5.7.4 Learning about visitors' information seeking preference based on the website layout

The site overlay feature allows the seeing of the clicks on the different links on the site. Looking at the site overlay gives an idea of how the visitor looks at each parts of the website when it opens up in their browser. This feature can further provide a look at the click report of the different links on the site, which can help draw a connection of how the site's graphical design could prompt more clicks on one link more than others.

Figure 31 provides the site-overlay report for the MU library. Looking at the report, it seems that most of the clicks that the MU library website got for the Fall 2008 semester were on the "Find Database" and "Search for Articles on a Topic" of 26% each. Visitors used the Library website mainly to access these two options that lead to the same resource. For links that lead to

the same page, analytics adds up the individual clicks for each of the links and shows the same percentage figures for the two links. Clicks on the other links of the website are negligible.

A drawback of site-overlay is the inability to look back on a past site, in case the site's look has changed over time. If the site looked different at a past time, then the data for the time selected in the report is juxtaposed on the current site. The data does not allow interpretations of the previous design of the site. It seems from the site overlay data that the visitors on the MU library website are drawn towards the left panel, where the two most used links- "Find Database" and "Search for Articles on a Topic" are embedded in the menu. It seems the visitors' necessity to go to those two links have prompted them to click on those links the most, even though these links are not visible. The long time spent on the home page probably is because the most used links are not obvious. One study has found that visibility of links on the library home page and 'Whats New' are the two most preferred links by users (Crowley, 2002).

Visitors' information seeking preferences on the website can be found looking at the website layout. To summarize:

- The site-overlay feature under analytics explains if the site has changed over time and the data reported is for a different site layout than the present.

- Under normal circumstances the site overlay feature of analytics can be used to understand the clicks made on the different links on the site.

- It is true that in the current case visitors would mostly use the two links – "Find Database" and "Search for Articles on a Topic".

- The links most used are embedded among other menu options. It might be one reason why users ended up spending more time on this page.

It would be interesting to further investigate if the visitors use certain links motivated by their need for that link or because it merely catches their attention.

5.7.5 Understanding visitors sequence of actions using navigation summary

The *navigation summary* option provides the page report that indicates the entrances to a page, the pages that were visited before that page, the exits from that page, and the pages the visitor went to after that page (Google Analytics, 2007). The navigation summary reports a continuity of the aggregate users' actions. The report helps in representing what paths the visitors might take in their information seeking process.

Figures 32, 33, 34, and 35 provide the sequence of users' behavior when they visit the online databases to which the Library subscribes. The figures indicate that:

- Most of the visitors on the Library website home page go there to access the Library database.

- When on the database listing page, the visitors would like to look at the Library's listing of popular databases.

- Of 88 % of visitors who go to next pages from the database page, 11% went to popular databases sections

- Other users who went to business psychology or education databases constituted lesser than 5% of the users.

- Of the visitors who chose to look at the popular databases section, a considerable 43% went to "Academic Search Premier" and the second most popular database the visitors went to was "Lexis Nexis Academic."

- The navigation summary shows 56% exits from the Library catalog link.

The visitors' access to popular databases listed by the Library further helps visitors decide which databases they would like to use for information seeking. This further reaffirms Head's (2007) postulation that visitors rely on the library for quality resources. The Library catalog is external to the system, so the higher exit rates come from there. Though there were more users who visited the databases than the Library catalog, this refutes the common

assumption that visitors usually see the online catalog as a sole way to find information on the library website as in a search engine (Cockerell and Jayne, 2002).

Over time, the perceptions of Library use are changing. Visitors see the Library as more of a quality source that will direct them to quality articles through databases, or journals. Crowley (2002) also found that visitors would prefer to see sections of the popular databases on the top of the list of databases.

The *navigation summary* provides useful information. Based on the above discussion it can be said that:

- The visitors' sequence of actions mostly leads them to the default database listing from the Library's home page from where most of them would like to look at popular databases that the Library recommends.

- The navigation feature can benefit the Library by helping it understand the specific pages that would prompt the visitors to actually visit a resource made available to them.

- It can also be noted whether the visitors prefer one resource over the other or if visitors make use of the special information services provided to them such as listings of popular databases.

- The Library could also see what popular databases are preferred by the user such as Academic Search Premier or Lexis Nexis Academic.

The navigation summary feature can further help in understanding how visitors' activities match the MU library's expectations.

5.8   *Recommendations*

Web analytics has been commonly used in the commercial sector, mainly for the purpose of maximizing revenues. However, libraries have a different purpose of providing information to patrons and their main concern is to see that their patrons can get to all the resources and find the

information they need. Analytics metrics have been traditionally interpreted for corporate purposes. The interview with the MU library web committee and exploration of the Library's Google analytics implementation brought out some of the issues that hindered interpretation of the users' information behavior using analytics. Following are some of the primary factors that need to be kept in mind when implementing and interpreting data reported using web analytics.

- Complicated URLs – The Library website has innumerable pages that are further complicated by long URLs. With multiple URLs the report gets longer and, hence gets unnecessarily populated with more figures. Long URLs further make the pages they refer to unintuitive. In the top content report, the URLs further complicated interpretation. For interpreting the URLs one needs to be aware of the destination page for these URLs. There needs to be a reference that has a list of the URLs of the Library's website and the corresponding destination pages these URLs lead to. It is easier to interpret intuitive URLs in analytics reports. The relations between the pages can be interpreted by looking at them.

- Organization of pages – The MU library websites has many pages that have been accumulated over time. Presence of numerous pages has made organization of the pages difficult. Some of these pages are also redundant. At times an analyst needs to refer to a long list of URLs, which further complicates the interpretation process. Pages are noted by multiple URLs such as URLs that lead to the database resources have multiple denotations as '/search/database/default.aspx', '/search/databases/main.asp,' and 'http://mulibraries.missouri.edu/search/databases/main.asp.' Branch URLs are denoted by multiple URLs such as the engineering library URL consists of the terms engineering or engr. The naming conventions further make it difficult to segregate the branch specific information from the main Library report. Organizing the pages based on a

168

folder structure will allow for a systematic investigation using analytic's Content Drilldown feature.

- Title naming conventions - The Library has many interior pages that make interpretation difficult. For example the "Find Articles and Databases" group of page is used frequently by the Library's online visitors. There are many interior pages within each title page that have titles that match the main title page. When users click on the main title page, in this case "Find Articles and Databases," performance of the list of URLs that belong to the interior pages are revealed. The number of interior pages for this title is 1071 pages. However, the views for the URLs after the 232nd row is 10 or less, the rows after the 105th is 50 or less, and from the 73rd row onwards is 100 or less. These pages that are hidden in the Library websites are rarely accessed by the visitors. There is a possibility that these pages are accessed by the Library staff only, so the fewer number of visits might be as a result of the staff visits.

- Branch library data – The Library would like to know usage of the branch library gateways. In order to understand the branch usage in the current implementation the Library needs to search for the particular branch URL from the content report. However, this method is not efficient for a thorough report of the branch website performance, as the information obtained will be restricted to the content metrics for that particular page. It is necessary to have a separate profile created for each of the branch library implementations. Analytics reports, when visitors access from the branch library websites, can help provide useful insights to the branch and main library connections.

- Data flaws – Though analytics provides a one-stop shop for all the usage activity happening on the Library website, it still needs to be interpreted wisely. Analytics

reports are available using scripting that can differentiate between different user accounts even when they are from the same IP. There still remain questions when multiple users are working on the same machine or are influenced by a group of users collaborating on adjacent machines. Analytics data will still report each access as one independent access. However, if the Library keeps these possible issues in mind, it can still use analytics reports for generic interpretation and look at possible areas for further investigation.

- Site overlay issues – The Library website has multiple links, and some of these links may lead to the same page. In cases where two of the links point to the same destination URL, analytics adds up the usage of each of the links and present the added visits for each link. This might indicate there were equal percent of hits under the two links that point to the same destination page. Such a flaw in link usage might confuse any interpretation when the Library would like to eliminate of one of the links. The Library needs to be cautious when making such interpretations regarding website links.

- Users' actions that might affect interpretation – Users' possible actions need to be considered when interpreting analytics. Visitors who refresh a page will show as a visit from a page to the same page in analytics navigation summary; for visitors who use the browser back button to go to the previous page visited, their access paths will show as going to a page from which they came; some visitors might leave an ongoing session, distorting the time on page metric. Visitors on campus who access the MU library might access through multiple machines, and when off campus access through a different machine which will count as different users. Sometimes many users may access through a single machines when collaborating on a project or when they are consulting on adjacent machines. There can also be situations when a user

170

bookmarked a Library resource that is accessed through the Library but is external to it. The analytics for the Library will not track such a visit. These contexts of users' actions need to be kept in mind when interpreting the analytics reports.

