A Data-Driven Approach to Measure Web Site Navigability

XIAO FANG, PAUL JEN-HWA HU, MICHAEL CHAU, HAN-FEN HU, ZHUO YANG, AND OLIVIA R. LIU SHENG

XIAO FANG is an assistant professor of information systems at the David Eccles School of Business, University of Utah. He received his Ph.D. in management information systems from the University of Arizona. His current research interests are in the areas of business intelligence, social computing, and data mining. His research has been published in journals such as *Journal of Management Information Systems, ACM Transactions on Information Systems, ACM Transactions on Internet Technology, Communications of the ACM, INFORMS Journal on Computing,* and *Journal of the American Society for Information Science and Technology.*

PAUL JEN-HWA HU is a professor and David Eccles Scholar at the David Eccles School of Business, the University of Utah. His current research interests include information technology in health care, electronic commerce, digital government, knowledge and innovation management, and technology-mediated learning. Hu has published papers in *Journal of Management Information Systems, Information Systems Research, Journal of the Association for Information Systems, Decision Sciences, Decision Support Systems, Journal of the American Society for Information Science and Technology, Journal of Information Systems (AAA), Journal of Service Research,* and various IEEE and ACM journals and transactions.

MICHAEL CHAU is an associate professor in the School of Business at the University of Hong Kong. He received a Ph.D. in management information systems from the University of Arizona and a B.Sc. in computer science and information systems from the University of Hong Kong. His current research interests include information retrieval, Web mining, data mining, social media, electronic commerce, and security informatics. His research has been published (or is forthcoming) in journals such as *ACM Transactions on Management Information Systems, ACM Transactions on Information Systems, Communications of the ACM, Decision Support Systems, IEEE Computer, IEEE Transactions on Knowledge and Data Engineering, Journal of the Association for Information Systems, Journal of American Society for Information Science and Technology,* and *MIS Quarterly.* He is the author of more than 100 articles and has been ranked in a research productivity study as the fourteenth most productive researcher in the field of information science in the period 1998–2007.

HAN-FEN HU is an assistant professor of information systems in the Department of Management, Entrepreneurship and Technology, University of Nevada, Las Vegas. She received her Ph.D. in information systems at the David Eccles School of Business, University of Utah. Her research interests include e-commerce and behavioral issues in online settings. Her research has been published in *IEEE Intelligent Systems, Journal of the American Society for Information Science and Technology, Communications*

of the Association of Information Systems, and *Electronic Commerce Research and Applications.*

ZHUO YANG is a principal business intelligence analyst at Symantec Corporation. He earned a Ph.D. in information systems at the David Eccles School of Business, University of Utah. His research focuses on data mining and business intelligence. His research has been presented at INFORMS Annual Meeting, International Conference on Information Systems (ICIS), Workshop on Information Technologies and Systems (WITS), and Workshop on e-Business (WeB).

OLIVIA R. LIU SHENG is Emma Eccles Jones Presidential Chair of Information Systems at the David Eccles School of Business, University of Utah. She received her Ph.D. in computers and information systems from the University of Rochester. Her research focuses on global information technology and management including distributed enterprise data and knowledge management, electronic commerce, telemedicine, and distributed learning. She has published over 40 journal articles and has received funding from the U.S. Army, National Science Foundation, IBM, Toshiba, Sun Microsystems, Hong Kong Research Grants Council, Asia Productivity Organization, and SAP University Alliance.

ABSTRACT: Web site navigability refers to the degree to which a visitor can follow a Web site's hyperlink structure to successfully find information with efficiency and ease. In this study, we take a data-driven approach to measure Web site navigability using Web data readily available in organizations. Guided by information foraging and information-processing theories, we identify fundamental navigability dimensions that should be emphasized in metric development. Accordingly, we propose three data-driven metrics—namely, power, efficiency, and directness—that consider Web structure, usage, and content data to measure a Web site's navigability. We also develop a Web mining–based method that processes Web data to enable the calculation of the proposed metrics. We further implement a prototype system based on the Web mining–based method and use it to assess the navigability of two sizable, real-world Web sites with the metrics. To examine the analysis results by the metrics, we perform an evaluation study that involves these two sites and 248 voluntary participants. The evaluation results show that user performance and assessments are consistent with the analysis results revealed by our metrics. Our study demonstrates the viability and practical value of data-driven metrics for measuring Web site navigability, which can be used for evaluative, diagnostic, or predictive purposes.

KEY WORDS AND PHRASES: data-driven navigability metrics, Web metrics, Web mining, Web site navigability, Web site navigation.

WEB SITE DESIGN IS CRUCIAL TO THE SUCCESS of virtually all applications for e-commerce, digital government, and online learning. A well-designed site can attract visitors and help them find target information effectively and quickly [92]. In contrast, poorly designed sites hinder visitors' information seeking and can lead to dissatisfaction and lost business revenues [34]. Navigation and search represent two principal means for finding information on a Web site [87]. This study focuses on navigation, which in

general refers to the process of navigating through a Web site to find target information [59, 84]. According to Palmer [59], navigation is a fundamental and crucial way to locate information on Web sites.

Essential to navigation is the structure of the hyperlinks that connect different pages on a Web site [5, 80], which can significantly affect user experience and satisfaction [17, 57, 68, 76, 84]. However, many people still have difficulty finding information on a Web site [84], often because of its ineffective hyperlink structure [33, 57, 84], which inhibits visitor from finding target information, thereby leading to dissatisfaction [17, 57, 68, 76, 84].

We examine Web site navigability, defined as the degree to which a visitor can follow a Web site's hyperlink structure to successfully find information with efficiency and ease [19, 51, 59]. People often find information on a Web site by sifting through its hyperlinks [57]; thus, a well-designed structure helps people locate information more effectively and quickly because it provides a mental model of the type and location of information that facilitates their path selections through the interconnected pages [84].

Measurement is crucial to navigability. A review of extant information systems literature suggests the common use of perceptual measurements that target people's self-reports (e.g., [19, 59]). Although perceptual measurements are valuable for conveying a person's assessment, we can analyze and measure navigability with a data-driven approach that considers readily available Web data, including Web content, structure (i.e., hyperlink structure), and usage data (i.e., Web logs). Such a data-driven approach reflects visitors' browsing behaviors and enables development of metrics that can be used in combination with perceptual measurements for a fuller depiction of a site's navigability. Toward that end, analyzing Web data is promising and has been applied to develop navigability metrics (e.g., [9, 93, 94]). However, most previous studies that use Web data to develop navigability metrics have several limitations. First, choices of the focal aspect of navigability they target in metric development seem driven by observations or intuitions (e.g., [93, 94]). Thus, the resultant metrics tend to focus on a select navigability dimension, providing limited insight into a Web site's navigability. Second, most existing metrics consider partial Web data, typically Web structural data (e.g., [9, 93]), and therefore cannot fully convey a site's navigability. The navigability of a Web site conceivably involves not only its hyperlink structure depicted by Web structure data but also visitors' browsing behaviors informed through Web usage and content data. In addition, many previous studies tend to concentrate on metric development, often placing less attention on evaluations with actual users; consequently, they offer limited empirical evidence about the viability or pragmatic value of the proposed metrics (e.g., [9, 90]). To address these limitations, we aim to achieve the following objectives in this study:

1. to propose data-driven navigability metrics that consider more comprehensive Web data, including Web content, structure, and usage data;
2. to develop a viable Web mining–based method that enables the computation and use of the proposed metrics to measure Web site navigability systematically; and

3. to perform a rigorous evaluation study to produce empirical evidence that a Web site's navigability (e.g., high versus low), as revealed by our metrics, is congruent with actual user performance and assessments.

We propose three metrics: *power,* which indicates the likelihood that a visitor can successfully find target information on a site by traversing its hyperlink structure; *efficiency,* which conveys the extent to which a visitor can quickly find target information on a site by navigating its hyperlink structure (e.g., using less time or fewer clicks); and *directness,* which reflects the ease with which a visitor can decide where to move from the current page to the target information. Overall, our metric development is guided by information foraging theory [62, 64] and information-processing theory [52], which together point to several user-centric dimensions fundamental to Web site navigability: the likelihood of successfully, efficiently, and easily finding target information. Our metric formulations, anchored in the law of surfing [33], take as inputs a focal site's hyperlink structure (derived from Web structure data) and user browsing behaviors (derived from Web usage and content data) to measure the site's navigability quantitatively and comprehensively. To measure a site's navigability, we develop a Web mining–based method that models the site's hyperlink structure as a directed graph, distinguishes between content and index pages, and incorporates appropriate Web mining techniques to discover user visit patterns from Web logs. We empirically examine the proposed metrics by conducting an evaluation study that involves two large, real-world Web sites and 248 voluntary participants.

Literature Review

INVESTIGATIONS OF NAVIGABILITY APPEAR IN SEVERAL RESEARCH STREAMS. In this section, we review prior research on navigation in general and navigability in particular, summarize representative studies that develop navigability metrics with Web data, and highlight the differences between our navigability metrics and existing ones.

Web Site Navigation and Navigability

Web site design includes multiple key elements that affect user performance and experience [35, 87]. Among them, navigation is crucial. People often visit Web sites to find information, whether their purpose is to buy directly, build knowledge, search, or deliberate [54, 75]. Many people attempt to locate target information on a Web site [45]. As Spink and Cole [71] note, people visit Web sites to fulfill their information needs or interests.

Navigation is crucial because it can significantly affect people's assessments of a Web site [68, 79]; it also has substantial effects on user performance or actual use of a site [24]. To find information on a Web site, many people choose to follow the site's hyperlink structure; therefore, by linking the different pages about various resources, products, or services with an effective structure, organizations can make their site more "navigable" by better facilitating visitors' information seeking on the site [55].

According to Webster and Ahuja [84], navigation systems provide an important means for supporting people's browsing and path selections to locate target information on a Web site.

Navigability is closely related to navigation, but with some subtle differences. In general, navigation refers to the overall process of browsing a Web site to find information, whereas navigability, as we define herein, is specific to a Web site's hyperlink structure and emphasizes the notion that visitors can follow the structure to find information successfully, efficiently, and easily. We note the efforts for measuring navigability with perceptual items. For example, de Castro et al. [19] measure navigability with question items germane to perceived ease of navigation, effectiveness, and efficiency; Palmer [59] proposes items that emphasize page sequencing, layout organization, and navigation protocol consistency.

A review of previous research examining navigation or navigability converges on the importance of hyperlink structure and shows the availability of perceptual measurements for navigability. A Web site's navigability, as we define herein, represents the confluence of its hyperlink structure and visitors' browsing behaviors; therefore, navigability measurements may benefit from a data-driven approach that considers different, related Web data and targets the dimensions critical to visitors' finding information on a Web site.

Measuring Navigability with Web Data

Web data can be classified broadly as content, structure, or usage data [73]. Content data consist of Web page contents, such as texts on Web pages; structure data depict the hyperlink structure connecting different pages; and usage data are the records generated by users' browsing on a Web site (i.e., Web logs).[1] Accordingly, Web mining can be classified on the basis of the particular Web data employed. Web content mining employs text mining techniques [85] to analyze Web content data and can support applications such as sentiment analysis, stylometric analysis, fake Web site detection, and Web information retrieval [1, 2, 3, 4, 24, 25, 85]; Web structure mining discovers patterns from hyperlink structures [13, 41]; and Web usage mining reveals visitors' browsing patterns from Web usage data [26, 73]. Prior research has applied Web mining to study different aspects of Web sites, such as overall Web site success [69] and the effectiveness of automatically created index pages [61]. Web mining also has been used to measure site navigability, toward which we review representative studies in the following section.[2]

Previous research has attempted to mine Web data to develop navigability metrics, with a common focus on Web structure data. Botafogo et al. [9] propose two metrics, compactness and stratum, to assess the connectedness and structural organization of a hypertext system. Several extensions also have been undertaken for hypertext or hypermedia systems [18, 50, 89]. For example, to measure the navigation of a hypermedia system, Yamada et al. [89] extend the metrics of Botafogo et al. [9] by considering interface distance. Zhang et al. [93] measure navigability from a structural complexity perspective and propose several navigability metrics based on the total number of

hyperlinks. Yen [90] develops a metric that considers a page more accessible than other pages if more hyperlinks point to that page and its source pages are located closer to the homepage. Zhou et al. [94] use a Web site's structure data, in combination with an emulated surfing model, to calculate its navigability. Efforts have also been taken to develop navigability metrics with Web usage or content data, perhaps to a lesser extent. For example, Gupta et al. [30] analyze usage data and propose a click-ratio measure (i.e., the ratio of the number of clicks to the number of distinct pages in a visit session). Bayesian network models have been developed to measure navigability at the site or page level; these models take Web content and structure data as inputs to calculate the navigability of a Web site [48, 82].

