Classifying and Coding Online Actions

ESZTER HARGITTAI
Northwestern University

Research on how the Internet is diffusing across the population has broadened from questions about who uses the medium to what people do during their time online. With this change in focus comes a need for more detailed data on people's online actions. The author provides a method for coding and classifying users' online information-seeking behavior. The author presents an exhaustive list of ways in which a user may arrive at a web page. The proposed methodology includes enough nuanced information to distinguish among different search actions and links. In its entirety, the coding scheme makes it possible to understand many details about the users' sequence of actions simply by looking at the spreadsheet containing the information proposed in this article. The author also demonstrates the utility of this scheme with findings from a study on the information-seeking behavior of 100 randomly selected Internet users to exemplify the utility of this coding and classification scheme.

Keywords: Internet; web; methodology; coding; classification; online behavior

An increasing body of literature documents what segments of the population are Internet users (e.g., Hargittai, 2003a; National Telecommunications and Information Administration, 2002), and there is a growing emphasis on exploring what these users do online (see, e.g., Howard & Jones, 2003; Katz & Rice, 2002; Wellman & Haythornthwaite, 2002). As the research agenda moves toward examining in depth people's online behavior, we need new methodologies to collect the type of data that allow for nuanced analyses of people's online actions. Although more traditional methods of data collection such as survey research can be very helpful in collecting valuable information about people's Internet uses, some questions require additional types of data. In this article, I present a methodology for the collection, classification, and coding of fine-grained data about how people move around from one page on the World Wide Web to the next.

Knowing details about people's web-use behavior can be important for a wide variety of research questions. Studies ranging from political participation online to cultural consumption using new media will benefit from details about what web sites users visit. Do people look for news on traditional news organizations' web sites, or do they visit alternative sources of political information? Do users rely on content aggregators such as big portal sites to channel them toward content, or do they reach the less well-known and more hidden corners of cyberspace? Does everybody rely on the same search engine and know how to use it well, or are there discrepancies in online abilities? Once on an intended web page, on what

AUTHOR'S NOTE: I would like to thank Paul DiMaggio for helpful suggestions. I am also grateful to Inna Barmash and Susan Lutz for their assistance with coding. I also express my gratitude to the many people who took time from their busy schedules to participate in this study. Generous support from the Markle Foundation and NSF grant no. IIS0086143 is kindly acknowledged. The project has also been supported in part by a grant from the Russell Sage Foundation. I am also grateful to the Dan David Foundation for its support.

Social Science Computer Review, Vol. 22 No. 2, Summer 2004 210-227
DOI: 10.1177/0894439303262560

parts of the presented content do users focus? By collecting details about people's online behavior, we can answer such and many other questions about how people are using the web. But how do we make sense of the large amounts of data we have once we gather details about users' online actions?

The classification scheme I present in this article represents an exhaustive list of ways in which a user may arrive at a web page. A unique contribution of the project on which this article draws is that it did not impose on participants the use of any particular computer interface, web browser, or types of online services they could access to perform online actions. Given this flexibility, users turned to a myriad of tactics to find information on the web. By online action, I refer to the mode of moving from one web page to the next.

The proposed methodology is flexible. Not all projects may require the level of detail the classification scheme makes possible. Because codes are presented grouped according to conceptual categories, a researcher may decide to use only the higher level categories in his or her research project. Nonetheless, having a list of codes with which to start should facilitate the coding of log data about people's online actions.

In this article, I first review the literature on web searching. I then describe the study in which I collected data on people's online information-seeking behavior. I discuss in detail the methodology I used to classify and code people's online actions. I start by outlining the categories of tactics and then describe in detail the types of actions under each category. I share the results of coding one segment of a user's actions to showcase an example of how the coding scheme works. I also describe through examples some of the ways in which the classification and coding scheme helps analyze data about users' online actions. I present examples of hurdles users encounter during their online experience and also show how few people possess advanced search skills. The coding and classification system presented in this article facilitates aggregating and analyzing data about users' online information-seeking behavior making such description of the data possible.