- Making optimal use of all features provided by analytics – Google analytics provides many features apart from the default reports that it presents. Some of these features are available whereas others need to be set in order to enable it. The features that could be useful for the MU library are as follows:

  - The *site search* feature enables the Library to capture visitors search behavior on the Library website, it provides information on the search terms used, categories the visitor used to search, search navigation that includes the pages from which they started their search, the pages they visited and so on.

  - The *user-defined* feature provides a way for analytics to differentiate between the user profiles that visit the Library website.

  - The *advanced segmentation* feature allows dissecting the data to look for specific instances of user activity. This feature is helpful in that it operates on the data post collection, so no pre-planning is needed to set this up. It can be used to look up different angles of the existing data by dimensions of visitors, traffic sources, content, and systems.

5.9   *Future work*

This study attempted to identify the use of analytics for interpreting users' behavior in a Library environment. Findings in this study are based on investigations in one Library. The implications of this study are in the multiple areas of *Human Information Behavior*, *Human Computer Interaction* and *Library Evaluation.*

171

From an HIB perspective this study can help gain an idea of how analytics can be useful as a tool for learning users' information seeking process at a broader level. Certain other aspects of correlation between different users' actions can provide valuable insights. Also, supplementary studies that qualify findings from analytics figures are important. Thelwall et al. (2005) discussed strategies that could work when it comes to interpreting log data. Two of these strategies are through correlation to data sources of known value such as correlating number of links to research activity and by conducting qualitative studies.

Analytics is constantly upgrading its features and metrics. This study was conducted over a period of two years during which time many features were added. The study has pointed to certain elements of the metrics that provide researchers some ideas and data to start thinking about analytics in newer ways. An example of a feature not adequately covered in this study, though it could have generated more discussions, is the "advanced segmentation" option that was added during the period of the study. The option could be used to look at various relations between the different metrics that could help to interpret specific angles of users' information behavior.

The study provides academic libraries a reason to start preparing their infrastructure for analytics implementation. More work is required to look at certain domains in analytics that can be specifically useful for libraries. Studies in the library sciences area need to focus more on how analytics can be adapted to the library's work culture and to be used as a part of the managerial process of academic libraries' decision-making system.

Chapter 6

Conclusion

The study investigated the use of web analytics for interpreting MU library website users' information behavior. Research Question One was to understand how web analytics can inform Library decision-making and Research Question Two investigated the use of analytics to understand users' information behavior. There were three stages in the study. Data were collected by interviewing the Library's web committee at the MU academic library and by investigating the web analytics tool by *Google*, *Google analytics*.

Web analytics implementation in the MU library indicates the Library's recognition of it as a useful tool. However, it has not utilized its full potential. The Library has mostly used analytics for improving the hyperlinks within the website. The respondents recognized the importance of analytics in making decisions that improve their web design and services; however, they also expressed concerns about certain features. Additionally, since analytics has been used extensively in the commercial sector, this seems to have created a resistance from the Library to adapt to it fully. What seems positive is that the Library recognizes the benefits of analytics barring some aspects that relate to lack of context.

The Library has to make decisions that improve its services; nevertheless, there are constraints to effective decision-making. Using advanced technology the Library can enhance its services; however, the Library needs to be aware of the visitor machine capabilities so that they can provide users machine-compatible resources. Analytics implementation in the Library can provide data only on the pages stored within the Library's systems, and eJournals are external to the system. A thorough report on visitors' eJournal usage is difficult to get from the vendors. The Library has multiple search options, but analytics can only capture information on one search feature of the website.

There has been a mixed reaction to the use of analytics in the Library. Analytics has been used for changing links, but the Library is not comfortable using analytics extensively for making major web-design related changes. However, web design plays an important role in determining users' behavior as website affordances can hinder visitors' online information behavior.

The Library is curious to see the activities of their visitors and how they access the Library and its branch resources. Knowledge of the user might not seem to have any direct consequence but can lead to interesting discovery about the users' behavior on the Library website. The changes in users' information patterns over time across the Library's resources are interesting to the respondents. How visitors can access the Library system on their own without any help has enthralled the librarians.

The Library management, however, is skeptical about suiting new technology for its use. The Library has always upheld privacy of its users and relying on an external system may violate privacy measures. Also, the web analytics tool is not set for use for some purposes of academic libraries. Libraries provide services from proprietary systems. Analytics cannot access proprietary systems. Analytics can provide the required strength to Library service evaluation and decision-making, eradicating the gaps in traditional EBL. However, the management might not consider evidence from analytics that require the Library to think in new ways or demand a change in habitual decisions that they make.

User profile information would give the Library added advantage of providing more directed services to its users. Users' information access and seeking behaviors might vary if the visitor is an undergraduate, graduate, faculty, or staff. Analytics offers a way to know the users profile by setting up the "user-defined" feature. However, the feature needs to be set up for it to be operational. Users' access from multiple computers, or multiple users through same computer, can still shroud the user's profile.

174

Analytics can help the Library in quickly interpreting the Library user's behaviors. The metrics and features provided by analytics can be used to understand any user behavior trends and changes. Metrics such as pageviews, visits, bounce rates, time on page and exits can further inform interpretation of users' behavior. Comparing different metrics, metrics of the site and page, and metrics over a past period aid in a better understanding of the users' behavior.

The Library has reservations concerning the use of analytics because the Library's philosophy is mostly in contrast to commercial goals. However, the Library needs to realize that it can garner as much benefit from analytics as the commercial sector and yet not have to deal with all the adversities that the commercial sector has undergone in developing analytics in its form at present. The advantages of analytics are mainly the aggregate data that it provides about users real visits across the website pages, and the various features that enable viewing and comparing graphs. Once the Library starts accepting analytics for its use, it can further develop its use over time to better suit the Library's goals or even consider designing its own analytics.

The typicality of the users' behavior in an academic library can bring up issues with implementing analytics. Though analytics has been identified as useful except for some restrictions, it still might be a challenge to use it for understanding users' information seeking behavior in the academic environment. Supplementary studies data can provide a more accurate picture of the information behavior of users.

The visual reports provided by analytics present a quick and clear way to understand users' information-seeking behavior. Reports indicate visitors' use of the Library website were regular though there was constant rise and fall in information seeking. The average pageviews for each day and the time the users spend on the site remains constant. Such an observation suggests the importance of looking at metrics reports that provide the total usage rather than the average. Constant average figures indicate that visitors' quality perception of the Library website was consistent throughout the Fall semester.

175

Looking at visitors machine settings indicate their machine preferences. Visitors' machine settings would enable them to make use of certain applications. The report indicates the users' preference of IE and Windows, which will influence their information behavior.

Knowing the network location of the user can provide generic information on the users' information access points. A majority of visitors access the Library website when they are on campus. Visitors' information behavior can vary based on the profile of the visitor. Analytics could be set to capture visitors' profile. However, literature on academic users' behavior can help in interpretation of the online visitors' behavior.

Most users access directly and spend lesser times with lesser pages compared to other traffic through referrals and search engines. There is also a chance that visitors who are new to the university take some time to get used to accessing the Library website directly. They may take some time to formulate their information access strategy. The visitors accessing through referring sites might have an undefined information need or they might not have a clear information strategy in their minds. There is a possibility that users who search to the website are habitual searchers.

How the Library decides to organize the content in the pages, will be a determinant of how the users plan their strategy to fulfill their information need. The visitors perceive the Library as an entrée to the different information resources that can be found in one place. A majority of visitors would like to go to the external resources that the Library provides directly from its website. The visitors would like to access most of the information through the databases the libraries subscribe to indicating their preferences towards subject databases that provide domain specific selection of articles.

A useful Library-provided service is the popular database listings. Visitors appreciate quality and visibility of resources, and would like help in their information seeking strategy. The

navigation summary feature can help the Library to see paths of visitors' activities and match the Library's expectations of it.

Analytics has provided a means for the MU library to capture users' real time data visually. Looking at users' real actions through the website pages and their specific behavior provides a way for the Library to expand on the users' information behavior theories. These theories have been established in the past without much input from real visitors' actions. Analytics provides the library a way to validate these theories regarding the new digital users. Finally, looking at visual formats of the data instead of having to go through techniques of data mining and dealing with logs of files provides the Library an added advantage as it might not need to recruit additional staff with advanced data-mining skills.