Our study differs from previous research on developing navigability metrics with Web data in several ways. First, our metrics emphasize three fundamental navigability dimensions: the likelihood, efficiency, and ease of finding information on a Web site. In contrast, many existing navigability metrics developed with Web data measure a select navigability dimension. For example, Gupta et al.'s [30] metric focuses only on the efficiency of finding information, and Zhou et al.'s [94] metric stresses the likelihood of finding information. Second, our metrics consider more comprehensive Web data (i.e., structure, usage, and content) than many existing metrics, such as Zhang et al. [93] and Zhou et al. [94]. Because navigability sits at the confluence of hyperlink structure and user browsing, metrics developed with partial Web data may not fully reveal a Web site's navigability. Third, the formulations of our metrics are novel. Our metric formulations specify how to use the hyperlink structure and Web browsing behaviors extracted from Web structure, usage, and content data to measure a Web site's navigability, according to fundamental navigability dimensions.

Theoretical Foundations

Conceptually, people's information seeking on a Web site can be understood with information foraging theory [39, 62, 63, 64]. This theory extends optimal foraging theory [47] to explain how people find information in general; it posits that people develop strategies that optimize the utility of their information gains in relation to the costs they incur. In other words, a person moves through different states toward the goal state by choosing the most cost-effective paths and taking advantage of the support or cues available [62]. This process manifests a utilitarian consideration, grounded in the rational utility model for a spreading activation process that moves people through the hyperlink structure to find information on a Web site [27, 62]. In this connection, visitors likely switch to another "path" if they expect their information gains (i.e., toward finding target information) along the current path to be lower than those along another path [27].

According to information foraging theory, people are likely to modify their browsing strategies to maximize the rate at which they gain information while minimizing the associated costs [39, 63, 64]. The selection or modification of a browsing strategy for finding information on a Web site can be enlightened by the information-processing theory [52], which stresses cognitive, mental processing and posits that people process

the information they receive rather than merely respond to stimuli, that is, people normally process information through different thinking, analysis of stimuli, situational modifications, and obstacle evaluations, based on the information-processing model [78].

Because of the enormous number of traversal paths available on a Web site, visitors might employ heuristic processing by using a mental model to choose an appropriate path to find the information they need [12] instead of engaging in systematic explorations that are tedious and demand stringent time and effort requirements. Taking online retail as an example, a vendor usually offers a wide array of products and provides voluminous, detailed service information, which makes a "brute force" way of finding information on the site impractical for visitors. People often have various time constraints (e.g., too many things competing for the finite time available) and are limited in their cognitive, mental-processing capacity, which constitutes their bounded rationality [67]. With heuristic processing [12], visitors make judgments about their traversing paths (e.g., analysis of stimuli, situational modification, obstacle evaluation) to mitigate the associated costs, as measured by time, physical efforts, and cognitive processing.

These theories jointly imply that when attempting to find information on a Web site, visitors care about not only the likelihood of locating target information but also the efficiency and ease of doing so. The importance of efficiency and ease is consistent with a human–task interaction view [49], which suggests that the personal costs of performing a task include time and cognitive resources. To find information on a site, people must exert cognitive, mental efforts to analyze, estimate, and select appropriate paths to traverse. Their browsing behavior can be further informed by personal construct theory [40], which describes the phase-based process of construction people experience as they build a worldview by assimilating information on a Web site. Through this process, which is intricately interwoven with cognitive, mental, and physical activities, visitors move toward the target information using sense-making, which involves several realms of activity, such as physical actions and cognitive efforts, associated with the process and its phases [43]. The underlying sense-making process is congruent with cost–benefit theory [74], which posits that rational actors select a path only if the benefits of doing so outweigh the associated costs.

In summary, these theories explicitly highlight the importance of successfully, efficiently, and easily finding information, which in turn reveals several dimensions to be emphasized in our metric development. Efficiency can be evaluated in terms of physical effort or time [43, 68], and ease can be assessed with cognitive, mental processing [49]. The physical efforts a person spends to find information on a Web site can be measured as the number of clicks, which is highly correlated with the time a person needs to find information on the site [68]. Accordingly, our metric development targets the following: how likely a person is to find target information successfully on a site, the number of clicks needed to do so, and the ease of choosing among possible links from a current page. The navigability dimensions we target reflect a user-centric orientation; people, because of their time and cognitive-processing constraints, prefer paths that maximize their likelihood of successfully finding information yet still require

few clicks and minimal cognitive processing. Guided by these theories, we develop three data-driven metrics that correspond to the respective navigability dimensions, use comprehensive Web data, and analyze them in the light of the law of surfing.

Method and Metrics for Measuring Navigability

IN THIS SECTION, WE FIRST DESCRIBE A WEB MINING–BASED METHOD that enables the calculation of navigability with our metrics and then detail our development of each metric.

A Web Mining–Based Method for Measuring Navigability

Measuring a Web site's navigability requires proper representation of its hyperlink structure, effective discovery of users' information-seeking targets, and rigorous assessment of how well the hyperlink structure facilitates their achievement of such targets. Because a Web site resembles a complex graph [20], we apply graph theory [86] to represent a site's hyperlink structure. Web mining offers a viable means for analyzing visitors' browsing behaviors on a site [44, 53, 88]. Thus, we apply appropriate Web mining techniques to Web usage and content data to reveal important regularities in users' browsing—namely, access patterns. We then use access patterns to approximate users' information-seeking targets. Finally, we build on the law of surfing [33] to develop data-driven metrics for navigability by examining how well a Web site's hyperlink structure enables visitors to find target information on the site successfully, efficiently, and easily.

As Figure 1 shows, our method first processes Web logs to identify visit sessions and then parses a focal site to produce parsed pages. We classify each parsed page as either a content or an index page; the content pages, together with the identified visit sessions, serve as inputs to access pattern mining to discover access patterns. Our method constructs a distance matrix that represents the site's hyperlink structure; this matrix, in combination with the identified access patterns, allows for calculation of the focal site's navigability, according to the proposed metrics. In the following, we detail each step of the method, except the navigability calculation step, which we defer to the next subsection.

Web Log Preprocessing

Analysis of Web logs requires preprocessing, which includes cleaning, session identification, and session completion [16]. Our method cleans Web logs by removing accessory log records created in response to objects embedded in a user-requested page (e.g., pictures). Log records produced by Web spiders (i.e., programs launched by a search engine to gather Web pages) are also removed because they do not reflect visitors' browsing. We analyze the cleaned logs to identify sessions that constitute the basic units for access pattern discovery. A session is a sequence of Web page accesses during a visit to a Web site [16]. Web log records have no explicit session designations; to identify sessions, our method takes a 30-minute time-out approach that has been
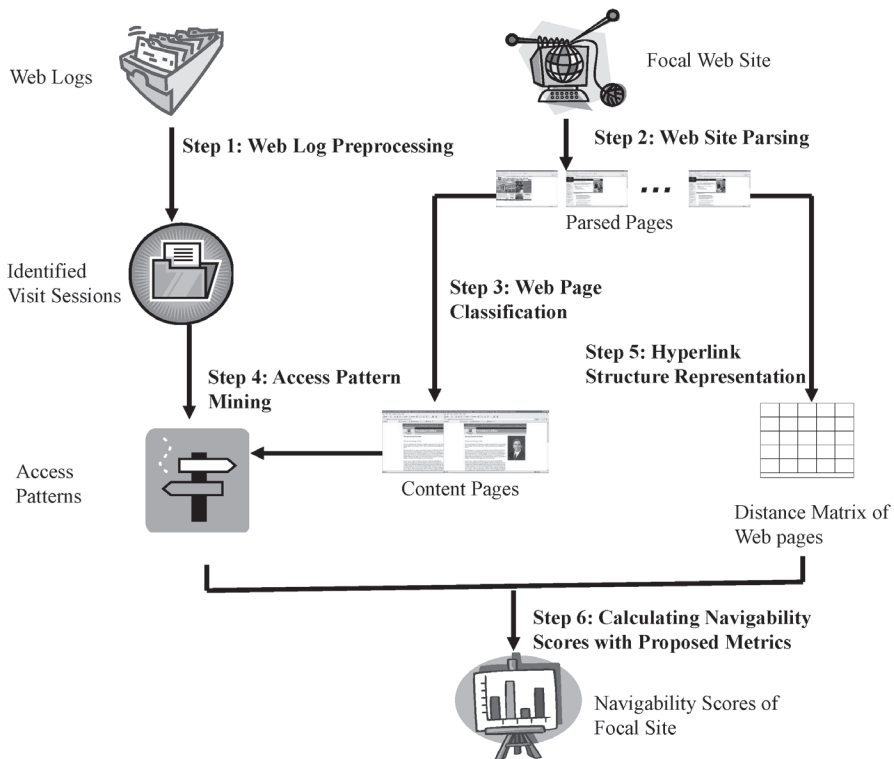
Figure 1. Web Mining–Based Method for Measuring Navigability

shown to be effective for session identification [70]. Each session is then completed by identifying the pages accessed by the visitor but not recorded in logs (i.e., cached pages)[3] and including these pages in the session. To identify accesses to cached pages, our method adopts the heuristics of Cooley et al. [16].

Web Site Parsing

Our method parses the focal site by gathering its pages and extracting their important features. It uses a spider program to gather all the pages on the site; each page is then parsed for important features, including the number of outgoing links, the number of internal outgoing links (i.e., hyperlinks pointing to pages on the same site), the number of external outgoing links (i.e., hyperlinks pointing to pages not on the same site), the size (measured by the number of bytes), the number of words, and the number of anchor text words. We do not include hyperlinks that point to anchors on the same page when analyzing the features of a Web page.

Web Page Classification

A page can be classified as a content page or an index page [72]. Index pages help people navigate a site and usually have many hyperlinks pointing to other pages but

little content [61], whereas content pages contain specific information (e.g., product description) and normally have fewer hyperlinks than index pages. Content pages often constitute targets of visitors' information seeking [72], whereas index pages are means to such targets; we therefore use sequences of content pages that users frequently access to approximate their information-seeking targets.[4] The sheer number of pages available on a Web site makes manual classification of pages tedious and ineffective. Thus, our method builds an automatic classifier with a support vector machine [81], which is effective in classifying tasks [37].[5] The classifier classifies pages as index or content pages on the basis of their respective features identified previously. To apply the classifier to classify pages, we assembled a training sample of 200 randomly selected pages from the focal site. Two domain experts highly knowledgeable in Web site design manually classified each page; they then met to reach an agreement on each classification result through face-to-face discussions. The resultant tagged pages were used to train the classifier, which was then applied to classify each page of the site.

Access Pattern Mining

Our method uses frequently accessed sequences of content pages (i.e., access patterns) as proxies for information-seeking targets. For example, people frequently visit a university's Web site to find tuition information and then payment instructions. In this case, the information-seeking target can be approximated by the access sequence of the two content pages: the tuition information page and the payment instructions page. Access patterns can be discovered from Web logs. In particular, preprocessed Web logs are represented as $S = \{s_i\}$, $i = 1, 2, ..., k$, where s_i is a session and k denotes the number of sessions. For each session s_i in S, index page accesses are removed from s_i. Thus, a session s_i can be represented as a sequence of m content pages consecutively visited in the session, $s_i = <p_{i,1}, p_{i,2}, ..., p_{i,m}>$, where $p_{i,j}$ refers to the jth visited content page in s_i, $j = 1, 2, ..., m$. An access sequence of content pages, $u = <q_1, q_2, ..., q_n>$, denotes a sequence of content pages accessed consecutively, where q_l is the lth visited page in u, $l = 1, 2, ..., n$. An access sequence, $u = <q_1, q_2, ..., q_n>$, is defined as contained in a session $s_i = <p_{i,1}, p_{i,2}, ..., p_{i,m}>$ if and only if there exists z, $1 \le z \le m - n + 1$, such that $q_l = p_{i,l+z-1}$, for $l = 1, 2, ..., n$. If an access sequence is contained in a session, it occurs. We calculate the occurrence rate $v(u)$ of an access sequence u as the ratio of the number of sessions that contain the sequence to the total number of sessions. Key access sequences have an occurrence rate that exceeds a prespecified threshold; the set of all key access sequences entails the access patterns. Extracting all the key access sequences from logs is a nontrivial task. The problem space is enormous because of the exponential number of candidate sequences attainable by exhausting all plausible permutations of content pages and the sheer volume of records in logs. Major algorithms for the efficient discovery of frequent sequences from large amounts of data include the AprioriAll [6], SPADE [91], and PrefixSpan [21, 60]. Among them, PrefixSpan is generally the most advantageous in terms of running time and memory usage [21, 60]. We therefore adopt PrefixSpan to discover key access sequences from logs.