RESEARCH ON ONLINE INFORMATION SEEKING

Scholars from many fields have explored how people use the web for information retrieval. Advertising and marketing specialists often refer to users as "consumers," emphasizing their particular interest in people's online actions, namely their shopping behavior (Bell & Tang, 1998; Jarvenpaa & Todd, 1996; Moe, 2003; Montgomery, 2003; Novak, Hoffman, & Yung, 2000). Work conducted in the human-computer interaction field tends to focus on features of web sites and how these hinder or facilitate web site navigation (Shneiderman, Lazar, & Melody, 2003; Shum & McKnight, 1997). Alternatively, computer scientists and those in the information science field draw on large-scale aggregate logs about people's web use by analyzing all web activity over a specified period (Catledge & Pitkow, 1995; Huberman, Pirolli, Pitkow, & Lukose, 1998; Spink, Jansen, Wolfram, & Saracevic, 2002; Spink, Wolfram, & Jansen, 2001). However, these studies rarely dwell into the details of people's online actions beyond looking at sites visited (Goldfarb, 2002) or number of search terms used in search queries (Spink et al., 2002). Additional information about users' online actions is important for a deeper understanding of how people are navigating online content.

Researchers in the library and information science field have conducted studies closest to the type of method used in this project (Wildemuth, 2002). Wang, Hawk, and Tenopir (2000) collected data by observing how respondents search for information specified by the research team. Their project generated synchronized video-audio data, which were then analyzed for detailed information about respondents' search techniques. However, as often is the case in such studies, the participants for the study were graduate students and faculty in

TABLE 1
Descriptive Statistics About Study Participants.

Item	*Mean*	SD	*Median*	*Minimum*	*Maximum*
Age	42.96	15.86	42	18	81
Education[a]	NA	NA	College	Less than high school	Ph.D.
Family income[a]	NA	NA	$80-$89,000	$17,500-$19,000	$250,000
Number of years since first use of the Internet	6.28	3.38	6	0	16
Number of hours browsing the web weekly	8.62	9.39	7	8 minutes	70 hours

a. Education and family income have no means as those variables were collected categorically.

an information science program. Moreover, many details about the users' actions were not coded.

To gain a better understanding of how the general population is using the Internet, it is important to include people from beyond the academic community in such studies (Rice, McCreadie, & Chang, 2001). Anigbogu and Rice (2001) did just that in their in-depth case-study of a user's actions taken to find health-related information. However, given how time- and labor-intensive such one-on-one case studies can be, these projects usually have no more than a few respondents, making quantitative analyses and larger scale comparisons and generalizations impossible.

The methods used in the studies cited above provide important information for a baseline understanding of how certain people navigate particular parts of the web. However, existing studies either limit their scope to specific user populations (e.g., undergraduate or graduate students, IT professionals, or people who go to libraries), do not collect background information about user attributes, or look at use patterns on an aggregate level without collecting data about the specific goals of a web user. This project remedies these shortcomings by collecting information about all these attributes of users and their online actions concurrently in one study. In this article, I focus in particular on the classification and coding scheme I developed for quantifying users' online actions. I describe that scheme in detail after a brief description of the research project in the next section.

THE WEB-USE PROJECT

I developed the classification and coding scheme presented in this article for a project conducted on the online skills of Internet users between the summers of 2001 and 2002. Participants represent a random sample of Internet users in Mercer County, New Jersey. One hundred people took part in the study, representing a 54% response rate, considerably high given that respondents were asked to come to the university research center to participate in the study and spent on average 1½ hours at the study location. Table 1 presents some basic demographic information about the participants, including how long they have been Internet users and how often they browse web content.

The researcher first orally administered a 25-minute questionnaire to obtain baseline information about users' usual web use patterns. Next, the respondents sat at a machine and performed some assigned tasks, such as looking for information about tax forms, political

candidates, health information, and local cultural events. Finally, participants filled out an online survey that included demographic details and some additional questions about usual web-use patterns and their Internet-use history.

Respondents sat at the machine of their choice (PC or Mac) and used their preferred browsing software (Internet Explorer, Netscape, or America Online). Using an audio-video software recording program, every action and comment of participants was recorded for subsequent coding and analysis. The focus of this article is to describe in detail how this rich material was coded, allowing quantification, aggregation, analysis, and comparison of information about users' online actions.