The study had limitations related to the analytics implementation of the Library. The organization of the Library website affected the interpretation of analytics report in terms of how the library website was structurally organized. The URLs and the paths to the different library pages and resources are determined by the places these pages hold in the overall structure of the library website. Analytics was implemented in the Library before the study was proposed. The Library's current website and the analytics implementation limited the scope of the study. Also, the library was in the process of changing the gateway and instituting a new resource gateway that further restricted the possibility of applying any changes to the then current Library website and the implementation of analytics on it.

Analytics has been widely used in commercial set ups and the metrics have connotations that are interpreted towards meeting organizational goals of generating more revenue. Analytics development has mostly kept in mind profit making institutions. Academic libraries are not-for-profit institutions that provide users services. On the other hand, commercial websites are revenue-driven and want to increase the "stickiness" of their website by making their users stay on the library website for as long as possible. The interview with the Library's web usability

committee was intended to take the Library's context in mind. The application of the interview findings for selecting appropriate analytics metrics provided a way to make analytics applicable and meaningful for the Library's purposes. However, the inherent meaning of the metrics in terms of what it captures and how it captures still need to be investigated. Analytics for academic library's purposes need to be redefined and appropriate interpretations of the metrics need to be established that will be useful for academic libraries.

Academic libraries have a different set of needs and expectations that analytics could help provide for. However, looking at the existing analytics application, some of the metrics and features need to be customized in order to meet the library's needs. Some of the metrics that an analytics tool implemented in academic libraries should consider are the ones related to the user-visits –the number of unique visitors, the sources they use to visit the Library's website, time they spent on each page and its contribution to the time spent on the site, pages viewed per visit, the visitor profiles, exit and bounce rate; the location of the users – if they are visiting from within the library or from outside the library, if outside the library, are they using the campus machines to access the Library website, which department or building the users are accessing from; visitors machine settings –Operating Systems, browser and plug-ins used, network location, connection speed, and the pages that the visitors use to go to the resources; Search behavior –how do the visitors use the multiple-search options provided by the library, what menu items do they choose, what keywords do they use, are their search successful and do the searches lead the user to results that the visitor was expecting. With time more studies need to look at how benchmarks can be developed that look into what is a good measure of performance for the academic library.

The data captured by analytics are aggregated, which hide the visitors' context. Visits that are of very short or long duration also influence the aggregate data. The Library needs to conduct further supplemental studies for an in-depth investigation of context to identify what the data appears to reveal. The current study on the MU library website was conducted on its *Google*

analytics implementation. A holistic account of the users' behavior on the Library's website was not possible because of the incompleteness of analytics implementation. Though the Library website is undergoing constant improvement, the current vastness of content and disorganization further complicated interpretation. Advanced technology has led to accessing instant information through mobile devices. Internet can be accessed using mobile phones. The behavior of mobile users when accessing the library website needs in-depth understanding. How many users access using their mobiles, from where, what pages they visit, how long they stay on the website, what resources they access, are some of the questions that need to be answered. Studies should also aim to investigate implementation issues of analytics in academic libraries. Improved organization of the Library website by providing appropriate titles, simplifying the URLs to make sense by looking at it, and reducing redundant pages are some of the steps that will help in improved interpretation of analytics reports. This study was conducted on the MU library website; more studies on similar academic library website are required.

Appendix A

Metrics and Features offered by Google Analytics (adapted from Google Analytics, 2009a)

1. Visitor Overview: Provides information about visitors that relate to their location, machine preferences, frequency of access
   - Map Overlay – Provides information on the location of the users accessing the webpage
   - New vs. Returning – Distinguishes visitors who access the first time to the ones who are frequent
   - Languages – Reports the language preferences set in the visitor's computer

2. Visitor Trending: Provides information on the patterns of visitor information over time.
   - Visits - Number of visits on the site made by the visitors.
   - Absolute Unique Visitors - Number of unique visitors on the site.
   - Pageviews - Total number of pages viewed by the visitors.
   - Average Pageviews – Total number of pages viewed in a given day divided by the number of visitors that day. It is considered one way of measuring visit quality. A high value suggests that visitor interacts extensively.
   - Time on Site - Time spent on site.
   - Bounce Rate - Percentage of single page visits.

3. Visitor Loyalty: Is a way to know the quality of the visit in terms of the frequency and duration of visit
   - Loyalty - Number of multiple visits.
   - Recency - Frequency with which visitors return to the site.

- Length of Visit - Entire distribution of visit time and not the 'average time on

  site'

- Depth of Visit - Entire distribution of visits instead of simply the 'average

  pageviews per visit'.

4. Browser Capabilities: Provides information on the visitors machine settings

   - Browsers - Which browsers do the visitors use

   - Operating Systems - Which operating systems do the visitors use

   - Browsers and OS - Which browsers/operating system combinations do the

     visitors use

   - Screen Colors - How many screen colors can the visitors see

   - Screen Resolutions - Which screen resolutions do the visitors use

   - Flash Versions - Which versions of Flash do the visitors have installed

   - Java Support - Is Java supported on the visitors' platforms

5. Network Properties: Presents information on the network visitors use to connect

   to the website.

   -Network Locations - Which internet service providers do the visitors use

   -Hostnames - From which hosts are people visiting the site

   -Connection Speeds - Which connection speeds are the visitors using

6. Traffic Sources: Provides information on the paths of access to the website.

   -Direct Traffic – Visitors who clicked a bookmark to come to the site or typed the

     site URL into their browser.

   -Referring Sites - Visitors who are referred from other sites.

   -Search Engines - The search engine entries to the site.

   -All Traffic Sources - Visitors from search engines, sites, and tagged links.

- Keywords – Provides information on all metrics for each keyword used to access the site.

7. Content: Presents reports on the website information in terms of the pages of the site visited.

   - Top Content - Most commonly viewed pages.

   - Content by Title - Most commonly viewed groups of pages.

   - Content Drilldown - Most commonly viewed content, viewed by drilling down.

   - Top landing pages - Pages that entice visitors to click further.

   - Top Exit Pages - Pages from where people exit the site.

   - Site Overlay - This feature allows navigation of the site as a visitor would, allowing the analyzer to view metrics for the different links on the site.

   - Navigation Summary – The report shows how often a page in the website was an entrance page, what pages were viewed before this page, how many visitors left the site from this page, and what pages were viewed after this page.

There are other features that allow comparison between two metrics of a page, and metrics of the page and the site. Comparison is also possible between two time frames. Other utilities are also provided to view the graphs by day, hour, week, and month. Options to view the graphs by table, percentage, performance, comparison, pivot table, and detail are also available. However, not all views are provided for all the metrics. Analytics also provides options to change views based on the contribution of a particular metric to – Visits, Pageviews, Time on Site, New Visits, Bounce Rate. Analytics provides certain elements called dimensions to base the reports on. The reports can be adjusted in terms of analytics dimensions such as source, medium, keyword, landing page, visitor types, landing page and so on.

Other features available in analytics can be availed only after applying the respective settings for those features such as

- Filter – Filters help in manipulating the final data such as excluding visits from particular IP addresses.

- Goal – A 'Goal' is considered a website page that serves as conversions for the site.

- Funnel visualization - A funnel represents the path that the visitors are expected to take on their way to converting to the goal

- User defined - This report allows comparing visitors from custom segments that have been defined.

- Site search - It contains reports about the visitors using the 'search' functionality on the website.

- Event Tracking - Event tracking is a method available in analytics that enables recording user interaction with website elements, such as a Flash-driven menu system.

Google analytics also added a new feature on *Advanced Segmentation* that allows slicing and dicing the analytics data with great precision by selecting dimensions of visitors, traffic source, content, ecommerce, and systems along with metrics on site usage, content, ecommerce, and goals.