Hyperlink Structure Representation

A site's hyperlink structure is represented by a distance matrix in which pages are indexes and the distance between any two pages is an element. We model the focal site as a directed graph (i.e., pages as vertices and hyperlinks as edges) and measure the distance between two pages according to the distance definition in graph theory [86]; that is, the distance from page A to page B on a site is the length of the shortest path from A to B, as measured by the number of hyperlink clicks. The distance is ∞ if no path exists from page A to page B. To construct a distance matrix, we define $Y(p, y)$ as a set of pages, where the distance from page p to any page in $Y(p, y)$ is y clicks. By the definition of $Y(p, y)$, we have

$$Y(p, 0) = \{p\}. \tag{1}$$

We represent the set of Web pages pointed to by the hyperlinks on page p as $T(p)$, which can be discovered by parsing page p. The distance from page p to any page in $T(p) - Y(p, 0)$ is one click, that is,

$$Y(p, 1) = T(p) - Y(p, 0). \tag{2}$$

Generalizing Equation (2), we have

$$Y(p, y) = \begin{cases} \{p\} & \text{if } y = 0 \\ \bigcup_{\forall k \in Y(p, y-1)} T(k) - \bigcup_{j=0}^{y-1} Y(p, j) & \text{if } y > 0. \end{cases} \tag{3}$$

In Equation (3), $\bigcup_{\forall k \in Y(p, y-1)} T(k)$ consists of all pages that are distant from page p by at most y clicks, because the distance from page p to any page k in $Y(p, y-1)$ is $y-1$ clicks and any page in $T(k)$ is one click away from k. Excluding the pages for which distance from page p is less than y clicks (i.e., pages in $\bigcup_{j=0}^{y-1} Y(p, j)$), we find that $\bigcup_{\forall k \in Y(p, y-1)} T(k) - \bigcup_{j=0}^{y-1} Y(p, j)$ consists of all pages distant from page p by y clicks. Figure 2 summarizes the procedure for calculating $Y(p, y)$ for all the pages on a site, which also identifies the shortest path between any two connected pages. Constructing a distance matrix from $Y(p, y)$ is straightforward and not described.

Data-Driven Metrics for Measuring Navigability

Corresponding to the fundamental dimensions of navigability suggested by the guiding theories, we develop three metrics—power, efficiency, and directness—that can be formally defined, quantitatively calculated, and empirically examined. As mentioned, these navigability metrics, which use the focal site's hyperlink structure and the identified access patterns as inputs, are premised in the law of surfing [33], which identifies regularities in Web surfing behaviors by characterizing the number of hyperlinks clicked during a visit session with a probabilistic distribution. Specifically, the law of surfing states that the probability $p(k)$ of surfing k hyperlinks in a session can be expressed as

Input: a set of Web pages in a Web site
Output: $Y(p, y)$ for each Web page p in the Web site

for each Web page p
   find $T(p)$ by parsing Web page p
end for
for each Web page p
   $i = 0$
   $Y(p, i) = \{p\}$
   while $Y(p, i) \neq \emptyset$
     $Y(p, i+1) = \emptyset$
     for each $k \in Y(p, i)$
       for each $h \in T(k)$
         if $h \notin Y(p, j)$ for all $0 \leq j \leq i$
            $Y(p, i+1) = Y(p, i+1) \cup \{h\}$
         end if
       end for
     end for
     $i = i+1$
   end while
end for

Figure 2. Calculation of $Y(p, y)$

$$p(k) = \sqrt{\frac{\beta}{2\pi k^3}} \exp\left[\frac{-\beta(k-\alpha)^2}{2\alpha^2 k}\right] \qquad k = 1, 2, ..., \qquad (4)$$

where the mean of the probability distribution (i.e., average number of hyperlinks surfed in a session) is α and the scale parameter β determines the shape of the probability distribution. The parameters α and β can be estimated from visit data recorded in Web logs [33]. The law of surfing offers a robust analysis of Web browsing behaviors and formally reveals the prominent regularities of Web surfing depth [27, 31, 46].

We define $G(l)$ as the probability of surfing *at least* l hyperlinks during a session, which is the sum of $p(k)$, where $k \geq l$,

$$G(l) = \sum_{\forall k \geq l} p(k) \qquad l = 1, 2, ... \qquad (5)$$

Thus, we can obtain the following from Equation (5):

$$G(l) = \begin{cases} 1 & \text{if } l = 1, \\ G(l-1) - p(l-1) & \text{if } l > 1. \end{cases} \qquad (6)$$

Using Equations (4) and (6), we can derive $G(l)$ from $l = 1$, and $G(\infty) = 0$. We can now propose the three metrics.

Power

We use the key access sequences, discovered from Web logs, to approximate visitors' information-seeking targets. Let U be a set of n key access sequences discovered from logs, $U = \{u_i\}$, $i = 1, 2, ..., n$, and $u_i = <p_{i,1}, p_{i,2}, ..., p_{i,m}>$, where $p_{i,j}$ is the jth visited content page in u_i, $j = 1, 2, ..., m$. For an information-seeking target approximated by a key access sequence u_i, power $R(u_i)$ can be measured as the probability of locating all the content pages in u_i sequentially, from $p_{i,1}$ to $p_{i,m}$. Let p_s denote the start page of seeking for u_i. If $p_s \neq p_{i,1}$, the distance from p_s to the first sought page $p_{i,1}$, $d(p_s, p_{i,1})$, can be derived from the distance matrix (constructed in the hyperlink structure representation step).[6] Visitors willing to surf at least $d(p_s, p_{i,1})$ hyperlinks can locate $p_{i,1}$ from p_s. According to Equation (6), $G(l)$ is the probability of surfing at least l hyperlinks; therefore, the probability of surfing at least $d(p_s, p_{i,1})$ hyperlinks is $G(d(p_s, p_{i,1}))$. Accordingly, the probability of locating $p_{i,1}$ from p_s can be approximated as $G(d(p_s, p_{i,1}))$. After locating $p_{i,1}$, a visitor can continue to locate $p_{i,2}$, and the probability of locating $p_{i,j}$ from $p_{i,j-1}$ also can be approximated as $G(d(p_{i,j-1}, p_{i,j}))$, where $2 \leq j \leq m$. If $p_s \neq p_{i,1}$, the power $R(u_i|p_s)$ of locating u_i becomes

$$R\left(u_i \,\middle|\, p_s\right) = G\left(d\left(p_s, p_{i,1}\right)\right) \prod_{j=1}^{m} G\left(d\left(p_{i,j-1}, p_{i,j}\right)\right) \qquad \text{if } p_s \neq p_{i,1}. \qquad (7)$$

Likewise, if $p_s = p_{i,1}$, we obtain

$$R\left(u_i \,\middle|\, p_s\right) = \prod_{j=2}^{m} G\left(d\left(p_{i,j-1}, p_{i,j}\right)\right) \qquad \text{if } p_s = p_{i,1}. \qquad (8)$$

Let P(start of seeking for $u_i = p_s$) be the probability of seeking for u_i starting from page p_s, which can be estimated from surfing data recorded in Web logs. For example, if we observe in Web logs the following sessions that record the seeking of target information u_i—<homepage u_i>, <homepage u_i>, <page1 u_i>, <page1 u_i>, <page2 u_i>—seeking of u_i would start from the homepage, page1, or page2, with probability P(start of seeking for u_i = homepage) = 0.4, P(start of seeking for u_i = page1) = 0.4, and P(start of seeking for u_i = page2) = 0.2. Accordingly, we can calculate the $R(u_i)$ as

$$R\left(u_i\right) = \sum_{\forall p_s} P\left(\text{start of seeking for } u_i = p_s\right) R\left(u_i \,\middle|\, p_s\right). \qquad (9)$$

Not all key access sequences are equally important. We therefore introduce a weight $w(u_i)$ of u_i in U, calculated as

$$w\left(u_i\right) = \frac{v\left(u_i\right)}{\displaystyle\sum_{\forall u \in U} v\left(u\right)}, \qquad (10)$$

where $v(u_i)$, the occurrence rate of u_i, is identified from the access pattern mining step. Therefore, the power $R(U)$ of a Web site can be measured as the weighted probability of achieving each information-seeking target in U on the Web site:

$$R\left(U\right) = \sum_{i=1}^{n} w\left(u_i\right) R\left(u_i\right). \qquad (11)$$

Power, $R(U)$, falls inclusively between 0 and 1 and generally can reveal the probability that a visitor achieves an information-seeking target by navigating through a Web site's hyperlink structure. When $R(U) = 0$, no information-seeking targets can be achieved; when $R(U) = 1$, all the targets can be achieved with a probability of 1. The higher the value of $R(U)$, the more powerful is a Web site's hyperlink structure design for helping visitors locate target information on the site.

Efficiency

In general, the closer a page is to the currently visited page, the more efficient it is to locate that page. For an information-seeking target approximated by a key access sequence $u_i = <p_{i,1}, p_{i,2}, ..., p_{i,m}>$, given that seeking for u_i starts from page $p_s \neq p_{i,1}$, the *efficiency* $Q(u_i | p_s)$ of achieving the information-seeking target can be measured as $d(p_s, p_{i,1}) + \Sigma_{j=2}^m d(p_{i,j-1}, p_{i,j})$, where $d(x, y)$ denotes the distance from page x to page y. By normalizing the efficiency metric onto $[0, 1]$, we obtain

$$Q(u_i | p_s) = \frac{m\gamma - \min\left(d(p_s, p_{i,1}) + \sum_{j=2}^m d(p_{i,j-1}, p_{i,j}), m\gamma\right)}{m(\gamma - 1)} \quad \text{if } p_s \neq p_{i,1}, \tag{12}$$

where m is the number of content pages in u_i; the function $\min(x, y)$ returns the smaller value between x and y, and $\gamma > 1$ is a constant. A page is considered most efficient to locate if it is one click away; it is least efficient to locate if it is γ or more clicks away. We set γ to an appropriate value such that the probability of surfing γ or more clicks (i.e., $G(\gamma)$) becomes trivial. Similarly,

$$Q(u_i | p_s) = \frac{(m-1)\gamma - \min\left(\sum_{j=2}^m d(p_{i,j-1}, p_{i,j}), (m-1)\gamma\right)}{(m-1)(\gamma - 1)} \quad \text{if } p_s = p_{i,1}. \tag{13}$$

We can then derive $Q(u_i)$ as follows:

$$Q(u_i) = \sum_{\forall p_s} P\left(\text{start of seeking for } u_i = p_s\right) Q(u_i | p_s). \tag{14}$$

In turn, we can measure the efficiency $Q(U)$ of a Web site as the weighted efficiency of locating each information-seeking target in U on the Web site, that is,

$$Q(U) = \sum_{i=1}^n w(u_i) Q(u_i). \tag{15}$$

The term $Q(U)$ indicates the efficiency of locating an information-seeking target on a scale of $[0, 1]$, with 0 being the least efficient (i.e., average distance to the visitor-sought content pages is γ or more clicks away) and 1 being the most efficient (i.e., all visitor-sought content pages are only one click away). The higher the value of $Q(U)$, the more efficient it is for a visitor to locate the target information on a Web site.

Directness

Visitors are likely to find target information with fewer clicks if we add more hyperlinks pointing to content pages on each page. At an extreme, efficiency $Q(U)$ becomes 1 when each page has hyperlinks pointing to all content pages on the site, that is, all content pages are only one click away from any page, which obviously is not a good design. Placing more hyperlinks on a page makes it increasingly difficult for visitors to decide on their next move. Given an information-seeking target approximated by a key access sequence $u_i = <p_{i,1}, p_{i,2}, ..., p_{i,m}>$ and that seeking for u_i starts from $p_s \neq p_{i,1}$, directness $L(u_i|p_s)$ can be measured as $N(p_s, p_{i,1}) + \Sigma_{j=2}^{m} N(p_{i,j-1}, p_{i,j})$, where $N(x, y)$ denotes the average number of hyperlinks on the pages located on the shortest path from x to y (identified in the hyperlink structure representation step) and $N(x, y)$ is ∞ if there is no path from x to y. By normalizing the directness measure onto [0, 1], we obtain

$$L\left(u_i \mid p_s\right) = \frac{m\delta - \min\left(N\left(p_s, p_{i,1}\right) + \sum_{j=2}^{m} N\left(p_{i,j-1}, p_{i,j}\right), m\delta \right)}{m\left(\delta - 1\right)} \quad \text{if } p_s \neq p_{i,1}, \tag{16}$$

where the function $\min(x, y)$ returns the smaller value between x and y and δ is a constant, $\delta > 1$. Visitors have less difficulty deciding on their next move if the current page contains only one hyperlink but more difficulty if the current page contains δ or more hyperlinks. The value of δ can be user specified or set to an adequate constant, according to a generally accepted usability guideline (e.g., [83]). Similarly,

$$L\left(u_i \mid p_s\right) = \frac{(m-1)\delta - \min\left(\sum_{j=2}^{m} N\left(p_{i,j-1}, p_{i,j}\right), (m-1)\delta \right)}{(m-1)(\delta - 1)} \quad \text{if } p_s = p_{i,1}. \tag{17}$$

We can derive $L(u_i)$ as

$$L\left(u_i\right) = \sum_{\forall p_s} P\left(\text{start of seeking for } u_i = p_s\right) L\left(u_i \mid p_s\right). \tag{18}$$

We can then calculate the directness $L(U)$ of a Web site as the weighted directness of achieving each information-seeking target in U on the site:

$$L(U) = \sum_{i=1}^{n} w\left(u_i\right) L\left(u_i\right). \tag{19}$$

Directness, $L(U)$, is within [0, 1] and indicates the degree of ease of deciding on the next navigation move: 0 indicates the most difficulty, and 1 indicates the least. The higher the value of $L(U)$, the easier it is for a visitor to decide on the next move.