The goal of the study was to see how people find information on the web and to measure their online skills. Skill was operationalized as the ability to locate different types of content on the web, and people's efficiency (measured as time to completion) in doing so. Hargittai (2002) described the recruiting methodology and the observation sessions in more detail. Hargittai (2003b) provides copies of the survey instruments.

THE CLASSIFICATION AND CODING SCHEME

Here, I describe the coding scheme I developed for classifying and quantifying online actions. This information is recorded on a spreadsheet on which each action that results in a new web page being viewed is designated by a separate line of code. The line includes information on the action, the web page the user is viewing, time of the action, and other details about how the user got to the page—all of which I discuss in detail below. I start by offering a detailed description of the actions by which users arrive at a web page.

At the most basic level, a user may arrive at a web page by (a) typing in its address directly into the location/address bar of the browser; (b) using a feature of the browser program; (c) filling out and submitting a form including a search query; or (d) clicking on a link, broadly defined. The link a person clicks on can have very different origins. The link may have been generated by a search engine in response to a query submitted by the user, or it may be a link present on a web page the user is already viewing.

If the user possesses previous knowledge about a web site, then he or she can simply type in the address of the page. Alternatively, the user can draw on a feature of the browser program (e.g., a Favorites link or Bookmark) to arrive at a web page. There are also numerous forms a user may fill out, the most popular of which is the search engine query form. A user types in a term or multiple terms, pushes the Submit button, and arrives at a result. Finally, a user may simply click on a link (that may have different origins) and arrive at a web page in that way. This link can be a portal site category directory prominently featured on popular web sites (e.g., Yahoo!), it may be a sponsored advertisement link, or it may be a simple link in the middle of some text. Finally, there are some miscellaneous ways in which a user may get to a web page such as being automatically redirected.

Material on a web page can have two origins: content relevance or sponsorship. These may influence the content of various browser settings and web page substance in different ways. The coding scheme presented in this section attempts to code information about people's online actions in a way that retains the nuances of what underlying motivations on the part of site- and browser-software creators leads users to a web page. Each of the possible actions can happen in various ways. When creating the coding scheme for classifying actions, I distinguished between different types of actions that can come under each of these categories. Next, I describe them in detail. The appendix presents all the action codes with their respective codes.

Directly Accessing a URL (Web Address)

A user may arrive at a specific uniform resource locator (URL) directly by typing the address into the location bar of the browser. More specifically, the user may type in the entire address, but he or she may also start typing in a URL and then, through the autocomplete option, finish it quickly (the program remembers previous actions and allows the user to complete the web address by choosing the automatically recalled address). Instead of relying on the autocomplete feature, a user may also pull down the location bar and select a URL he or she had visited a few minutes earlier. It is also possible for the user to press down the Back button and pull down a list of previously visited sites. These are all various shortcuts that may make browsing and getting around web sites more convenient, quicker, and shows a certain level of sophistication in using the medium.

There are some additional features of browsing programs that allow easy direct access to web sites. It is also possible to go quickly to the preset homepage by clicking on the Home button. Use of Bookmarks or Favorites listings also allows direct movement to a predesignated web page. Finally, if the user closes the browser program (either by accident or intentionally), the default homepage will show up automatically when the browser is restarted, and this action gets a separate code to signal the particular context. This first set of codes in the appendix refers to these actions.

Use of Browser Feature

The browser programs all have buttons additional to the ones already mentioned that allow for quick navigation. The most commonly used is the Back button; in fact, some have studied it extensively and have noted that use of the Back button composes a significant amount of users' online actions (Cockburn, Greenberg, Jones, McKenzie, & Moyle, 2003; Greenberg & Cockburn, 1999). For this reason, the coding scheme includes refined measures of using the Back button with separate codes for clicking on it once versus clicking on it numerous times (the latter is an action whose effect can be easily achieved by alternative tactics and thus suggests a lower level of user sophistication and can slow down navigation). A user may get to a site quickly by using a link accessed through the History button of the browser or a customized link. Other browser buttons such as the Forward button, Reload/Refresh button, and Search button also take the user to various sites and thus are included in the coding scheme.