Appendix B

Example for Transcript Analysis

Participants Comments	Reformulation
Respondent 5.2: It doesn't always reflect the view which is part of the problem. It does not record the outgoing links for one thing. It is part of the problem that you have to look at. Even if you tag outgoing links they are not in site overlay..	▪ Web analytics does not always show the right figures (Issues-GA) ▪ Web analytics does not record the outgoing links that are on the webpage (Issues-GA) ▪ Tagging outgoing links (such as outgoing_XXXX) does not show the different links to the same page as different. (Issue-GA)

Participants Comments	Reformulation
Respondent 1.2: That's one overall sort of problem with GA on library site, because most GA was created for companies that want to keep people on their websites as long as they can show them ads. The whole purpose of the gateway (library) is to get them off the site as fast as you can because then where they are going because it's a gateway. Its not trying to be sticky and trying to put all this information, I mean like the bounce rates on the other pages..if a company looked at the library's bounce rates they would think it was a terrible failure, but for us high bounce rates is good., is they are getting to what they need. They don't have to come back and try.	■ Most of analytics were created to be used in commercial settings (Library Related) ■ A library gateway's purpose is to get the user to the user- intended resources (Library Related) ■ A higher bounce rate may be a concern for commercial websites as opposed to library's website (Library Related)
Respondent 5.3: There is no way to tell if they stay on the site for long, it could be that they are lost or because they are looking at a sourcepack and figure out what it is they want to use. So I was not sure because if I tagged outgoing links on the gateway and on some more pages directly underneath it. I don't know how those figure out in the bounce rates because they are tag , are they counted as	■ Web analytics does not provide information on the user's context such as when they spend a long time on a page whether they are looking for some information or they are lost when finding information they want (Service, Issue-GA)

185

Participants Comments	Reformulation
bounces if I tag them or not. I do not know that. If you can find that out that would be nice.	▪ How does tagging work with analytics reports (Issues – GA) ▪ Does that show as a bounce and hence reflect in the bounce rate (Issues – GA)

Appendix C

Memo for Each Category

Category	Memo
	▪ Includes items or statements that has to do with the library services such as
	▪ Issues that can improve library users access of the library services
	▪ Actions or statements towards improving efficiency of library services
Decisions about Services	▪ Library initiatives towards understanding their users behavior in order to provide them better services
	▪ Identification of issues that raise problems with access to library services
	▪ User side factors that might affect service provided by libraries
	▪ Information on user side machine features that could be used to improve services
	▪ Support for library staffing in order to provide better services online as well as offline (reference desk)

Category	Memo
	▪ Efficient services to meet users support needs
	▪ Service related enhancements that libraries would like to make
	▪ Issues with tools that may lead to misinterpretation of service related issues
	▪ Information needs of libraries regarding the service provided
	▪ Issues with web analytics that libraries have in terms of service use information that it provides
	▪ Library related decisions that relate to user use of links
	▪ Increase in user satisfaction of the use of library services
The Library's Interest in User Behavior	▪ Any instances that indicate
	▪ User specific actions
	▪ User habits of information access

Category	Memo
	▪ Users characteristics when accessing information on the library website
	▪ Users use of the library website
	▪ Users changes in online behavior on the library website
	▪ Information needed by library management regarding their users behavior
	▪ Aspects of user behavior that libraries would like to know
	▪ Use of analytics to understand users behavior on specific resources provided by the website
	▪ Information on user profile who access the library website
	▪ Users actions on the web pages
	▪ Impression that library managers would be interested to know about the user activity in different bits and parts of the library website.
	▪ what areas the user visit more frequently

Category	Memo
	▪ Basic uses of analytics for the library's purpose
	▪ Anything to do with users perception of the content of the library website that they express in terms of the search keywords
	▪ Actions or indications of how users of the website seek the information that they need from the website
	● This category includes any statement or indication that involves
	● The navigation of the website either in terms of user's navigation behavior or the website's architecture
	● Navigational problems related to links on the library website
	● Library measures to improve navigation on their website
Decisions about Web Design	● Different ways of accessing the library website
	● User habits of accessing the library website
	● Analytics features that help in understanding the navigation behavior of libraries.
	● Navigational changes made by library
	● Indications of the links user click
	● Libraries expectation of users navigation behavior
	● Navigation design of the library website

190

Category	Memo
	• The category contained statements or indications -
	• That might only apply to libraries and their needs
	• Library related concerns with library resources
	• Library's specific concerns
	• Library specific metric interpretation regarding website usage
	• This category included-
The Library's	• Any management related actions
Management and	• Any decision related to management
Function	• Policy/promotional related statement , action or indication
	• Use of analytics for supporting management decisions
	• Library concerns regarding interpretation of analytics on usage statistics of their website
	• Issues that might crop up with analytics implementation within libraries
	• Library reports are provided by their vendors
	• Use of analytics for library management purposes
	• Knowledge and evidence for supporting decision-making

191

Appendix D

Example of Themes and Subthemes Coded With the Google Analytics Metrics

Themes	Subthemes	Respondent statements (rephrased)	Google analytics metrics applicable
Decisions about Services	Providing for specific user needs	Libraries would like to know what problems do the users face and would try to solve those. (Respondent 3 Service Evaluation)	Navigation Summary, Exit Rates, Bounce Rates
Decisions about Services	Providing machine compatible resources	What browsers do people use and if they are compatible with the resources provided (such as audio clips) (Respondent 5 service)	Connection Speed, Browser Capabilities (Browsers, Screen Resolution)
Decisions about Services	Providing machine compatible resources	What are the configuration of users machines and if the resources and website was compatible to these configurations (Respondent 3 service)	Network Properties (Connection Speed)(W2)
Decisions about Services	Providing a satisfying search experience	They put and removed some links on the gateway (Respondent 5 Service)	Site Overlay, Content by Title, Top Content (J1)

Themes	Subthemes	Respondent statements (rephrased)	Google analytics metrics applicable
Decisions about Services	Supporting staffing decisions	Peak times of user visits would help libraries to set chat reference as it is minimally staffed (Respondent 6 service)	Visitor trending—Visits, Absolute unique visits, time on site (R4)
Decisions about Services	Supporting staffing decisions	GA would help libraries to make a case for staffing the reference desk (online or desk)(Respondent 6 Service)	Visitor trending—Visits, Absolute unique visits, time on site (R5)

Appendix E

Example of analysis using axial coding

Themes	Subthemes	Statements rephrased	Explanation
Decisions about Services	Providing for specific user needs	The library has used the search function in analytics to look at specific link visits (R5.4 user behavior, service)	Libraries tend to make service related decisions based on information on how users use the different links presented on their website.
Decisions about Services	Providing for specific user needs	It might or might not be necessary to link pages from the gateway as users could be bookmarking certain pages (R6.10 User Behavior, Services)	
Decisions about Services	Providing for specific user needs	Any information on where the users are directly going to (probably by bookmarking) would help libraries to place links in appropriate places.(R3.7 Services)	
Decisions about Services	Providing for specific user needs	Libraries would like to know which links are redundant and hence get rid of those links (R6.11	

194

| Themes	Subthemes	Statements rephrased	Explanation
		Service Navigation)	
Library's Interest in User Behavior	User activity on the library website	There is a dip in the online usage of the library website during spring break (R3.6 User Behavior)	Libraries would be interested to know how users use the library website.
Decisions about Services	Providing for specific user needs	It might or might not be necessary to link pages from the gateway as users could be bookmarking certain pages (R6.10 User Behavior, Services)	