Our proposed metrics are distinct but related. For example, although efficiency and directness measure different fundamental aspects of navigability, they can be correlated because visitors' cognitive load may increase as they click more to find information on a Web site. Figure 3 summarizes the procedure for using access patterns and the
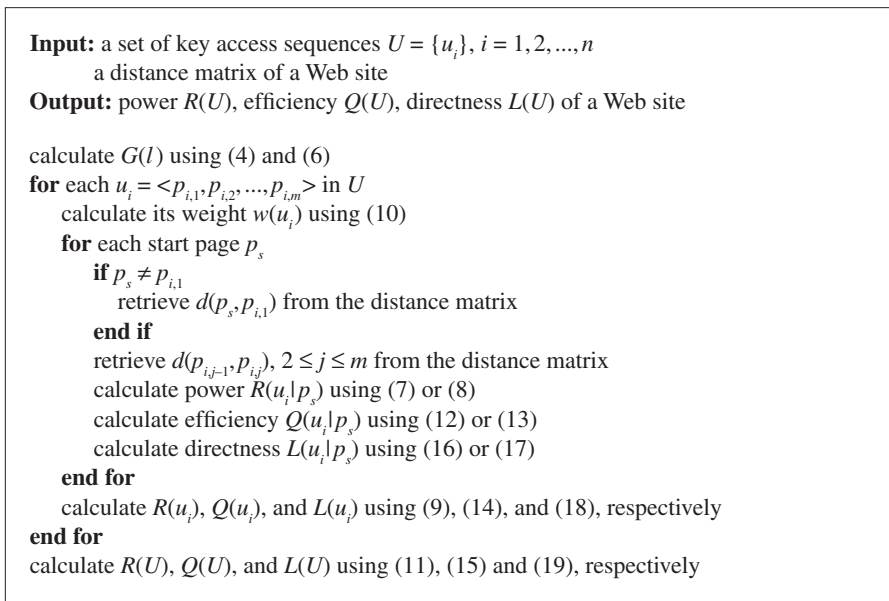
Input: a set of key access sequences $U = \{u_i\}$, $i = 1, 2, ..., n$
        a distance matrix of a Web site
Output: power $R(U)$, efficiency $Q(U)$, directness $L(U)$ of a Web site

calculate $G(l)$ using (4) and (6)
for each $u_i = <p_{i,1}, p_{i,2}, ..., p_{i,m}>$ in U
    calculate its weight $w(u_i)$ using (10)
    for each start page p_s
        if $p_s \neq p_{i,1}$
            retrieve $d(p_s, p_{i,1})$ from the distance matrix
        end if
        retrieve $d(p_{i,j-1}, p_{i,j})$, $2 \leq j \leq m$ from the distance matrix
        calculate power $R(u_i|p_s)$ using (7) or (8)
        calculate efficiency $Q(u_i|p_s)$ using (12) or (13)
        calculate directness $L(u_i|p_s)$ using (16) or (17)
    end for
    calculate $R(u_i)$, $Q(u_i)$, and $L(u_i)$ using (9), (14), and (18), respectively
end for
calculate $R(U)$, $Q(U)$, and $L(U)$ using (11), (15) and (19), respectively

Figure 3. Measuring a Web Site's Power, Efficiency, and Directness

distance matrix of a Web site as inputs to produce power, efficiency, and directness scores that jointly convey the site's navigability.

Implementation and Illustrations

To demonstrate the viability of our method and metrics, we developed a prototype system that follows the method described in Figure 1. We employed SpidersRUs [14] to parse a Web site (step 2 in Figure 1).[7] We used the prototype system to analyze and compare the navigability of two large sites in the higher education domain: Site A and Site B.[8] The sites are built and administered by two comparable public research universities in the United States (Site A university and Site B university hereinafter) that offer comprehensive degree programs at graduate and undergraduate levels and have approximately 30,000 faculty, staff, and students. These two sites serve similar user populations and have a comparable number of pages (i.e., Site A has 4,277 pages: 3,840 content pages and 437 index pages; Site B has 4,118 pages: 3,738 content pages and 380 index pages). Overall, we chose these Web sites primarily because they are from the same domain, are comparable in size, serve similar purposes and user populations, and have highly similar content and traffic, which can reduce the threat of potential confounding factors pertaining to Web site domain, size, objectives, or user groups.

   From each site, we collected Web logs generated over a four-week window. The logs from Site A contained 35,966,494 records, from which we preprocessed and identified 732,321 sessions. The logs from Site B consisted of 32,170,062 records,

Table 1. Key Parameter Values Used for Each Investigated Web Site

	Site A	Site B
α	2.68	2.99
β	10.97	10.40
γ	10	10
δ	100	100

from which we identified 555,299 sessions. We first derived the actual distribution of session size (i.e., the number of clicks per session) from the preprocessed logs to estimate the parameters α and β in Equation (4). By applying nonlinear regression to the actual distribution of session size, we obtained least square estimates for α (= 2.68) and β (= 10.97) for Site A. The fit between Equation (4) and the actual distribution of session size was statistically significant, $p < 0.0001$ and $R^2 = 0.92$. For Site B, the estimated α and β were 2.99 and 10.40, respectively. The fit between Equation (4) and the actual distribution of session size was also statistically significant, $p < 0.0001$ and $R^2 = 0.97$. We then used the α and β estimates to calculate $G(\cdot)$, according to Equations (4) and (6). For both sites, we set γ in Equation (12) to 10 because $G(10)$, or the probability of accessing pages 10 or more clicks away, seems trivial (i.e., < 0.005). Consistent with Google's Webmaster guidelines [83], we set δ in Equation (16) to 100 for both sites. Table 1 summarizes the specific parameter values used to assess the navigability of each site.

After determining these parameter values, we used the prototype system to assess each site's navigability. When mining key access sequences from Web logs, we initially set the threshold at 0.05 percent because a large number of sessions in Web logs favor small thresholds to keep a reasonable number of discovered sequences. To ensure the reliability and robustness of our evaluation results, we conducted a series of trials at threshold values between 0.05 percent and 0.175 percent, in increments of 0.025 percent. Table 2 summarizes the metric scores calculated for each site.

According to our metrics, the navigability of Site A seemed better than that of Site B, across the range of threshold values. On average, Site A recorded higher scores in power, efficiency, and directness, showing 19.6 percent, 12.8 percent, and 11.7 percent differentials, respectively. To provide a proper anchor for interpreting the between-site differences, we identified an additional 18 comparable university sites and calculated the range of each metric's value across Site A, Site B, and these additional sites with a threshold value of 0.05 percent. According to our analysis, power ranged between 0.50 and 0.77 (average = 0.71, standard deviation = 0.06), efficiency between 0.75 and 0.89 (average = 0.85, standard deviation = 0.03), and directness between 0.20 and 0.60 (average = 0.40, standard deviation = 0.12). In this light, the difference between Site A and Site B in each metric seems substantial, that is, at least 15 percent of the maximum difference across the 20 sites. For example, across the 20 sites, the maximum difference of power is 0.27 (i.e., 0.77–0.50). The difference in power between the two sites we studied is 0.12 (i.e., 0.75–0.63), which is 44 percent of the maximum difference.[9]

Table 2. Evaluation Results for Navigability (Site A Versus Site B)

Threshold value	Power		Efficiency		Directness	
	Site A	Site B	Site A	Site B	Site A	Site B
0.05%	0.75	0.63	0.87	0.77	0.42	0.36
0.075%	0.77	0.64	0.88	0.77	0.42	0.37
0.1%	0.78	0.65	0.88	0.78	0.41	0.37
0.125%	0.78	0.66	0.89	0.79	0.41	0.37
0.15%	0.79	0.66	0.89	0.79	0.42	0.38
0.175%	0.80	0.66	0.89	0.79	0.42	0.38

We evaluated power and efficiency according to the distance between the starting page of a session and the first content page sought in a key access sequence and then the distance between successively sought content pages in that sequence. In general, the greater these distances, the lower are the power and efficiency. Our analysis showed that these distances were consistently greater on Site B than on Site A. For example, on average, a content page appearing in the top 10 most frequently visited key access sequences was 1.7 clicks away on Site A, whereas a similar content page on Site B was 2.3 clicks away. Both sites had low directness scores and the directness difference between Site A and Site B was smaller than that of power or efficiency. The low directness scores could be partly attributed the fact that each site contains many content pages that inevitably increase the likelihood of having many hyperlinks on a page; as a result, it is difficult for visitors to decide on their next move from the currently visited page. Moreover, both sites feature frequently visited index pages, each of which contains many hyperlinks. While serving as springboards for navigation, these index pages also make it difficult for visitors to decide on which hyperlink to click next.

To empirically examine the navigability analysis results revealed by our metrics, we collected user performance and assessments by conducting an evaluation study. According to our metrics, Site A has higher navigability than Site B; accordingly, we anticipate that people will be more likely to find information successfully with fewer clicks and less cognitive processing on Site A than on Site B. Furthermore, as an ex post facto comparison, we used the collected user performance and assessment data to further assess our metrics versus prevalent navigability metrics such as that developed by Zhou et al. [94]. In the following, we describe our evaluation study design and data collection.

Evaluation Study and Data Collection

IN OUR EVALUATION STUDY, WE ASKED VOLUNTARY PARTICIPANTS to complete 12 information-seeking tasks on an investigated site. Our overall objective was to assess whether visitors are more likely to find target information successfully, efficiently, and easily

on a Web site of high navigability than on one of low navigability, as revealed by our metrics. In this section, we describe our study design, tasks, participants, measurements, and data collection.

Study Design

Despite their similarity in purpose, populations served, content, and traffic, Site A has considerably higher navigability than Site B. We recruited participants from both universities and asked them to complete specific information-seeking tasks on the respective sites. Each participant used one of the investigated sites to complete the tasks, and the Web site assignment was random, that is, a participant might or might not use his or her own university's site.

Our evaluation also addressed the significance of user familiarity. In general, users' familiarity with a site can influence their performance and satisfaction. For example, people highly familiar with a Web site understand its structural design and content layout; they are more likely to find information on the site than on an unfamiliar site and are less likely to become confused or "lost" [29]. High familiarity enables visitors to minimize the number of clicks necessary to find information [29], partly because of reduced trial-and-error behaviors [23]. According to Galletta et al. [28], people can complete more information-seeking tasks on a familiar site than otherwise. Visitors usually choose a surfing path because they believe it will lead them to target information; familiarity with the site's structure facilitates their choice making [7]. The performance improvement resulting from users' familiarity with a Web site aligns with the power law of practice [38], which suggests search effectiveness and efficiency gains through repeated visits, which reduce information-seeking difficulty and complexity. When visiting a familiar site, people can find information with less cognitive attention [8] and thus exhibit greater satisfaction [56]. As Galletta et al. [29] note, when navigability is problematic, its adverse influences on people's navigation can be mitigated by user familiarity with the site. Furthermore, the negative effect of low familiarity can be reduced when the site offers high navigability, that is, people are more likely to find information on an unfamiliar site when the site is more navigable. Our analysis showed that participants in general were more familiar with their university's site than with the other site. With navigability (i.e., high versus low) and familiarity (i.e., high versus low), we created four experimental conditions and assigned participants to each randomly, while mindfully balancing the total number of participants in each group.

Tasks

We analyzed each site's Web logs to identify frequently sought content page sequences (i.e., key access sequences). We conducted a pretest with 256 participants from Site A university and 165 participants from Site B university. All the participants reviewed 20 pages from their own university site (10 content and 10 index pages), randomly

selected from the pages on the site, and indicated whether the presented page provided information they searched for frequently, on a five-point Likert scale (1 = "not at all," 5 = "very high"). The pretest participants from both universities were comparable in age, number of years at the current university, gender composition, and self-reported familiarity with their own university site. According to the paired sample t-test results for Site A, content pages were more likely to constitute information-seeking targets than index pages (3.48 versus 1.90, $t = 31.15$, df [degrees of freedom] = 255, $p < 0.001$); we noted similar results for Site B (2.99 versus 1.91, $t = 17.07$, df = 164, $p < 0.001$). These participants then received 10 key access sequences of content pages and 10 nonkey access sequences of content pages (i.e., sequences of low visit frequency) from their own university site, all randomly selected from our access pattern mining results. For each sequence, a participant specified his or her need, desire, or interest in seeking the pages in that sequence, on a five-point Likert scale (1 = "no need, desire, or interest at all," 5 = "great need, desire, or interest"). According to our analysis, for both universities, the need, desire, or interest was significantly higher for key than for nonkey access sequences: Site A: 3.48 versus 1.30, $t = 59.23$, df = 255, $p < 0.001$; and Site B: 2.99 versus 1.33, $t = 37.24$, df = 164, $p < 0.001$. According to our pretest results, content pages were more likely to constitute information-seeking targets than index pages, and key access sequences identified from Web logs were consistent with users' common information-seeking needs, desires, or interests.