Search Engines

The use of search engines is of particular interest to the field of information seeking as so much of how people find information on the web happens by means of search engines. Accordingly, I include a particularly detailed coding scheme for use of search engines. A separate list of codes exists for the type of search engine respondents used versus the types of search engine results they pursued. That is, if a user went to a search engine and performed a search, the first code records information about the search engine itself (e.g., whether the search engine was a major search engine, a site-specific search engine, or whether it was a feature of a browser program such as American Online's (AOL) keyword search, which is directly accessible on the browser and does not require going to a particular site), whereas the subsequent code specifies information about which of the search results the user chose to pursue. Information about the brand of the search engine is also available on the spreadsheet

under a different variable (it is included in the specific URL associated with the action as opposed to the code of the action itself).

In addition to noting what type of search engine a user queried, it is also important to distinguish between searches users knew they were performing versus searches they performed by using the location bar of the browser in less-informed ways. Many users type terms into the location bar thinking they are going directly to a web site, but they get automatically redirected because they did not enter an actual web address, rather, just terms (e.g., typing in "new york times" instead of "nytimes.com"). In such cases, newer versions of some browsers (as opposed to their first versions from the mid-1990s) redirect users to preset search engines. That is, browsers take the terms, submit them to a search engine, and present the user with the results automatically. AOL redirects users to its own search engine automatically. Some versions of Internet Explorer redirect to MSN search. Some versions of Netscape also redirect to a preset search engine. Noting that a search was performed in this manner is also important when studying online actions and, thus, a separate code exists to signal such moves. Codes identifying these tactics are numbered 41-50 in the appendix.

After users perform searches, they usually click on one of the search results. What particular search result they click on is of interest as it signals information about people's understanding of various search engine features and advertisements. It is desirable to distinguish to the greatest extent possible links that are featured by search engines because of sponsorship or advertising deals from seemingly genuine results thought to be most relevant to the user's query. Thus, the coding scheme includes separate codes for sponsored matches, featured matches, and other types of results. Detailed information in this realm allows us to ascertain the popularity of sponsored and featured matches, helping us infer the extent to which commercial considerations influence users' online actions (Hargittai, 2004, p. 10). Moreover, because Google, Yahoo!, AOL, and the MSN search bar were the most common search engines used by participants in the study, and because they all include various link options in their results listings, these all have separate codes of their own in the coding scheme (these are listed as 500 codes in the appendix).

Portal Site and Directory Categories

It is of additional interest to distinguish between general links on a page versus directory categories available on portal sites (big point-of-entry content-aggregator sites such as Yahoo!). These are not just links, rather, they are content prominently featured whose popularity may be interesting and important to discern. For this reason, it is worthwhile to note how often users turn to directories featured on such sites for selecting what content to view. Accordingly, the coding scheme includes a separate list of codes for actions that involve pursuing links by way of directory categories. In particular, AOL has a list of very visible channels on the left side of the browser window to which AOL users may turn often. It was important to specify use of these features as the information can tell us about particular sources of online content popular for users. These codes range from 60-66 in the appendix.

Advertisements

Little focus has been paid to the origins of links. Do users know about advertisement-sponsored results? I consider advertisements as a special type of link, as the reason for their presence on a site is not merely content relevance. They can have different sources, and documenting the particular type of ad users click on can be informative. It may help us ascertain better whether users consciously click on these links or whether they are being tricked into

clicking on advertisements. To understand in detail what types of ad links users pursue, I created a separate list of codes for advertisements (codes in the 70s). I also include codes for closing pop-up windows as that is also an action as well as noting when someone resets the browser. Although these may be of less interest, to have an exhaustive list of online actions, it is necessary to have codes for such tactics as well.

Miscellaneous Links and Features

Having accounted for special links accessed through directory categories and search engines, we are left with the multitude of links that make up the web. Most web sites include links to other pages, other sites, or simply to other parts of a web page. I include separate codes for each such possible link (noting whether it was within page, within site or off-site, and whether it was a text link or a graphical link). These codes range from 80 to 89.

There are some additional miscellaneous actions that may occur while a user is surfing the web. At times, people will close windows accidentally (or on purpose), they will select another open window (from the bottom explorer bar on their screen in the Windows environment), or may be redirected automatically to another site. By closing a window or minimizing one, a user may end up on a new site via an already open window that they had either previously visited or one that popped up in the background (a so-called pop-under). Again, to account for all possible movements from one page to the next, it is necessary to code such actions as well. These are listed under Miscellaneous and range from 90 to 99 in the appendix.