Appendix F
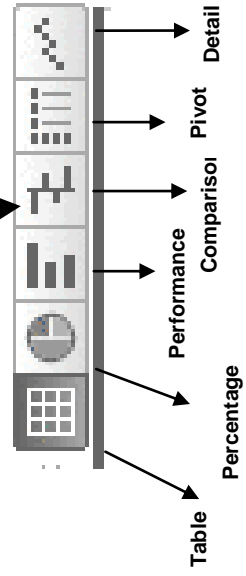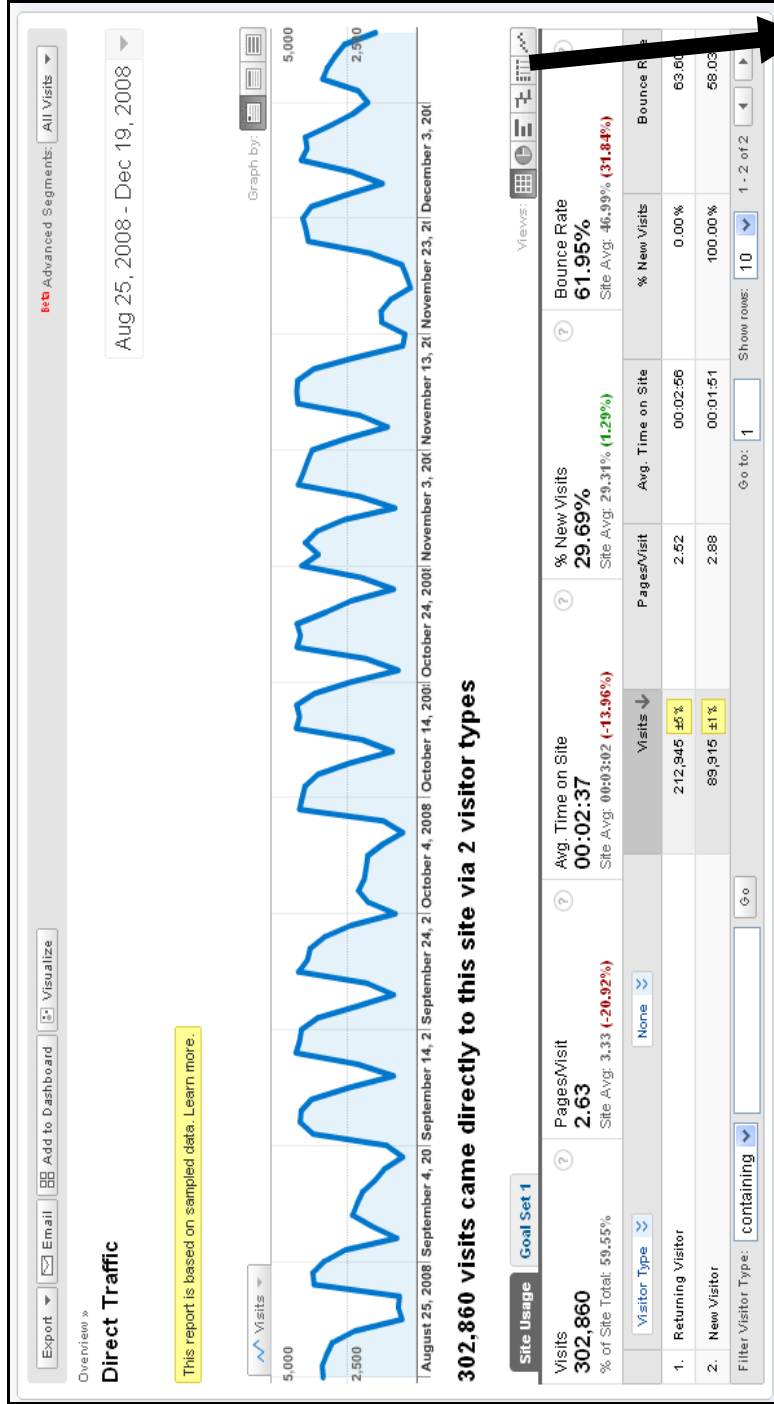
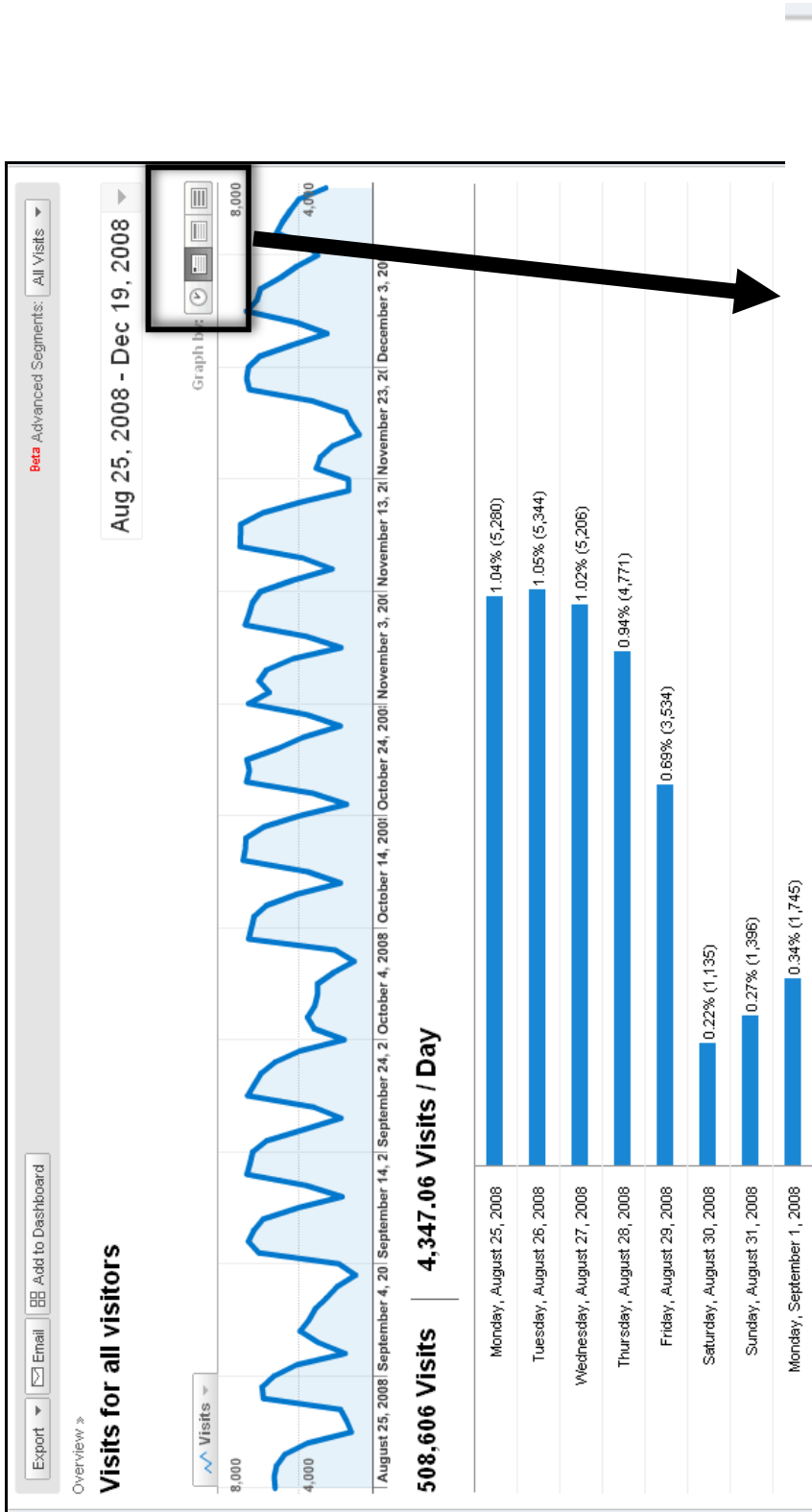*Figure F1*. Different Views Available in Analytics

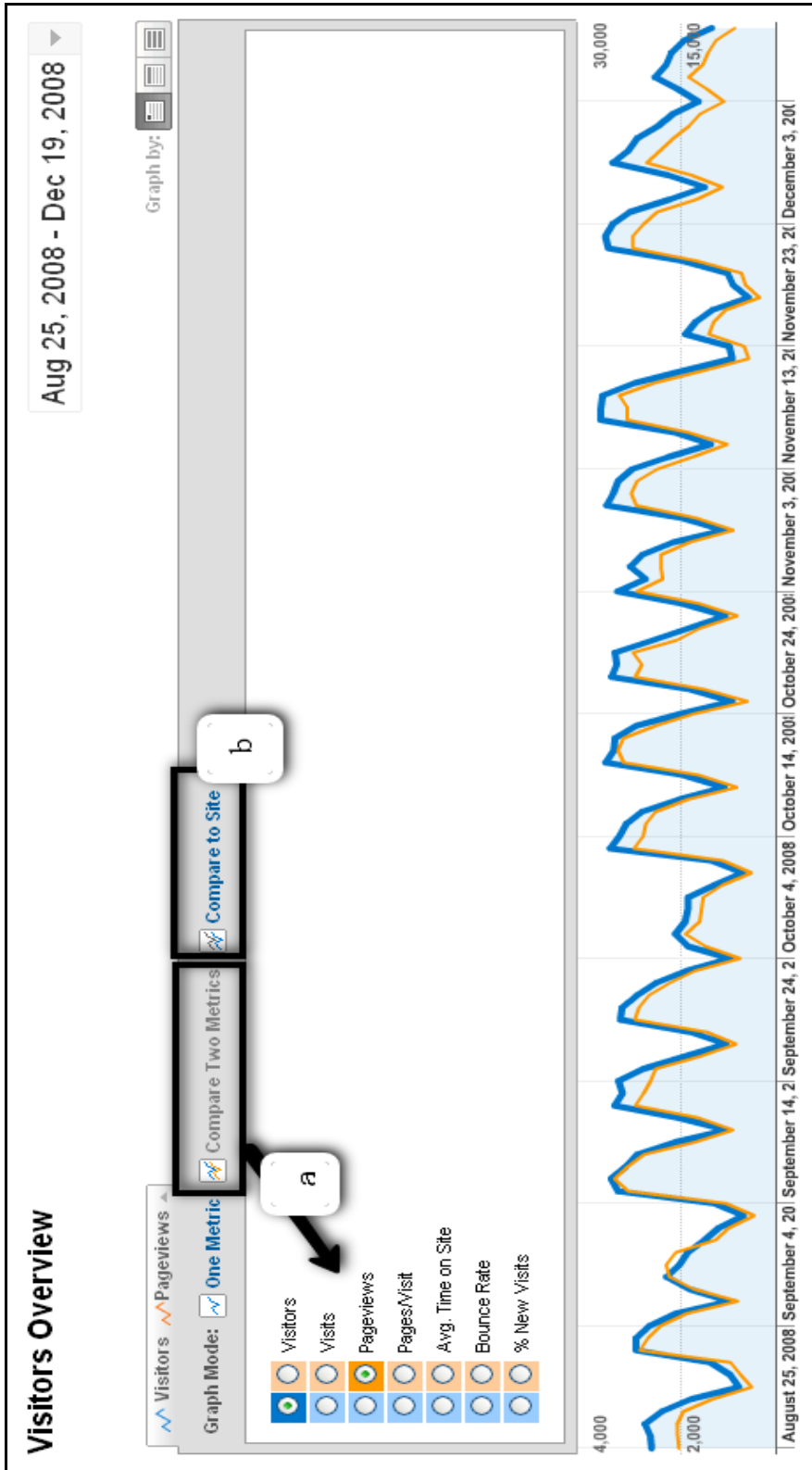*Figure F2.* Graph by Hour, Day, Week, Month