Next, we verified the frequently sought key access sequences by surveying another random sample of 20 participants from each university. Each reviewed 20 frequently sought key access sequences we discovered from the participant's own university's Web logs and specified the frequency with which he or she would access each sequence, on a five-point Likert scale (1 = "never" and 5 = "once or several times a month, or more frequently"). According to our analysis, participants from both universities frequently sought 15 key access sequences, that is, students shared some similarity in their information needs and interests (e.g., the operating hours of the campus medical center).[10] In our evaluation study, we used 3 of these 15 key access sequences as warm-up exercises and the remaining as the information-seeking tasks to be performed by participants in the evaluation study. Among the 12 tasks included in our study, half were low in complexity (e.g., only one page in the access sequence) and the others high in complexity (e.g., multiple pages in the sequence). Appendix A lists all the tasks.

Participants

We recruited participants among the business undergraduate students enrolled in similar information systems or operations classes in both universities.[11] Participation was voluntary, and each participant received $10 for his or her time and efforts. To solicit best efforts, we offered substantial, additional monetary incentives to top performers in the study, that is, those who successfully completed the greatest number of tasks in the least amount of time. Instructors assisted in our recruitment, and a student's decision to participate had no effect on his or her class grade.

Measurements

We examined user performance by focusing on effectiveness and efficiency. Specifically, we used three measures: task success rate, task time, and the number of clicks to complete a task. We assessed effectiveness with task success rate, which is the ratio of the number of successfully completed tasks (i.e., participant finding the target page) to the total number of tasks to be performed in the study. We measured the click requirement with the exact number of clicks a participant used in a task. We also measured the amount of time (in seconds) a participant spent on a task, which closely relates to the number of clicks and offers another perspective on user performance efficiency. Participants had up to four minutes to complete each task.[12] We collected participants' self-reports about the cognitive-processing load required, after they had completed all the tasks, which indicates the cognitive processing they underwent when choosing paths to access the target information. We used five items adapted from Hong et al. [32] and Palmer [59] to measure cognitive-processing load, with minor wording changes appropriate for our context; all the items employed a seven-point Likert scale (1 = "strongly disagree," 7 = "strongly agree"). Appendix B lists the measurement items for cognitive-processing load. To various extents, these measurements corresponded with our metrics: task success rate for power, the number of clicks for efficiency, and cognitive-processing load for directness.

Data Collection

We conducted the evaluation study in multiple sessions, all administered in a designated computer laboratory at each university.[13] Before each session, we read a script to inform the participants of the study's purpose, answered questions, and addressed any privacy-related concerns. We explicitly stated our intent and commitment to performing data analyses at aggregate levels without using any personally identifiable information. We obtained written consent from each participant and promised convenient access to the data he or she provided in the study. We then collected some demographic data from the participants, described the overall study flow and the tasks to be performed, and provided warm-up exercises until each participant signaled readiness for the study. Then, the participants received packets that detailed each task, together with target page screenshots they could use to verify whether they had completed that task. We used client-side monitoring software to record the exact starting and ending time of each task, the number of clicks, and the specific pages accessed. By comparing the recorded pages a participant accessed with the target page(s), we determined whether he or she had successfully completed a task. After completing all the tasks, the participants filled out a questionnaire survey that gathered their cognitive processing.

Data Analyses and Results

WE PERFORMED A PILOT STUDY TO REEXAMINE THE ITEMS for cognitive-processing load and to fine-tune our study flow and data collection procedure. In total, 39 undergraduate

Table 3. Comparative Analysis of Participants from the Studied Universities

Dimension	Participants from Site A university	Participants from Site B university
Average age	24.79; range: 19–52	23.53; range: 16–60
Gender	Male: 90 (70.3%)	Male: 72 (60%)
	Female: 38 (29.7%)	Female: 48 (40%)
Major	Business: 127 (99.2%)	Business: 110 (91.7%)
	Not declared: 1 (0.8%)	Not declared: 10 (8.3%)
Status in university	Freshman: 43 (33.6%)	Freshman: 10 (8.3%)
	Sophomore: 31 (24.2%)	Sophomore: 28 (23.3%)
	Junior: 29 (22.7%)	Junior: 26 (21.7%)
	Senior: 25 (19.5%)	Senior: 56 (46.7%)
General computer efficacy	5.72 (out of 7)	5.85 (out of 7)
Internet technology competence	5.3 (out of 7)	5.5 (out of 7)
Web browsing capability	5.4 (out of 7)	5.8 (out of 7)
Familiarity with studied Web site	3 (out of 7)	3 (out of 7)
Internet usage	14.8 hours a week	18.2 hours a week

students—17 from Site A university and 22 from Site B university—enrolled in a required information systems or operations management class took part in the pilot study voluntarily. Each participant followed the described study flow and completed the 12 information-seeking tasks. According to our analysis, these participants understood the tasks clearly and knew exactly what they needed to do in each task; more than 99 percent were able to complete a task within three minutes. We used their responses to assess the reliability of the items for cognitive-processing load. The Cronbach's alpha was 0.95 for cognitive-processing load, exceeding the common threshold of 0.7 [58].

We then conducted an evaluation study with 248 participants (128 from Site A university and 120 from Site B university). As Table 3 summarizes, the participants were comparable in terms of age and gender composition; we noted no significant between-group differences in general computer efficacy, Internet usage, or familiarity with the randomly assigned site they used in the study. In addition, participants reported higher familiarity with their university's site than with the other site, and the difference was statistically significant.[14] We reexamined the reliability of our cognitive-processing measurements and noted a Cronbach's alpha greater than 0.90, suggesting adequate reliability.

To compare user performance and assessments between the two Web sites, we aggregated the data associated with each site, across participants and tasks, and used them to perform a series of paired t-tests. As Table 4 shows, the task success rate associated with Site A was significantly higher than that of Site B (0.90 versus 0.78, $p < 0.001$). The time requirements associated with Site A were significantly lower than those of Site B (40.25 versus 65.70, $p < 0.001$), as was the number of clicks (3.69 versus 6.08,

Table 4. User Performance Comparisons: Site A Versus Site B

Measurement	Site	N	Mean	Standard deviation	t-value	p-value	Difference
Task success rate	A	121	0.90	0.10	7.37	< 0.001	0.12
	B	127	0.78	0.15			
Time requirements	A	121	40.25	20.29	−9.67	< 0.001	−25.45
	B	127	65.70	21.11			
Number of clicks	A	121	3.69	1.40	−11.54	< 0.001	−2.39
	B	127	6.08	1.84			
Cognitive-processing load	A	121	2.83	1.06	−7.20	< 0.001	−1.10
	B	127	3.94	1.34			

Note: 121 participants used Site A (61 from Site A university and 60 from Site B university); 127 participants used Site B (67 from Site A university and 60 from Site B university).

$p < 0.001$) and the cognitive-processing load (2.83 versus 3.94, $p < 0.001$).[15] Overall, our results showed that participants were more likely to find information successfully, efficiently, and easily on Site A than on Site B, which is congruent with the analysis results revealed by our metrics.

We further compared user performance between the two sites by separating tasks germane to high versus low complexity. We then combined the data associated with tasks from each site to create two data sets: one pertaining to low-complexity tasks and the other to high-complexity tasks. We used each data set to further compare user performance between sites by performing paired t-tests. For low-complexity tasks, user performance seemed better on Site A than on Site B, and the differences were statistically significant: task success rate (0.97 versus 0.79, $p < 0.001$), time requirements (27.35 versus 36.61, $p < 0.001$), and number of clicks (2.35 versus 3.15, $p < 0.001$). We found similar results for high-complexity tasks, and the differentials appeared greater in magnitude: task success rate (0.83 versus 0.67, $p < 0.001$), time requirements (53.14 versus 94.79, $p < 0.001$), and number of clicks (5.03 versus 9.01, $p < 0.001$). Table 5 summarizes the comparative user performance results by task complexity, which does not include the cognitive-processing load because we gathered participants' assessments of cognitive processing after they completed all the tasks. Our results showed that participants were more likely to find information successfully and efficiently on Site A than on Site B, regardless of task complexity.

We also compared user performance between the two sites by taking into account participants' familiarity with the site they used, that is, their own university's site or the other site. We combined the data from each site that were generated by the participants familiar with the site versus those produced by their unfamiliar counterparts, thereby creating two data sets: one germane to high familiarity and the other to low familiarity. We used each data set to perform paired t-tests that enabled us to understand user performance between the two sites when participants were familiar versus not familiar with the site. When participants were familiar with the site, user performance was better on Site A than on Site B, and the differences were statistically significant: task success rate (0.93 versus 0.80, $p < 0.001$), time requirements (29.76 versus 56.06, $p < 0.001$), number of clicks (3.39 versus 5.76, $p < 0.001$), and cognitive-processing load (2.62 versus 3.48, $p < 0.001$). We found similar results when participants were not familiar with the site they used: task success rate (0.87 versus 0.77, $p < 0.001$), time requirements (50.90 versus 74.33, $p < 0.001$), number of clicks (3.99 versus 6.36, $p < 0.001$), and cognitive-processing load (3.05 versus 4.35, $p < 0.001$). Table 6 summarizes our comparative results in terms of participants' familiarity with the site, which shows that participants were more likely to find information successfully, efficiently, and easily on Site A than on Site B, regardless of whether they were familiar with the site.

We examined the performance of the participants from each university who used different Web sites; specifically, we aggregated the participants by university to create two data sets: participants from Site A university versus those from Site B university. In each data set, approximately half the participants used their own university's site, and the others used the other site. For example, among the 120 participants we recruited from Site B university, 60 used their own university's site in the study. With these data,

Table 5. User Performance Comparisons in Low- Versus High-Complexity Tasks

Task complexity	Measurement	Site	N	Mean	Standard deviation	t-value	p-value	Difference
Low	Task success rate	A	121	0.97	0.07	7.50	< 0.001	0.17
		B	127	0.79	0.25			
	Time requirements	A	121	27.35	15.31	−4.03	< 0.001	−9.26
		B	127	36.61	20.56			
	Number of clicks	A	121	2.35	1.31	−4.13	< 0.001	−0.80
		B	127	3.15	1.71			
High	Task success rate	A	121	0.83	0.18	5.70	< 0.001	0.17
		B	127	0.67	0.27			
	Time requirements	A	121	53.14	30.75	−10.34	< 0.001	−41.65
		B	127	94.79	32.60			
	Number of clicks	A	121	5.03	2.24	−11.90	< 0.001	−3.98
		B	127	9.01	2.99			

Table 6. User Performance Comparison When Users Have High Versus Low Familiarity

Familiarity	Measurement	Site	N	Mean	Standard deviation	t-value	p-value	Difference
High	Task success rate	A	61	0.93	0.09	6.68	< 0.001	0.13
		B	60	0.80	0.13			
	Time requirements	A	61	29.76	12.21	−9.78	< 0.001	−26.30
		B	60	56.06	16.96			
	Number of clicks	A	61	3.39	1.17	−8.81	< 0.001	−2.37
		B	60	5.76	1.73			
	Cognitive-	A	61	2.62	1.09	−3.89	< 0.001	−0.85
	processing load	B	60	3.48	1.31			
Low	Task success rate	A	60	0.87	0.10	4.13	< 0.001	0.10
		B	67	0.77	0.16			
	Time requirements	A	60	50.90	21.37	−6.25	< 0.001	−23.43
		B	67	74.33	20.82			
	Number of clicks	A	60	3.99	1.55	−7.65	< 0.001	−2.37
		B	67	6.36	1.90			
	Cognitive-	A	60	3.05	1.00	−6.48	< 0.001	−1.30
	processing load	B	67	4.35	1.24			

we performed paired t-tests to examine user performance across the universities. For participants recruited from Site A university, we noted better user performance when they used Site A rather than Site B, with statistically significant differences: task success rate (0.93 versus 0.77, $p < 0.001$), time requirements (29.76 versus 74.33, $p < 0.001$), number of clicks (3.39 versus 6.36, $p < 0.001$), and cognitive-processing load (2.62 versus 4.35, $p < 0.001$). We found similar results for participants recruited from Site B university, though the differences were smaller in magnitude, with the exception of time requirements (50.90 versus 56.06, $p > 0.05$): task success rate (0.87 versus 0.80, $p < 0.001$), number of clicks (3.99 versus 5.76, $p < 0.001$), and cognitive-processing load (3.05 versus 3.48, $p < 0.05$). Table 7 summarizes the comparative results by grouping participants by university; as shown, our overall results imply that visitors' familiarity with a Web site cannot compensate for its low navigability.