Not all of the codes are mutually exclusive in the sense that some actions recorded by more specific codes could also be designated by more general codes. For example, if a user runs a search on Google using the "I'm feeling lucky" button, then that action could be coded simply as 30 for having run a search on a search engine. However, the code 301, which stands specifically for using the "I'm feeling lucky" feature in Google, allows for being more specific. The rule of thumb is that a more specific code always overrides a more general code to retain as much information as possible. At a later stage in the analysis, certain codes can be collapsed if the granularity of detail is not of interest.

Additional Information About Online Actions

The final spreadsheet includes information on every action in sequence taken by the respondent as per the coding scheme described above. In addition to these action codes, much other information is also recorded on the data sheet. Each web site the user visited is noted on the spreadsheet. Detailed information is coded about whether and, if yes, how much the person scrolled up and down a web page (these codes are listed at the end of the appendix under Scrolling). In cases in which users performed a search, the exact search query was recorded, including the use of Boolean operators (e.g., the word *and*), quotation marks, or any other specifications. These are coded on the sheet retaining any possible spelling mistakes the user may have made (in case of spelling mistakes, these are noted with "[sic]" to flag that they are deliberately included as such on the coding sheet and are not spelling mistakes introduced by the coder).

In cases in which users clicked on a text link, the exact text of the link was added to the data sheet. Similarly, if the user clicked on a button link, the text on the button was also transcribed. In cases in which respondents filled out an online form, a variable denotes the content entered into these form fields. If the participant selected an option from a pull-down menu, then the selected information is recorded. When the respondent came to a site from the

default homepage they had set up at the beginning of the study session, then this was noted in a column to signal how often users depend on features of their default page.

The data sheet also includes information about errors that may have occurred during the respondent's browsing activity and any other particularities that may have come up. For example, if the user engaged in a temporary action but then suspended it so that it did not lead to a new page, their actions were still noted. Such would be a case when the user clicked on a pull-down menu but then decided not to pursue that line of action and did not choose anything from the menu options. In cases in which users are performing specific tasks, it is useful to add an identification number for each task. If the question is whether the user succeeded in achieving a goal or not, then a binary variable can signal success or failure on the task. Having this information helps identify the end of a session.

Overall, the coding scheme allows us to distinguish among many different types of actions. After coding the videos, we retain information about whether the user clicked on a link (and whether that link was an advertisement, an image link, a within-page or within-frame link, a category directory item, or a search result), how the user got there, and whether he or she went to a web site directly by typing in a URL or pursuing another action.

Table 2 presents an example of what the coding looks like for a search task performed from start to finish. Enough details are coded to allow a reader of the output to recreate the steps of the user without having to see the video file. In the case of this example coding sheet, we can tell that the user took six actions to complete Task 6, which asked respondents to find a web site that compares different presidential candidates' views on abortion. The goal of creating the coding scheme was to quantify users' online actions to allow for easy aggregation and quantitative analyses across users.

This respondent—a 47-year-old woman who is a full-time school teacher—started out by hitting the Back button twice (Action 22) and arriving at the Google search engine. She started this task about 8 minutes into the observation session (time: 7.92). She ran a search using the word "abortion," but after seeing the results decided to refine her search query and typed in "abortion presidential comparison." She then clicked on a Google search engine result link (Action 510) stating "Candidate Comparisons," and arrived on a web page hosted on the www.familyvoice.org web site. That particular web page likely did not contain the information for which the user was looking because she hit the Back button once (Action 21) and returned to the results of her Google search. Next, she clicked on a link (Action 510) stating "Abortion: Presidential Candidates Views," which led her to a web page on the Issues2000.org web site. At this point, she successfully completed the task. This user spent less than 2 minutes looking for content that would satisfy this task. Thanks to the coding scheme, we know exactly how she got to the final web site.

People's online actions can vary widely. Some people may use similar actions regardless of the types of content they seek, others may turn to different tactics depending on the task. Users may also combine or switch actions if their initial attempts did not prove fruitful. In the next section, I present some descriptive statistics about the variety of actions people use and how these relate to user efficiency in navigating online content.