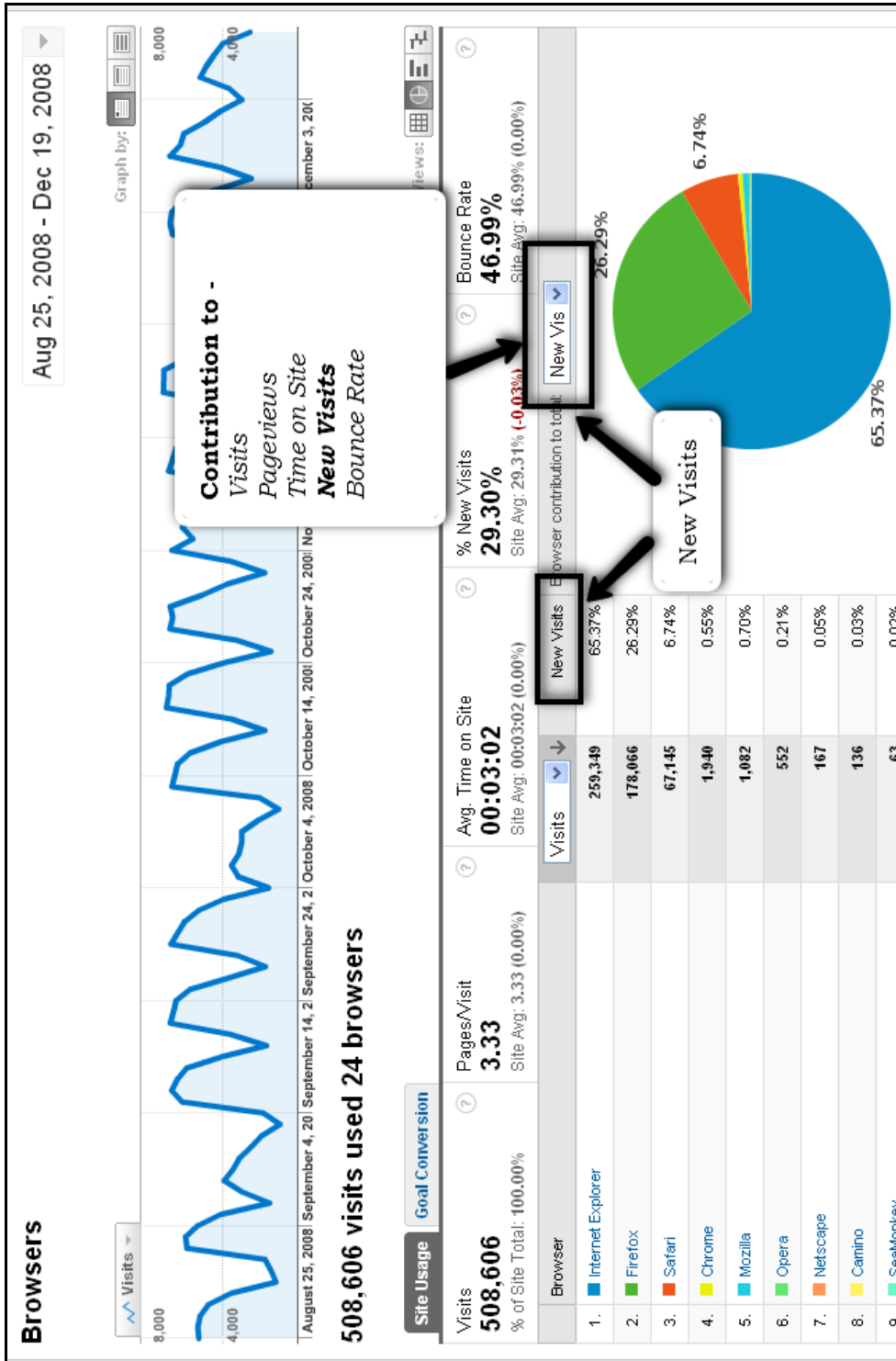*Figure F3.* a) Compare Two Metrics, b) Compare to Site

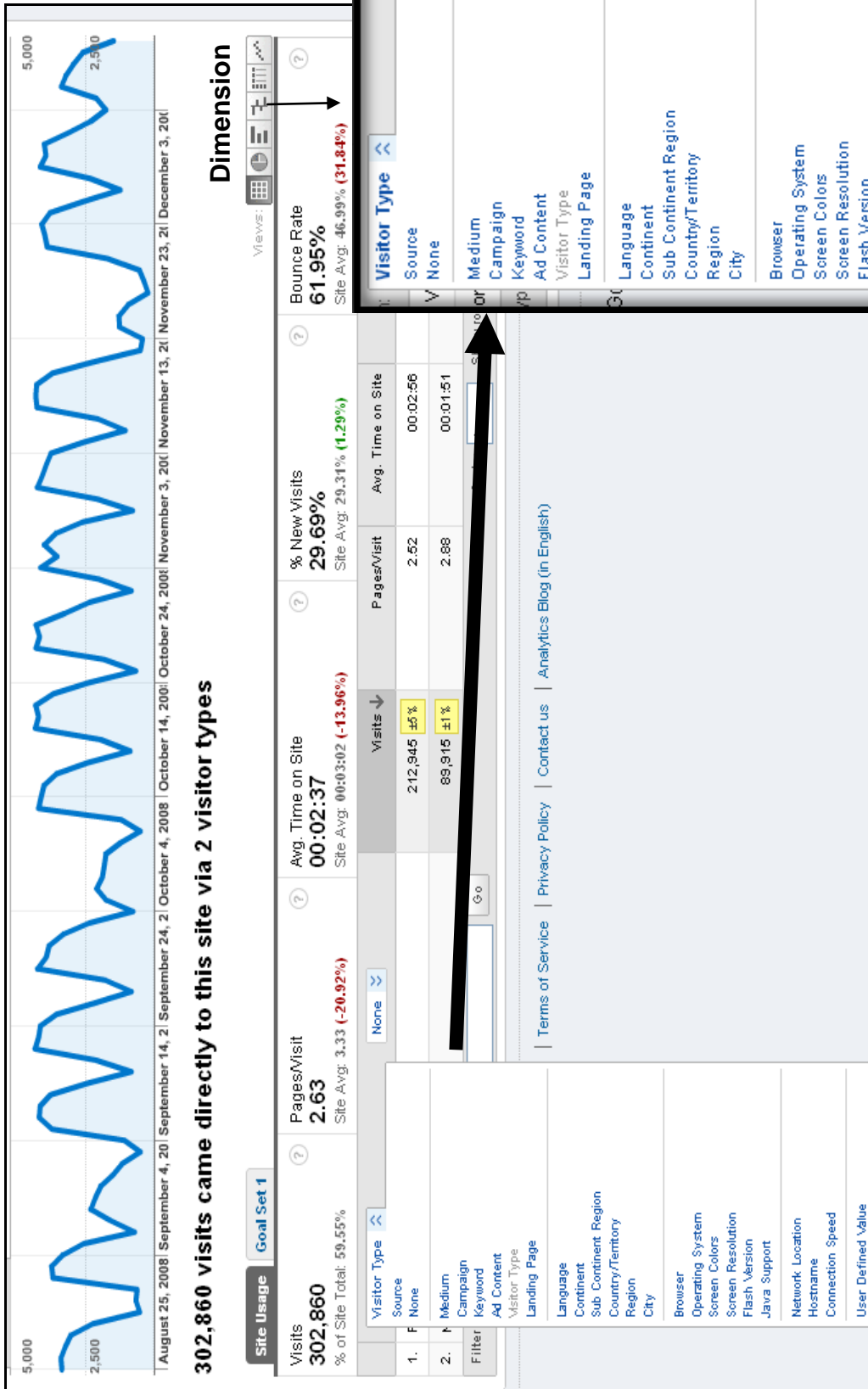*Figure F4.* Contribution of a Metric to the Visits, Pageviews, Time on Site, New Visits, and Bounce Rate

Figure F5. Reports by different Dimensions

References:

Bates, M. J. (2005). Berrypicking. In K. E. Fisher, Erdelez, S., & McKechnie, L (Ed.), *Theories of Information Behavior.* Medford, N.J.: Information Today.