In addition to empirically testing the comparative analysis results revealed by our metrics, we used the experimental data to perform an ex post facto comparison with a Markov model–based navigability measure (MNav), which has been shown to be more effective than many existing navigability measures [94]. This measure is developed with Web structure data and a Markov model that emulates users' browsing behaviors. Specifically, at time $t = 0, 1, 2, ...$, a visitor's current position is represented by a vector $p(t) = (p_1(t), p_2(t), ..., p_{n-1}(t), p_n(t))$, where $p_i(t)$, $i = 1, 2, ..., n – 1$, denotes the probability at page i, and $p_n(t)$ is the probability at the virtual stopping state, which indicates the termination of browsing on a site. Let P be the matrix of transition probabilities; each element of P represents the probability of transiting from one page to another. According to Zhou et al. [94], $p(t)$ is calculated as

$$p(t) = p(t – 1)P = p(0)P^t. \tag{20}$$

Equation (20) converges at time T; the MNav score of a site then can be calculated as the probability not in the virtual stopping state, that is, $1 – p_n(T)$. The greater the MNav score of a site, the higher is the site's navigability [94].

We used Equation (20) to calculate the respective MNav scores of Site A and Site B, following the procedure and parameter settings by Zhou et al. [94]. We obtained an MNav score of 0.66 for Site A and 0.67 for Site B, suggesting comparable navigability between the two sites, although Site B has a slightly higher navigability. This result contradicts the comparative analysis revealed by our metrics and is not supported by our user evaluation data that show that Site A has higher navigability than Site B. The comparative analysis by the MNav measure may have several limitations. First, this measure does not consider Web usage and content data in its calculation. To derive the transition probability matrix P, the MNav measure instead assumes that each hyperlink on a page has equal probability of being clicked. This assumption may not always hold, because a hyperlink a visitor will click on may depend on his or her information-seeking target as well as his or her estimate of the hyperlink's likelihood of leading to the target. As a result, some hyperlinks could be clicked more frequently than others rather than equally. By analyzing visitors' browsing behaviors recorded in Site A's Web logs, we noted that, on average, 10 percent of hyperlinks on a page drew nearly 60 percent of the clicks on that page. Taking the most frequently visited page of

Table 7. User Performance Comparison: Participants from Site A Versus Site B University

University	Measurement	Site	N	Mean	Standard deviation	t-value	p-value	Difference
Site A university	Task success rate	A	61	0.93	0.09	6.86	< 0.001	0.16
		B	67	0.77	0.16			
	Time requirements	A	61	29.76	12.21	−14.93	< 0.001	−44.57
		B	67	74.33	20.82			
	Number of clicks	A	61	3.39	1.17	−10.72	< 0.001	−2.97
		B	67	6.36	1.90			
	Cognitive-processing load	A	61	2.62	1.09	−8.35	< 0.001	−1.73
		B	67	4.35	1.24			
Site B university	Task success rate	A	60	0.87	0.10	3.58	< 0.001	0.08
		B	60	0.80	0.13			
	Time requirements	A	60	50.90	21.37	−1.46	0.150	−5.16
		B	60	56.06	16.96			
	Number of clicks	A	60	3.99	1.55	−5.92	< 0.001	−1.78
		B	60	5.76	1.73			
	Cognitive-processing load	A	60	3.05	1.00	−2.01	0.047	−0.43
		B	60	3.48	1.31			

Site A as an example, among 120 hyperlinks on this page, the most frequently clicked hyperlink attracted 19 percent of the total clicks. We also analyzed Site B's logs and noted a similarly skewed distribution of hyperlink clicks. Second, the MNav measure assumes that all the visits start from the home page and accordingly sets the vector $p(0)$. By examining the logs of Site A and Site B, we found that more than 20 percent of visit sessions started from pages other than the homepage. Furthermore, prior Web usage analyses do not support this assumption [42]. These questionable assumptions can constrain the use of $p(0)$ and P to represent Web surfing behaviors, which affects the accuracy of $p(t)$ as calculated by Equation (20) and thus hinders the effectiveness of the MNav measure for revealing a Web site's navigability. In contrast, our metrics consider Web structure and content data and are calculated with actual Web surfing behaviors recorded in Web logs, rather than making assumptions about surfing behaviors. Thus, the comparative analysis revealed by our metrics is congruent with the actual user performance and assessments observed in the evaluation study.

We further examined our metrics with additional sites beyond the education domain. Specifically, we identified from Website Review (www.websitereview.net), which provides user evaluations of navigation on different sites, three pairs of Web sites for shopping, sports, and recreation. In each pair of sites, one is high in navigation and the other low in navigation, according to the user evaluations.[16] We examined these sites to test whether the navigability results revealed by our metrics are consistent with the user evaluations of navigation. For each site, we used our method to extract its content and structure; we simulated Web logs by following the method of Liu et al. [46] because the usage data (i.e., Web logs) are not publicly available. This method, which simulates the random walk of information foraging agents (e.g., visitors) on a Web site to generate artificial logs for the site, has been shown to be capable of generating synthetic Web logs that demonstrate the regularities observed in actual Web logs, such as the law of surfing [46]. By applying our method to a site's content, structure, and simulated logs, we then calculated its scores for power, efficiency, and directness. Table 8 shows the power, efficiency, and directness scores of each site, in conjunction with user navigation evaluation of that site. By grouping the sites by domain (e.g., Shopping-1, Shopping-2), we note that the navigability revealed by our metrics is consistent with the user evaluations across all three pairs of sites that we analyzed. We also compared our metrics with the MNav measure by Zhou et al. [94] in relation to the user evaluations. According to the MNav measure, Shopping-1 has navigability comparable to that of Shopping-2, Sports-1 is more navigable than Sports-2, and Recreation-1 is more navigable than Recreating-2. These results are not consistent with the user evaluations. Overall, our analyses of these sites show that the navigability revealed by our metrics is consistent with user evaluation results available at Website Review, while the navigability analyses by the MNav measure contradict the user evaluations results. Last, we benchmarked our metrics against the measure by Yen [90] in Appendix C.

Extensions to Proposed Metrics

Our metrics can be extended in several ways. For example, our metrics could be extended to compare the navigability of Web sites that vary in scale (i.e., different

Table 8. Comparison of Proposed Metrics and MNav with User Evaluations of Navigation

	Shopping-1	Shopping-2	Sports-1	Sports-2	Recreation-1	Recreation-2
Power	0.74	0.79	0.36	0.61	0.73	0.82
Efficiency	0.86	0.88	0.73	0.83	0.86	0.88
Directness	0.16	0.45	0.32	0.51	0.36	0.53
User evaluations	Low	High	Low	High	Low	High
MNav	0.71	0.71	0.74	0.56	0.72	0.69

numbers of pages) or domain (e.g., commerce versus education). In general, it is more difficult to find information on a larger Web site than on a smaller one. In this case, we could introduce a scale factor to account for the difference in scale, as measured by the number of pages on a Web site. For example, consider two sites: site one with n_1 pages and site two with n_2 pages; we can define the scale factor as n_1/n_2. When comparing the navigability of the sites, we multiply each metric score of site one by the scale factor to accommodate for the scale difference. Similar accommodations may be needed for cross-domain comparisons because the scale may vary with domain, for example, a university site could have more pages than an e-government site. Visitors' browsing behaviors may also differ across domains. Thus, the navigability of a Web site, as suggested by our metrics, should be interpreted in terms of how well its hyperlink structure facilitates visiting behaviors specific to that domain.

In addition, our metrics can be extended with the combined use of search engine. Toward that end, several factors should be properly considered. For example, we need to estimate the probability of a visitor's use of search engine versus browsing the hyperlink structure to find information. Such probabilities could be estimated by analyzing Web logs. The effectiveness of the search engine is relevant as well (e.g., the probability that the links returned by a search engine point to the target information). Several measures commonly used in information retrieval (e.g., recall, precision) offer logical ways to assess a search engine's effectiveness [66]. The remaining distance, from the results returned by a search engine to the target information, also needs to be estimated. For example, the search results returned by a search engine could be displayed over multiple pages; in this case, if a hyperlink that points to the target page appears on the kth page of the results, the target page is then k clicks away, after using the search engine. We therefore should evaluate the average distance traversed to a target page after using a search engine.

Finally, we could integrate power, efficiency, and directness for a single, holistic measure. Specifically, we calculate *overall navigability* with the harmonic mean, a common approach of combining multiple measures to produce a holistic measure [22, 36, 66]. For power $R(U) > 0$, efficiency $Q(U) > 0$, and directness $L(U) > 0$, the harmonic mean of these metrics is

$$\frac{3}{\dfrac{1}{R(U)}+\dfrac{1}{Q(U)}+\dfrac{1}{L(U)}} = \frac{3R(U)Q(U)L(U)}{Q(U)L(U)+R(U)L(U)+R(U)Q(U)}.$$

Overall navigability metric $O(U)$ is 0 if $R(U) = 0$, or $Q(U) = 0$, or $L(U) = 0$. Therefore, we have

$$O(U) = \begin{cases} 0 & \text{if } R(U) = 0 \text{ or } Q(U) = 0 \text{ or } L(U) = 0 \\ \dfrac{3R(U)Q(U)L(U)}{Q(U)L(U)+R(U)L(U)+R(U)Q(U)} & \text{otherwise,} \end{cases}$$

where $O(U)$ is bounded within [0, 1]. In addition, $O(U) = 0$ if $R(U) = 0$, or $Q(U) = 0$, or $L(U) = 0$; $O(U) = 1$ if $R(U) = Q(U) = L(U) = 1$. Apparently, the greater the value of $O(U)$, the better is a site's overall navigability.

Discussion

OUR STUDY HAS SEVERAL IMPLICATIONS FOR PRACTICE. Although information seeking is a common reason people visit Web sites, many sites remain difficult to navigate, hindering user experience and satisfaction [57]. Toward that end, we provide three data-driven metrics and a viable method for assessing a Web site's navigability. Supported by our method and metrics, organizations can evaluate and monitor their sites' navigability continually or even implement an automatic alert mechanism if the navigability, as revealed by our metrics, falls below a specified threshold. Organizations also can use our metrics to record navigability longitudinally and analyze essential patterns or emerging trends to generate insights into the conditions that yield high or low navigability. Use of our navigability metrics can augment existing Web analytical tools (e.g., Omniture, WebTrends) by providing normative or diagnostic analyses. Also, organizations modify their sites periodically for improved user performance, such as by comparing alternative designs. Using our metrics and method, Web site administrators could quickly predict the impact of each design on navigability. To do so, they could synthesize artificial Web logs using Liu et al.'s [46] method, which simulates the random walk of information foraging agents (e.g., visitors) on a site and thereby generates artificial logs. Employing our metrics and artificial logs, Web site administrators could predict the navigability of alternative designs, choose the promising designs that yield high navigability scores, and then use existing evaluation methods to evaluate the promising designs only. This capability offers benefits because time constraints often make it difficult for Web site administrators to undertake full alternative design evaluations, particularly those involving a large number of users. In addition, our metrics and method can be used prescriptively by helping organizations identify promising areas for improving their existing hyperlink structure design. For example, by sorting key access sequences by their weights and then by their navigability scores (power, efficiency, directness), organizations could identify sequences that have high weights but are low in navigability. Such sequences in turn highlight important bottlenecks for navigability, that is, frequently visited sequences (high weight) that are not effectively facilitated by the existing hyperlink structure design (low navigability); mitigating these bottlenecks could improve a Web site's navigability substantially.

Our study has several limitations that deserve future research attention. First, we targeted university Web sites, and therefore our results may not be equally applicable to sites with very different structures, such as social networking or wiki sites. Our spiders and page parsers offer limited utilities for Web 2.0 sites and interactive contents (e.g., Flash or Ajax based); extensions are needed to process such contents. The two Web sites we studied have little Web 2.0 and interactive contents; however, to apply our methods to sites rich in such contents, our spiders must be able to download interactive contents and our page parser must extract links and other features from these contents. Second, although our study is appropriate for our objectives and intended comparison, it targets a specific scenario. Understandably, Web sites may vary in their relative strengths and weaknesses; therefore, we should perform empirical evaluations in different scenarios.