ANALYZING USERS' ONLINE ACTIONS

The above-described classification and coding scheme on these data yields a wealth of information about users' online information-seeking behavior. Here, I present some of the findings to show how the coded data can be aggregated and analyzed. For those who rely on the web for work or use it for various daily tasks and who have developed a core base of

TABLE 2
Example of Data Sheet With Detailed Information About a Participant's Online Actions

Action	*URL*	*Time*	*Task*	*Scroll*	*Search Query*	*Link Text*	*Error*	*From Home*	*Form Information*	*Select*	*Button Text*	*Autocomplete*	*Success*	*Comments*
201	www.google.com	7.92	6	0										
30	www.google.com/search? . . .	8.40	6	0	Abortion									
30	www.google.com/search? . . .	8.97	6	0	"Abortion Presidential Candidates"	Google Search								
510	www.familyvoice.org/candidate_comparisons.htm	9.18	6	1		Candidate Comparisons								
20	www.google.com/search? . . .	9.42	6	0										
510	www.issues2000.org/2000/Abortion.htm	9.65	6	2		Abortion: Presidential Candidates Views							1	

NOTE: URL = uniform resource locator.

TABLE 3
Examples of Mistaken Location Bar Use[a]

Princeton Packet.com, metropolitan museum of art.com
www.new york times.com, time warner.com
my wallet has been stolen.com
nursing spectrum.com, consumers reports.com
Philadelphia Phillies.com, J.C. Penneys.com

a. Each row represents a different user's actions. Examples on the same line were performed by the same user.

TABLE 4
Examples of Search Queries That Do Not Contain Spaces Entered Into Search Forms[a]

plumber'slocal9, capitalhealthsystem
presidentalcampaign2000
princetonhistoricalsocietyvolunteer
princeeeeetonartmuseum, employmentopportunities, fordescort, frickmuseum

a. Each row represents a different user's actions. Examples in the same row were performed by the same user.

online skills, it may be surprising to learn that some users never use certain arguably basic actions online such as typing the address of a web site directly into the location bar of the browser or running a query using a search engine.

Use of the Location Bar

In the classification scheme presented above, one of the core types of actions is to type in the address of a web site directly in the location bar of the browser. Among the 100 participants in this study, 14% never used this tactic. Instead, these users rely on recommendations from search engines or click on links from their default homepage—or, in the case of AOL users, pick an AOL channel from the left-hand menu and pursue links from there.[1]

Of these 14 users who never type in the address of a web site, some do use the location bar in one way or another, leaving 6 users who never do anything with this feature of the browser. Those who do use it will sometimes type in a few words and use it as though it were a search engine. Two of the users did attempt to go to web sites directly, but, in fact, they were guessing URLs and typing them incorrectly by leaving spaces in between words, resulting in errors. They were not the only 2 users to make this mistake. Several participants made the mistake of putting spaces in between words and then adding a top-level domain extension to their query. The problem here is that it neither works as a search query (because of the .com at the end) nor as a URL (because of the spaces). Table 3 presents some examples of this action.

Confused Use of the Search Form

An interesting opposite of this mistake is when users do not include a space in their search query while attempting to use a search engine. Table 4 presents examples of this mistake. It may be that these users are confused about which form they are filling out and think they are typing a URL instead of a search query. However, given that they do not include a domain

name extension, this is unlikely. When asked about the lack of spaces in her search queries, a 63-year-old woman who had been a web user for 2 years and spends 3 hours online weekly noted the following:

> I guess it's my thinking about web e-mail addresses where there's no space and that kind of thinking or thought pattern, and many web addresses have no space in them. And it's been my pattern to omit spaces unless there's an underscore. Sometimes they have underscore.

Another user, a 39-year-old, stay-at-home mother of three who had only been using the Internet for 1 year and goes online just a few minutes each week did not have a particular reason for not using spaces in her query on "princetonhistoricalsocietyvolunteer": "I was spacing, so I decided to try and not space between the words." During this session, she clicked on AOL's Help feature to get some assistance when she was not coming up with any results during a search but noted, "See, I never did any of this. . . . This just takes too much time." Although she was successful in locating various types of content online, she spent more time on the tasks than 90% of participants.