Borgman, C. L., Hirsh, S. G. & Hiller, J. (1996, July). Rethinking Online Monitoring Methods for Information Retrieval Systems: From Search Product to Search Process. *Journal of the American Society for Information Science* (1986-1998). Vol. 47(7), 568-583.

Bracke, P. J. (2004, October). Web usage mining at an academic health sciences library: an exploratory study. *Journal of the Medical Library Association*, 92 (4), 421-428.

Brantley, S., Armstrong, A., & Lewis, K. M. (2006). Usability testing of a customizable library web portal. *College & Research Libraries*, 67(2), 146-163.

Buckland, M.K. (2003). Five grand challenges for library research. *Library Trends*, 51(4), 675-686.

Caswell, J.V. (2004, June). A conceptual framework for gateways. *Information Technology and Libraries*. 73-81.

Caswell, J.V. (2006). Leveraging resources in a library gateway. *Library Hi Tech*, 24(1), 142-152.

Chau, M., Fang, X., & Sheng, O. R. L. (2005, November). Analysis of the query logs of a website search engine. *Journal of the American Society for Information Science & Technology*, 56(13), 1363-1376.

Chi, E. H. (2002). Improving Web usability through visualization. *IEEE Internet Computing*, 6(2), 64-71.

Clark, L., Ting, I. H., Kimble, C., Wright, P. & Kudenko, D. (2006, January).Combining ethnographic and clickstream data to identify user Web browsing strategies. *Information Research*. 11(2), 1-19.

Cockrell, B. J., & Jayne, E. A. (2002). How do I find an article? Insights from a web usability study. *The Journal of Academic Librarianship*, 28(3), 122-132.

Cothey, V. (2002, January 15). A longitudinal study of World Wide Web users' information-searching behavior. *Journal of the American Society for Information Science and Technology*. Vol. 53(2), 67-78.

Crowley, G. H., Leffel, R., Ramirez, D., Hart, J. L., & Armstrong, T. S. (2002). User perceptions of the library's web pages: A focus group study at Texas A&M University. *The Journal of Academic Librarianship*, 28(4), 205-210.

Creswell, J. W. & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage Publications, Inc.

Davis, P.M. & Solla, L.R. (2003). An IP-level analysis of usage statistics for electronic journals in Chemistry: Making inferences about user behavior. *Journal of the American Society for Information Science and Technology*, 54(11), 1062-1068.

Detlor, B. (2005). Web information behaviors of organizational workers. In K. E. Fisher, Erdelez, S., & McKechnie, L (Ed.), *Theories of Information Behavior*. Medford, N.J.: Information Today.

Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (ARIST), 21, 3-33.

Dwek, R. (2005, June 30). Online: Whatever way you phrase it, a website's only useful if it works. *Marketing Week*, 35.

Eirinaki, M. & Vazirgiannis, M. (2003, February). Web mining for Web personalization. *ACM Transactions on Internet Technology*, 3(1), 1-27.

Eldredge, J. (2006). Evidence-based librarianship: The EBL process. *Library Hi Tech* 24(3), 341-354.

Elsevier (2008). *User Behavior* (Chapter IV). Retrieved February 6, 2008, from http://www.elsevier.com/wps/find/librarians.librarians/tulip_fr_chap4 .

Eszter, H. (2002, December). Beyond logs and surveys: In-depth measures of people's Web use skills. *Journal of the American Society for Information Science and Technology*, 53(14), 1239- 1244.

Fang, W. (2007). Using Google Analytics for improving library website content and design: A case study. *Library Philosophy and Practice – LPP Special Issue on Libraries and Google*. Retrieved September 27, 2007, from http://www.webpages.uidaho.edu/~mbolin/fang.htm.

Fisher, K. E., Erdelez, S. & McKechnie, L. (E.F.) (Eds.). (2005). *Theories of information behavior*. Medford, NJ: Information Today, Inc.

Google analytics (2009a). *Google analytics help*. Retrieved October 30, 2009, from http://www.google.com/support/analytics/bin/topic.py?hl=en&topic=15337 .

Google analytics (2009b). *Enterprise-class web analytics made smarter, friendlier and free*. Retrieved October 30, 2009, from http://www.google.com/analytics/ .

Google. (2007). *Google analytics terms of service*. Retrieved September 27, 2007 from http://www.google.com/analytics/tos.html.

Google (2009). *What does bounce rate mean?* Retrieved August 24, 2009 from http://www.google.com/support/analytics/bin/answer.py?hl=en&answer=81986

Griffiths, J. R., & Brophy, P. (2005). Student Searching Behavior and the Web: Use of Academic Resources and Google. *Library Trends* 53(4, Spring 2005), 539-554.

Head, A. (2007). Beyond Google: how do students conduct academic research? *First Monday*, 12(8). Retrieved on Apr, 28, 2008 from http://www.firstmonday.org/issues/issue12_8/head/index.html

Hernon, P., & McClure, C. (1990). *Evaluation and library decision making*. Norwood,NJ: Ablex Publishing Corporation.

Hiller, S., & Self, J. (2004). From Measurement to Management: Using Data Wisely for Planning and Decision-Making. *Library Trends* 53 (1), 129-155.

Hightower, C., Sih, J., and Tilghman, A. (1998, January). Recommendations for Benchmarking web site usage among academic libraries. *College and Research Libraries*, 59(1), 61-79.

Huntington, P., Nicholas, D. and Jamall, H.R. (2008, April). Website usage metrics: A re-assessment of session data. *Information Processing and Management* 44(1), 358-372.

Ingwersen, P. and Jarvelin, K. (2005). The integrated IS&R framework. *The Turn. Integration of Information Seeking and retrieval in context.* (pp. 259 - 311). Netherlands: Springer

Ingwersen, P. & Kalervo, J. (2005). Implications of the cognitive framework for IS&R. *The Turn. Integration of Information Seeking and retrieval in context.* (pp. 313 - 357). Netherlands: Springer

Jackson, M. (2002). The advent of portals. *Library Journal*, 127(15), 36-39.

Jacoby, J. (2005). Optimal foraging. In K. E. Fisher, Erdelez, S., & McKechnie, L (Ed.), *Theories of Information Behavior.* Medford, N.J.: Information Today.

Jacoby, G. A. & Luqi (2007, February). Intranet model and metrics: Measuring intranet overall value contributions based on a corporation's critical business requirements. *Communications of the ACM*, 50(2), 43- 50.

Jansen, B.J. and Spink, A. (2000, October/November) Methodological approach in discovering user search patterns through web log analysis. *Bulletin of the American Society for Information Science*, 27(1), 15- 17.

Jansen, B. J., Spink, A. & Pedersen, J. (2005, April). A temporal comparison of AltaVista Web searching. *Journal of the American Society for Information Science and Technology,* 56 (6), 559- 570.

Jasra, M. (2006, August 16). Web analytics comparison – Google vs. VisiStat. *The Enquiro*. Retrieved September 27, 2007, from http://www.enquiro.com/marketing-monitor/Web-Analytics-Comparison-Google-VisiStat.asp.

Khoo, M., Pagano, J., Washington, A.L., Recker, M., Palmer, B., and Donahue, R. A. Using web metrics to analyze digital libraries. *International Conference on Digital Libraries. Proceedings of the 8th ACM/IEEE-CS joint conference on Digital Libraries*, 375-384.

Kim, J.-A. (2006a). Capturing metrics for undergraduate usage of subscription databases. *ONLINE*, 30(3), 32-37

Kim, J.-A. (2006b). Toward an understanding of Web-based subscription database acceptance. *Journal of the American Society of Information Science and Technology*, 57(13), 1715-1728.

King, D.W. & Montgomery, C.H. (2002, December). After migration to an electronic journal collection. Impact on faculty and doctoral students. *D-Lib Magazine*, 8(12). Retrieved October 1, 2007 from http://www.dlib.org/dlib/december02/king/12king.html .

Kyrnin, J. (n.d.). Web Analytics: How to use web analytics effectively. Web analytics can improve your web site. About.com: Web design/HTML. Retrieved May 25, 2008 from http://webdesign.about.com/od/loganalysis/a/web_analytics.htm .