For example, comparing a site high in power but low in directness with another site low in power but high in directness is essential; similarly, assessing user performance and assessments on two sites that are comparable in navigability is also critical because it allows us to demonstrate that user performances do not differ when there is no difference in navigability, according to our metrics. Third, our metrics and method should be extended by considering additional factors that affect navigation, such as navigation aids (e.g., back button) and information scent [15]. For example, the importance of a page could be estimated with anchor text or font size of the main text [10]. Links located in prominent locations on a site and with anchor text in a large font size or a sharp color may be more likely to be clicked by users. Multimedia, including images and movies, are important page properties as well and may affect the likelihood of a visitor's clicking on a link. Although our proposed metrics and method can shed light on probable problem areas of a site's hyperlink structure, formal algorithms and implementable methods are needed for prescriptive purposes. Finally, efforts are also needed to combine our data-driven metrics and salient perceptual measures for assessing Web site navigability, preferably by involving different user groups, work contexts, or domains (e.g., e-commerce, digital government, online learning).

Conclusion

THE MAIN RESEARCH CONTRIBUTION OF OUR STUDY lies in the development of three data-driven metrics for measuring Web site navigability. We use established theories to guide our choices of the fundamental navigability dimensions to emphasize and propose specific data-driven metrics (i.e., power, efficiency, and directness) that correspond to these dimensions respectively. Premised in the law of surfing, we formulate our metrics by considering Web structure, usage, and content data. By integrating appropriate Web mining techniques, we develop a method for calculating a Web site's navigability according to our metrics, which explicitly specifies the input Web data and their transformations and analyses. In addition, we demonstrate the viability and practical value of our metrics and method by implementing a prototype system, and we use it to assess the navigability of two sizable, real-world Web sites. We perform an evaluation study by comparing the user performance observed on the respective sites and thus produce empirical evidence suggesting that people are more likely to find information successfully, efficiently, and easily when the site has high navigability, as revealed by our metrics.

NOTES

1. A typical Web log record includes the IP (Internet Protocol) address, time, and URL (Uniform Resource Locator), which describe who accessed a page and when, the request status (i.e., success or failure), and the size of the data transmitted.

2. Other issues surrounding navigability also have been studied [11, 65, 77], such as the development of data-gathering agents for navigability evaluation [65]. Because our focus is on navigability metrics, we do not provide a review of these studies.

3. Previously accessed pages can be cached at a proxy server or a client machine; revisits to these pages therefore may not be recorded in server-side Web logs, although they are part of a session.

4. The results of our evaluation task selections in the Evaluation Study and Data Collection section also suggest that content pages are much more likely to constitute visitors' information-seeking targets than index pages and that frequently accessed sequences of content pages approximate common information-seeking targets reasonably well.

5. We constructed our classifier with the SVM (support vector mechanism)-light implementation [37].

6. We consider two scenarios of seeking for u_i: starting from any page but the first page in u_i (i.e., $p_s \neq p_{i,1}$) or starting from the first page in u_i (i.e., $p_s = p_{i,1}$).

7. To parse a site, SpidersRUs specifies the homepage URL as the seed. When accessing a page, SpidersRUs downloads the page and extracts all the hyperlinks on the page; it then downloads each of the pages pointed to by the extracted hyperlinks and extracts the hyperlinks on that page. SpidersRUs continues this process until exhausting all the pages on the site.

8. Both sites are the main sites of their respectively served universities, e.g., www.university-name.edu.

9. The difference between Site A and Site B in power, efficiency, and directness is about 2 times, 3 times, and 0.5 times of their respective standard deviation.

10. The two sites use different wordings to describe highly similar, if not identical, resources, services, or information. For example, Site A used "contact information and hours of the student health center" for information about the campus medical center's contact information and operating hours, and Site B used "contact information and operating hours of the main campus medical center."

11. Although students, faculty, staff, and external visitors could use a site, students constitute a crucial user group. They substantially outnumbered faculty and staff combined and often used the university's site to obtain information.

12. We administered this maximal time limit to keep the experiment at a reasonable duration. This time limit seemed adequate because the vast majority of pilot study participants completed each task within three minutes.

13. We conducted six experimental sessions at Site A university, which were administered by the same two investigators; seven sessions were conducted at Site B university, all administered by the same investigator. All the sessions followed the same procedure regardless of their locations or administrators.

14. We performed a manipulation check in the experiment, which indicated that participants were significantly more familiar with their university's site than with the other site; unpaired t-test results showed a significant between-group difference: 4.68 versus 1.10, $t = 28.66$, $p < 0.001$, on a seven-point Likert scale (1 = "extremely unfamiliar," 7 = "extremely familiar").

15. The time requirements and number of clicks we report do not incur any penalty for failed tasks, that is, for each task a participant failed to complete successfully, we used the exact amount of time and the number of clicks he or she took in the task. If considering any penalty, we expect greater differences in user performance between the two sites.

16. Each site is "low" or "high" in navigation, according to the user ratings on a five-point scale, with 1 being the lowest and 5 being the highest. In our study, we chose two sites from the same domain that have noticeable differences in the navigation rating.

References

1. Abbasi, A., and Chen, H. Cybergate: A system and design framework for text analysis of computer-mediated communication. *MIS Quarterly, 32,* 4 (2008), 811–837.

2. Abbasi, A.; Chen, H.; and Nunamaker, J.F. Stylometric identification in electronic markets: Scalability and robustness. *Journal of Management Information Systems, 25,* 1 (Summer 2008), 49–78.

3. Abbasi, A.; Chen, H.; and Salem, A. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems, 26,* 3 (2008), Article 12.

4. Abbasi, A.; Zhang, Z.; Zimbra, D.; Chen, H.; and Nunamaker, J.F. Detecting fake Websites: The contribution of statistical learning theory. *MIS Quarterly, 34,* 3 (2010), 435–461.

5. Abels, E.G.; White, M.D.; and Hahn, K. Identifying user-based criteria for Web pages. *Internet Research: Electronic Networking Applications and Policy, 7,* 4 (1997), 252–262.

6. Agrawal, R., and Srikant, R. Mining sequential patterns. In P.S. Yu and A.L. Chen (eds.), *Proceedings of the 11th International Conference on Data Engineering.* Los Alamitos, CA: IEEE Computer Society Press, 1995, pp. 3–14.

7. Ahmad, R.; Li, Z.; and Azam, F. Measuring "navigational burden." In D.-K. Baik, D. Primeaux, N. Ishii, and R. Lee (eds.), *Proceedings of the Fourth International Conference on Software Engineering Research, Management and Applications.* Los Alamitos, CA: IEEE Computer Society Press, 2006, pp. 307–314.

8. Bergman, O.; Beyth-Marom, R.; Nachmias, R.; Gradovitch, N.; and Whittaker, S. Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems, 26,* 4 (2008), Article 20.

9. Botafogo, R.A.; Rivlin, E.; and Shneiderman, B. Structural analysis of hypertexts: Identifying hierarchies and useful metrics. *ACM Transactions on Information Systems, 10,* 2 (1992), 142–180.

10. Brin, S., and Page, L. The anatomy of a large-scale hypertextual Web search engines. In H. Ashman and P. Thistlewaite (eds.), *Proceedings of the 7th International World Wide Web Conference.* Amsterdam: Elsevier, 1998, pp. 107–117.

11. Castro, C.; Meliá, S.; Genero, M.; Poels, G.; and Calero, C. Towards improving the navigability of Web applications: A model-driven approach. *European Journal of Information Systems, 16,* 4 (2007), 420–447.

12. Chaiken, S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 5 (1980), 752–766.

13. Chakrabarti, S.; Dom, B.E.; Kumar, S.R.; Raghavan, P.; Rajagopalan, S.; Tomkins, A.; Gibson, D.; and Kleinberg, J. Mining the Web's link structure. *IEEE Computer, 32,* 8 (1999), 60–67.

14. Chau, M.; Qin, J.; Zhou, Y.; Tseng, C.; and Chen, H. SpidersRUs: Automated development of vertical search engines in different domains and languages. In M. Marlino, T. Sumner, and F. Shipman (eds.), *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries.* New York: ACM Press, 2005, pp. 110–111.

15. Chi, E.H.; Pirolli, P.; Chen, K.; and Pitkow, J. Using information scent to model user information needs and actions and the Web. In J. Jacko and A. Sears (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* New York: ACM Press, 2001, pp. 490–497.

16. Cooley, R.; Mobasher, B.; and Srivastava, J. Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems, 1,* 1 (1999), 1–27.

17. Cyr, D. Modeling website design across cultures: Relationships to trust, satisfaction, and e-loyalty. *Journal of Management Information Systems, 24,* 4 (Spring 2008), 47–72.

18. De Bra, P., and Houben, G.J. Hypertext metrics revisited: Navigational metrics for static and adaptive link structures. Working paper, Department of Computing Science, Eindhoven University of Technology, 1997.

19. de Castro, V.; Genero, M.; Marcos, E.; and Piattini, M. Including routes in Web information systems as a way to improve the navigability: An empirical study. In B. Benatallah, F. Casati, D. Georgakopoulos, C. Bartolini, W. Sadiq, and C. Godard (eds.), *Proceedings of the Web Information Systems Engineering (WISE) 2007 Workshops.* New York: Springer, 2007, pp. 505–510.

20. Donato, D.; Laura, L.; Leonardi, S.; and Millozzi, S. The Web as a graph: How far we are. *ACM Transaction on Internet Technology, 7,* 1 (2007), 1–25.

21. Dong, G., and Pei, J. *Sequence Data Mining.* New York: Springer, 2007.

22. Duffie, D. Asset price dynamics with slow-moving capital. *Journal of Finance, 65,* 4 (2010), 1238–1268.

23. Edwards, D.M., and Hardman, L. "Lost in hyperspace": Cognitive mapping and navigation in a hypertext environment. In R. McAleese (ed.), *Hypertext: Theory into Practice,* Oxford: Intellect, 1999, pp. 90–105.

24. Fan, W.; Gordon, M.D.; and Pathak, P. Genetic programming-based discovery of ranking functions for effective Web search. *Journal of Management Information Systems, 21,* 4 (Spring 2005), 37–56.

25. Fan, W.; Luo, M.; Wang, L.; Xi, W.; and Fox, E.A. Tuning before feedback: Combining ranking function discovery and blind feedback for robust retrieval. In M. Sanderson, K. Järvelin, J. Allan, and P. Bruza (eds.), *Proceedings of the 27th Annual International ACM SIGIR Conference.* New York: ACM Press, 2004, pp. 138–145.

26. Fang, X.; Liu Sheng, O.R.; Gao, W.; and Iyer, B. A data-mining-based prefetching approach to caching for network storage systems. *INFORMS Journal on Computing, 18,* 2 (2006), 267–282.

27. Fu, W., and Pirolli, P. SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human Computer Interaction, 22,* 4 (2007), 355–412.

28. Galletta, D.F.; Henry, R.M.; McCoy, S.; and Polak, P. Web site delays: How tolerant are users? *Journal of the Association for Information Systems, 5,* 1 (2004), 1–28.

29. Galletta, D.F.; Henry, R.M., McCoy, S.; and Polak, P. When the wait isn't so bad: The interacting effects of Website delay, familiarity, and breadth. *Information Systems Research, 17,* 1 (2006), 20–37.

30. Gupta, R.; Bagchi, A.; and Sarkar, S. Improving linkage of Web pages. *INFORMS Journal on Computing, 19,* 1 (2007), 127–136.

31. Halvey, M.; Keane, M.T.; and Smyth, B. Mobile Web surfing is the same as Web surfing. *Communications of the ACM, 49,* 3 (2006), 76–81.

32. Hong, W.; Thong, J.Y.L.; and Tam, K.Y. The effects of information format and shopping task on consumers' online shopping behavior: A cognitive fit perspective. *Journal of Management Information Systems, 21,* 3 (Winter 2004–5), 149–184.

33. Huberman, B.A.; Pirolli, P.; Pitkow, J.E.; and Lukose, R. Strong regularities in World Wide Web surfing. *Science, 280,* 5360 (April 3, 1998), 95–97.

34. Ivory, M.Y., and Hearst, M.A. Improving Web site design. *IEEE Internet Computing, 6,* 2 (2002), 56–63.

35. Jarvenpaa, S.L., and Todd, P.A. Consumer reactions to electronic shopping on the World Wide Web. *International Journal of Electronic Commerce, 1,* 2 (1997), 59–88.

36. Jean, W.H. The harmonic mean and other necessary conditions for stochastic dominance. *Journal of Finance, 39,* 2 (1984), 527–534.

37. Joachims, T. Making large-scale SVM learning practical. In B. Schölkopf, C.J.C. Burges, and A.J. Smola (eds.), *Advances in Kernel Methods—Support Vector Learning.* Cambridge: MIT Press, 1999, pp. 169–184.

38. Johnson, E.J.; Bellman, S.; and Lohse, G.L. Cognitive lock-in and the power law of practice. *Journal of Marketing, 67,* 2 (2003), 62–75.