Advanced Skills

Advanced search skills include the use of quotation marks around search terms to force proximity of terms in some search engines, the use of multiple terms for more refined queries, and the use of other Boolean operators such as the word *not* or the minus sign. No one used the minus sign, which signals the exclusion of terms in several search engines (e.g., in Google).[2] Although people did use the sign in their search queries occasionally, in all such cases, it was meant as a hyphen between terms entered into the search form.[3] This suggests that not only do people not know about the role of the minus sign in some search engine algorithms, they do not realize that they are using signs that could influence the results they will get using the particular operators.

In one case, a respondent included the minus sign in her query, which had the exact opposite of the intended effect. When a 21-year-old bank customer service representative typed in "lactose intolerance -recipes," her goal was to find sites that included recipes for those with lactose intolerance. Ironically, such a query in Yahoo!—a portal whose search engine at the time was powered by Google—would exclude all pages that included the word "recipes." It is not surprising, then, that this respondent spent 6 ½ minutes on this task and was only able to complete it once she refined her search terms and excluded the minus sign from the query.

Only 16% of participants used quotation marks during their sessions. Moreover, in some cases, those who did use them used them incorrectly or superfluously. A 72-year-old female participant knew to use quotes but, in some cases, used them around just one word, which serves no function in any search engine, so her knowledge of the feature was probably spotty. Another respondent, a 43-year-old female, self-employed professional concierge, used quotes around almost all of her search terms, often resulting in much too restrictive queries with few, or in some cases, no results. When asked why she used quotation marks, she explained the following:

> *Respondent:* My friend and I were e-mailing back to each other and usually, when we come across a site, we will share it, either in e-mail or a phone conversation. So, she e-mailed me that Google was one of the best sites and Ask Jeeves was passé, and then she also taught me to put the quotations marks between the words and it would come up with different sites the way I would use the quotation marks also.

Figure 1. The Result of a Misspelled URL Guess gugenheim.com Accessed Using the America Online Interface

Researcher: What do quotation marks do? Why do you use them?
Respondent: I think it—let's see. I don't know—[hesitates, then begins laughing] because she told me to put quotation marks about almost everything. But the way you use the quotation marks you get different sites that come up.

These examples suggest that even those who use features that would be considered advanced search functions do not always know their purpose, and lacking this deeper knowledge, they sometimes misuse them, resulting in less-than-optimal results.

Some users exhibit advanced skills that few possess, and although they may not use them often, they draw on them at times. For example, there were only 2 users in the sample who took advantage of Google's cache feature during their entire sessions. This feature allows the user to pull up an old cached version of a web page from Google's archives. It is an extremely helpful feature when a site fails to come up and the user is faced with an error page. There was just 1 user who took advantage of the Find-on-Page feature of browser programs to find specific terms on long web pages. Some browser features are better known. Of the users, 43% knew to access a previously viewed URL by way of a shortcut feature such as clicking down on the Back button and selecting the address or clicking on the location bar and choosing the URL from among all previously viewed web sites. These can be helpful in cutting down on amount of time clicking from site to site.

TABLE 5
Examples of URL Guesses

www.princetonuniversity.edu, www.nassauchurch.com, www.arthistory.com
www.irs.com, www.irs.org, www.redcross.com, www.smithsonian.gov

Guessing

Users often guess URLs incorrectly either because they make spelling mistakes, they randomly put together words that do not add up to an existing web address, or they use the wrong top-level domain extension (i.e., they use .com instead of .org, .org instead of .edu, and so forth). Table 5 presents examples of all these mistakes from actual occurrences in the study. Part of the problem is that web sites sometimes do come up despite misspelled or wrongly guessed URLs, thereby reinforcing faulty actions.

Figure 1 shows an example of a web site that came up when the user—a 56-year-old man who works in business development and has been an Internet user for 4 years, surfing the web for less than 4 hours weekly—typed in "gugenheim.com" in search of the Guggenheim Museum's web site. By encountering this web site, the user did not realize he had mistyped the name of the museum. Rather, he thought the site was not yet fully accessible as per the note on the gugenheim.com site. He noted, "Well, it's their homepage, but I'm not getting any further action." He then proceeded to try the .org and .net version of gugenheim. One of these redirected him to a domain registration web site, further confusing the process.