Lakos, A. (2007, October). Evidence-based library management: The leadership challenge. *Portal: Libraries and the Academy*, 7(4), 431-450.

Larson, R. R. (1991, April). The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science* Vol. 42(3), 197-215.

Lau, E. P. & Goh, D. H.L. (2006, September).In search of query patterns: A case study of a university OPAC. *Information Processing & Management*. Vol. 42(5), 1316-1329.

Ledford, J. L. & Tyler, M.E. (2007).*Google analytics 2.0*. Indianapolis, IN: Wiley Publishing, Inc.

Lincoln, Y. S. & Guba, E.G. (1985). *Naturalistic inquiry*. California: Sage Publications.

Lin, X., Liebscher, P., and Marchionini, G. (1991, August). Graphical representations of electronic search patterns. *Journal of the American Society for Information Science*. 42(7),469-478.

Liu, N., Marenco, L. & Miller, P.L. (2006, July/August). Resource log: an embeddable tool for dynamically monitoring the usage of Web-based bioscience resources. *Journal of the American Medical Informatics Association*, 13(4), 432-437.

Lucas, W. and Topi, H. (2006). Learning and training to search. In A. Spink and C.Cole (Eds.), *New directions in Human Information Behavior*, (pp. 209-226), Netherlands: Springer.

Ma, W. (2002, May). A database selection expert system based on reference librarian's database selection strategy: A usability and empirical evaluation. *Journal of the American Society for Information Science and Technology*, 53(7), 567-580.

Marchionini, G. (2002, December). Co-evolution of user and organizational interfaces: A longitudinal case study of WWW dissemination of national statistics. *Journal of the American Society for Information Science and Technology*. Hoboken. Vol. 53 (14), 1192-1209.

Mission, Vision and Strategies for the MU Libraries (2006, June). Retrieved September 12, 2009 from http://mulibraries.missouri.edu/staff/policies/POL19.htm

Nasaoui, O., Zaiane, O., Spiliopoulou, M., Mobasher, B., Masand, B. & Yu, P. S. (2005). WebKDD 2005 – Web mining and Web usage analysis post-workshop report. *SIGKDD Explorations*, 7(2), 139-142.

Nicholas, D. and Huntington, P. (2003). Micro-Mining and segmented log file analysis: A method for enriching the data yield from internet log files. *Journal of Information Science*, 29 (5).

Nicholas, D., Huntington, P., and Watkinson, A. (2003). Digital Journals, big deals and online searching behavior: a pilot study. *Aslib Proceedings: New Information Perspectives*, 55(1/2), 84-109.

Nicholas, D., Huntington, P. and Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of Documentation*, 61(2), 248-280.

Nicholas, D., Huntington, P., Jamali, H.R. and Watkinson, A. (2006, September). The information seeking behaviour of the users of digital scholarly journals. *Information Processing & Management*, 42(5), 1345-1365.

Nicholas, D., Huntington, P., Lievesley, N. & Wasti, A. (2000). Evaluating consumer website logs: A case study of the Times/The Sunday Times website. *Journal of Information Science*. 26 (6), 399 - 411.

Nicholson, S. (2003). The Bibliomining process: Data Warehousing and Data Mining for library decision-making. *Information Technology and Libraries 22* (4)

Nicholson, S. (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Proceedings of the 67th ASIS&T Annual Meeting*, 41, 496-506.

Nicholson, S. (2005). Digital Library Archaeology: A Conceptual Framework for Understanding Library Use through Artifact-Based Evaluation. *The Library Quarterly, 75*(4), 496-520.

Nicholson, S. (2006a). Approaching librarianship from the data: Using Bibliomining for evidence-based librarianship. *Library Hi-Tech* 24(3). 369-375

Nicholson, S. (2006b). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services *Information Processing & Management*, 42(3), 785-804.

Norguet, J-P, Zimanyi, E. & Steinberger, R. (2006). *Proceedings of the 2006 ACM symposium on Applied Computing, France*, 525-529. NY: ACM Press.

Pace, A. K. (2004, October). Technically speaking. *American libraries*, 35(9), 68-69.

Pagano, J. (2006, November 11), Benefits and challenges of developing a Public Sector metrics program using commercial tools. *Proceedings of the 1st international workshop on Contextualized attention metadata: collecting, managing and exploiting of rich usage information*, USA, 5-8. NY: ACM Press.

Patton, M. Q. (Ed.). (2001). *Qualitative Research & Evaluation Methods* (3rd ed.). USA: Sage Publications, Inc.

Penniman, W. D. & Dominick, W. D. (1980). Monitoring and evaluation of on-line information system usage. *Information Processing and Management*, 16, 17-35.

Peterson, I. (March 2000). Beyond hits and page views. *Journal of Electronic Publishing*. 5 (3).

Phippen, A., Sheppard, L. & Furnell, S. (2004). A practical evaluation of Web analytics. *Internet Research*. Vol. 14(4), 284-293.

Rubin, J.H. (2004, September 16). Log analysis pays off. *Network Computing*, 15(18). 76-78.

Saraf, V. & Mezbah-ul-Islam, M. (2002). Measuring library effectiveness: A holistic approach. *Journal of Library and Information Science* 27 (2), 81-105.

Sen, A., Dacin, P. A. & Pattichis, C. (2006, November). Current trends in Web data analysis. *Communications of the ACM*, 49 (11), 85-91

Spink, A., Yang, Y., Jansen, J., Nykanen, P., Lorence, D. P., Ozmutlu, S. & Ozmutlu, H. C. (2004, March). A study of medical and health queries to Web search engines. *Health Information & Libraries Journal*, 21 (1), 44-51.

Spink, A., Park, M., and Cole, C. (2006). Multitasking and coordinating framework for human information behavior. In A. Spink and C.Cole (Eds.), *New directions in Human Information Behavior*, (pp. 137-154), Netherlands: Springer.

Srinivasan, S., Amir, A., Deshpande, P. & Zbarsky, V. (2004). Grammar-based task analysis of Web logs. *Proceedings of the thirteenth ACM international conference on information and knowledge management*, 244-245. NY: ACM Press.

Starkweather, W.M. & Wallin, C.C. (1999, Spring). Faculty response to library technology: Insights on attitudes. *Library Trends*, 47(4), pp. 640-668.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research. Grounded theory procedures and techniques.* Newbury Park, CA : Sage Publications.

Tenopir, C., Wang, P., Zhang, Y., Simmons, B., & Pollard, R. (2008). Academic users' interactions with ScienceDirect in search tasks: Affective and cognitive behaviors. *Information Processing and Management, 44*(1), 105-121.

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology*, 39(1), 81-135.

Todd, R. J. (2003). Adolescents of the information age: patterns of information seeking and use, and implications for information professionals. *School Libraries Worldwide*, **9**(2), 27-46.

Tsakonas, G., & Papatheodorou, C. (October 1, 2006). Analyzing and evaluating usefulness and usability in electronic information services. *Journal of Information Science*, 32(5), 400-419.

Turnbull, D. (2005). World Wide Web information seeking. In K. E. Fisher, Erdelez, S., & McKechnie, L (Ed.), *Theories of Information Behavior*. Medford, N.J.: Information Today.

Tyler, M. E. & Ledford, J. L. (2006). *Google analytics*. Indianapolis, IN: Wiley Publishing, Inc.

University of Missouri (2008). *MU Facts*. Retrieved May 07, 2008 from http://missouri.edu/about/mufacts.php

Wang, P., Hawk, W. B., & Tenopir, C. (2000). Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing and Management, 36*(2), 229-251.

Yi, K., Beheshti, J., Cole, C., Leide, J. E. & Large, A.(2006, July). User search behavior of domain specific informatin retrieval systems: An analysis of the query logs from PsycINFO and ABC-Clio's Historical Abstracts/America: History and Life. *Journal of the American Society for Information Science and Technology*. Vol. 57(9), 1208-1220.

Zhang, D., Zambrowicz, C., Zhou, H. & Roderer, N. K. (2004). User information seeking behavior in a medical Web portal environment: A preliminary study. *Journal of the American Society for Information Science and Technology*, 55(8), 670-684.

VITA

Anindita Paul is the daughter of Mahendra Prasad Paul and Purnima Paul. She was born in the city of Kanpur in India. She did her schooling from multiple cities in India and went on to study BSc. in Mathematics from the University of Burdwan. She then completed her MBA from the University of Visva-bharati following which she was working in the IT Industry. During her work period she got interested in users' information behavior and joined the PhD program in Information Science at the University of Missouri in 2004. She completed her PhD. in 2009.