39. Katz, M.A., and Byrne, M.D. Effects of scent and breadth on use of site-specific search on e-commerce Web sites. *ACM Transactions on Computer–Human Interaction, 10,* 3 (2003), 198–220.

40. Kelly, G.A. *The Psychology of Personal Constructs.* London: Routledge, 1992.

41. Kleinberg, J. Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46,* 5 (1999), 604–632.

42. Koch, T.; Golub, K.; and Ardo, A. Users browsing behavior in a DDC-based Web service: A log analysis. *Cataloging & Classification Quarterly, 42,* 3–4 (2006),163–186.

43. Kuhlthau, C.C. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42,* 5 (1991), 361–371.

44. Kwan, I.S.Y.; Fong, J.; and Wong, H.K. An e-customer behavior model with online analytical mining for Internet marketing planning. *Decision Support Systems, 41,* 1 (2005), 189–204.

45. Lazonder, A.W.; Biemans, H.J.A.; and Wopereis, I.G.J.H. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science, 51,* 6 (2000), 576–581.

46. Liu, J.; Zhang, S.; and Yang, J. Characterizing Web usage regularities with information foraging agents. *IEEE Transactions on Knowledge and Data Engineering, 16,* 5 (2004), 566–583.

47. MacArthur, R.H., and Pianka, E.R. On the optimal use of a patchy environment. *American Naturalist, 100,* 916 (1966), 603–609.

48. Malak, G.; Sahraoui, H.; Badri, L.; and Badri, M. Modeling Web quality using a probabilistic approach: An empirical validation. *ACM Transactions on the Web, 4,* 3 (2010), 1–31.

49. Marchionini, G. Interfaces for end-user information seeking. *Journal of the American Society for Information Science, 43,* 2 (1991), 156–163.

50. McEneaney, J.E. Graphical and numerical methods to assess navigation in hypertext. *International Journal of Human–Computer Studies, 55,* 5 (2001), 761–786.

51. McKinney, V.; Yoon, K.; and Zahedi, F.M. The measurement of Web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research, 13,* 3 (2002), 296–315.

52. Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 2 (1956), 81–97.

53. Mobasher, B.; Cooley, R.; and Srivastava, J. Automatic personalization based on Web usage mining. *Communications of the ACM, 43,* 8 (2000), 142–151.

54. Moe, W. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology, 13,* 1–2 (2003), 29–40.

55. Montoya-Weiss, M.M.; Voss, G.B.; and Grewal, D. Determinants of online channel use and overall satisfaction with a relational, multichannel service provider. *Journal of the Academy of Marketing Science, 31,* 4 (2003), 448–458.

56. Nadkarni, S., and Gupta, R. A task-based model of perceived website complexity. *MIS Quarterly, 31,* 3 (2007), 501–524.

57. Nielsen, J. *Designing Web Usability: The Practice of Simplicity.* Indianapolis: New Riders, 2000.

58. Nunnally, J. *Psychometric Theory.* New York: McGraw-Hill, 1978.

59. Palmer, J.W. Web site usability, design, and performance metrics. *Information Systems Research, 13,* 2 (2002), 151–167.

60. Pei, J.; Han, J.; Mortazavi-Asl, B.; Wang, J.; Pinto, H.; Chen, Q.; Dayal, U.; and Hsu, M.-C. 2004. Mining sequential patterns by pattern-growth: The PrefixSpan approach. *IEEE Transactions on Knowledge and Data Engineering, 16,* 11 (2004), 1424–1440.

61. Perkowitz, M., and Etzioni, O. Towards adaptive websites: Conceptual framework and case study. *Artificial Intelligence, 118,* 1–2 (2000), 245–275.

62. Pirolli, P. Rational analyses of information foraging on the Web. *Cognitive Science, 29,* 3 (2005), 343–373.

63. Pirolli, P., and Card, S. Information foraging in information access environments. In I.R. Katz, R. Mack, L. Marks, M.B. Rosson, and J. Nielsen (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* New York: ACM Press, 1995, pp. 51–58.

64. Pirolli, P., and Card, S. Information foraging. *Psychological Review, 106,* 4 (1999), 643–675.

65. Rodríguez, M.G.; Gayo, J.E.L.; and Lovelle, J.M.C. Web navigability testing with remote agents. In G. Goos, J. Hartmanis, and J. van Leeuwen (eds.), *Lecture Notes in Computer Science,* vol. 2016. Berlin: Springer, 2001, pp. 311–323.

66. Salton, G., and McGill, M.J. *Introduction to Modern Information Retrieval.* New York: McGraw-Hill, 1983.

67. Simon, H. A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 1 (1955), 99–188.

68. Song, J.H., and Zinkhan, G. Determinants of perceived Web site interactivity. *Journal of Marketing, 72,* 2 (2008), 99–113.

69. Spiliopoulou, M. Web usage mining for Web site evaluation. *Communications of the ACM, 43,* 8 (2000), 127–134.

70. Spiliopoulou, M.; Mobasher, B.; Berendt, B.; and Nakagawa, M. A framework for the evaluation of session reconstruction heuristics in Web-usage analysis. *INFORMS Journal on Computing, 15,* 2 (2003), 171–190.

71. Spink, A., and Cole, C. Human information behavior: Integrating diverse approaches and information use. *Journal of the American Society for Information Science and Technology, 57,* 1 (2006), 25–35.

72. Srikant, R., and Yang, Y. Mining Web logs to improve Website organization. In V.Y. Shen, N. Saito, M.R. Lyu, and M.E. Zurko (eds.), *Proceedings of the Tenth International World Wide Web Conference.* New York: ACM Press, 2001, pp. 430–437.

73. Srivastava, J.; Cooley, R.; Deshpande, M.; and Tan, P. Web usage mining: Discovery and applications of usage patterns from Web data. *ACM SIGKDD Explorations Newsletter, 1,* 2 (2000), 1–12.

74. Stigler, G.J. The economics of information. *Journal of Political Economy, 69,* 3 (1961), 213–225.

75. Su, B. Characteristics of consumer search on-line: How much do we search? *International Journal of Electronic Commerce, 13,* 1 (2008), 109–129.

76. Szymanski, D.M., and Hise, R.T. E-satisfaction: An initial examination. *Journal of Retailing, 76,* 3 (2000), 309–322.

77. Takagi, H.; Saito, S.; Fukuda, K.; and Asakawa, C. Analysis of navigability of Web applications for improving blind usability. *ACM Transactions on Computer–Human Interaction, 14,* 3 (2007), 311–323.

78. Tam, K.Y., and Ho, S.Y. Understanding the impact of Web personalization on user information processing and decision outcomes. *MIS Quarterly, 30,* 4 (2006), 865–890.

79. Tung, L.L.; Xu, Y.; and Tan, F.B. Attributes of Web site usability: A study of Web users with the repertory grid technique. *International Journal of Electronic Commerce, 13,* 4 (2009), 97–126.

80. Vance, A.; Elie-Dit-Cosaque, C.; and Straub, D.W. Examining trust in information technology artifacts: The effects of systems quality and culture. *Journal of Management Information Systems, 24,* 4 (Spring 2008), 73–100.

81. Vapnik, V. *Statistical Learning Theory.* Chichester, UK: Wiley, 1998.

82. Vaucher, S., and Sahraoui, H. Multi-level evaluation of Web site navigability. In G.A. Di Lucca and H.M. Kienle (eds.), *Proceedings of the 12th IEEE International Symposium on Web System Evolution.* Los Alamitos, CA: IEEE Computer Society Press, 2010, pp. 93–100.

83. Webmaster guidelines. Google, March 20, 2011 (available at www.google.com/support/webmasters/bin/answer.py?hl=en&answer=35769/).

84. Webster, J., and Ahuja, J.S. Enhancing the design of Web navigation systems: The influence of user disorientation on engagement and performance. *MIS Quarterly, 30,* 3 (2006), 661–678.

85. Wei, C.; Chiang, R.H.L.; and Wu, C. Accommodating individual preferences in the categorization of documents: A personalized clustering approach. *Journal of Management Information Systems, 23,* 2 (Fall 2006), 173–201.

86. West, D.B. *Introduction to Graph Theory*, 2d ed. Upper Saddle River, NJ: Prentice Hall, 2001.

87. Wolfinbarger, M., and Gilly, M.C. eTailQ: Dimensionalizing, measuring and predicting eTail quality. *Journal of Retailing, 79,* 3 (2003), 183–198.

88. Wu, H.; Gordon, M.; DeMaagd, K.; and Fan, W. Mining Web navigations for intelligence. *Decision Support Systems, 41,* 3 (2006), 574–591.

89. Yamada, S.; Hong, J.; and Sugita, S. Development and evaluation of hypermedia for museum education: Validation of metrics. *ACM Transactions on Computer–Human Interaction, 2,* 4 (1995), 284–307.

90. Yen, B.P.C. The design and evaluation of accessibility on Web navigation. *Decision Support Systems, 42,* 4 (2007), 2219–2235.

91. Zaki, M.J. SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning, 42,* 1–2 (2001), 31–60.

92. Zhang, P., and von Dran, G.M. 2000. Satisfiers and dissatisfiers: A two-factor model for Website design and evaluation. *Journal of the American Society for Information Science, 51,* 14 (2000), 1253–1268.

93. Zhang, Y.; Zhu, H.; and Greenwood, S. Web site complexity metrics for measuring navigability. In H. Weber, H. Ehrich, and K. Schewe (eds.), *Proceedings of the Fourth International Conference on Quality Software.* Los Alamitos, CA: IEEE Computer Society Press, 2004, pp. 172–179.

94. Zhou, Y.; Leung, H.; and Winoto, P. MNav: A Markov model-based Web site navigability measure. *IEEE Transactions on Software Engineering, 33,* 12 (2007), 869–890.

Appendix A: Warm-Up Exercises and Information-Seeking Tasks

Warm-Up Exercises

1. Find the location of the College of Business Administration and the dean's bio.
2. Find the university's president's name.
3. Find the page containing current campus news and then the page containing the information about the university (e.g., facts, history, etc.).

Information-Seeking Tasks

1. Find the location and operating hours of the Campus Main Library.
2. Find the page containing the description of the University Athletics and then the page containing the description of the University Football team.
3. Find the location and hours of the Office of Academic Advising and then the Office of Career Services.
4. Find the page containing a list of current campus events.
5. Find the location and store hours of the Campus Bookstore.
6. Find parking permit rates and how to buy parking permits.
7. Find the contact information and operating hours of the Campus Medical Center.
8. Find the Academic Calendar of the current academic year and then the dates for spring break.
9. Find the class schedule of the current semester and then the location of a specific course.
10. Find the page containing Campus Directory and then the page containing Campus Map and Directions.
11. Find the page containing Campus Recreation Services and then the page containing Campus Sports Clubs.
12. Find the Tuition and Rates of the current semester and how to pay tuition.

Appendix B: Measurement Items Used in the Study

Cognitive Processing Load (sources: [32, 59])

CPL-01: In the study, it generally took me a lot of processing efforts to figure out how to find a target page/content on the Web site.

CPL-02: I needed a lot of thinking when deciding how to navigate from a current page toward the target page/content on the Web site.

CPL-03: In general, I spent a lot of cognitive effort to find a target page/content on the Web site.

CPL-04: Generally speaking, my navigating the Web site to locate a target page/content was cognitively demanding.

CPL-05: Overall, I incurred a significant cognitive load when trying to find a target page/content on the Web site.

Table C1. Accessibility Measure

Page accessibility	Site A	Site B
Mean	0.98	1.34
Standard deviation	5.00	5.79

Appendix C: Comparing Proposed Metrics and Additional Benchmark Measure

WE BENCHMARKED THE PROPOSED METRICS against the accessibility measure developed by Yen [90], which evaluates a Web site's navigation according to the ease of accessing a page through the site's hyperlink structure. When a site is high in accessibility, it is easier for visitors to find information on the site. According to Yen [90], the accessibility of a page is higher if more hyperlinks point to it and its source pages (i.e., pages with a hyperlink pointing to that page) are located closer to the homepage. For Sites A and B, we calculated the accessibility of each page, as summarized in Table C1.

We performed Welch's t-test to compare the page accessibility of both sites; the page accessibility of Site A is significantly lower than that of Site B ($t = 2.64$, $p < 0.01$). This comparative analysis suggests that visitors can find information more easily on Site B than on Site A, a finding that contradicts our user evaluation results, in which participants were able to locate target pages more successfully and easily on Site A than on Site B. Web site navigation should be assessed at the confluence of the hyperlink structure and user browsing behaviors [11, 51, 59]; it manifests how well a Web site's hyperlink structure enables visitors to find information by navigating the site. In this light, the proposed metrics consider more comprehensive Web data and therefore can better reveal navigability by producing analytical evaluation results congruent with actual user performance, assessment, and satisfaction.