It is curious that some users will continue guessing URLs despite their lack of success with this method. Often, a simple search for the organization whose site they are trying to locate would get them to the desired destination more quickly, but some will continue to use the method regardless.

CONCLUSION

The aggregate statistics presented in the previous section are just some examples of why it is helpful to code and quantify users' online actions. As noted earlier, there are a myriad of research questions that could benefit from such detailed information about people's web-use behavior. Other analyses may focus on the types of web pages people visit, for example, looking at how often people go to some sites over others. Such data can help us understand whether the web is leading people to turn to alternative sources of information or whether they continue to get news and other material from more traditional sources. By having information about how people find sites, we can discern their level of understanding about navigating online content and what browser or web-site features are most popular with them.

Having such detailed information about users' online actions helps us understand the differences in people's ability to use the web effectively and efficiently. It also makes other details about people's online behavior quantifiable, allowing for aggregation. In addition to helpful descriptive statistics, the data can be used to inform analyses about what types of people engage in various online behaviors. Once we have quantitative measures of sites visited and online actions taken, we can include such information in analyses about skill, web-site usability, and content popularity.

As more and more daily tasks move online (e.g., accessing government documents, conducting financial transactions, obtaining political information), it will be increasingly important to assess and monitor whether people—average users—are able to make sense of

online materials and find their way to different types of content. Most policy discussions about inequalities in Internet use revolve around questions of access. However, as the examples in this article suggest, there are considerable discrepancies in people's ability to use the web efficiently, which may also have important implications for inequality. Those who cannot navigate online content are poised to be left behind and miss out on conducting important daily activities in the most efficient ways possible. The coding and classification scheme presented in this article helps conceptualize different types of online actions and makes numerous detailed analyses about people's web-use behavior possible.

APPENDIX
Coding and Classification Scheme

Code	*Description of Action*
	Directly accessing a URL (Web address) including default Home
10	Type in URL in location bar
101	Type in URL without .com extension and press Ctrl-Enter
11	File → Open
12	America Online (AOL) Internet www.
13	URL truncation
14	Pull down location bar for URL
15	Back button pull-down for URL
16	Go menu (Netscape) for previous URL
17	Home button
171	Default move to homepage
18	Favorites button/Bookmark
181	Favorites button/Bookmark within Favorites/Bookmark directory
	Browser buttons
20	Back button—simple click—once
201	Back button twice
202	Back button three times
203	Back button four or more times in a row
21	Forward button
22	Reload/Refresh button
23	Search button (Internet Explorer or Netscape)
231	AOL search button
24	Stop button
25	URL by way of History file link
26	Browser link bar (customized link)
	Search engines
30	Major search engine search
301	Google "I'm feeling lucky" search
302	Google "Search within results" search
31	Within-site search engine search
32	Location bar terms search
33	Confused use of location bar
34	Search with MSN search column on left
35	AOL keyword
351	AOL main search form
36	AOL "Internet" (on top section of browser)

37 AOL channel search form
38 Search with topic-specific search form
39 Major search engine search not from the search engine
40 Edit → Find (Ctrl F)—find function within page

Search engine results

41 Search engine result link
42 Search engine result link—sponsored link
43 Search engine result link—related searches link
44 Search engine result link—recommended category
45 Within search engine result link (site accessed through internal site search)
46 Topic-specific search result link
47 Search engine "More results/next 20" link
48 Yahoo!: "Go to Web Page Matches"
49 Search engine "Previous results/previous 10/20" link
50 Suggested links below search results link

Particular search engine results

500 AOL search results link
501 AOL Recommended Sites link
502 AOL Sponsored Links link
503 AOL Matching Sites link
504 AOL Narrow Your Search link
505 AOL "Search for 'x' on:" link
506 AOL Other Searches
507 AOL "Show me more like this" link
510 Google results link
511 Google results link Cached
512 Google result Category link
513 Google result Similar Pages link
514 Google Sponsored link
515 Google Categories
516 Google "Did you mean?" link
520 Yahoo! Category Matches link
521 Yahoo! Inside Yahoo! Matches link
522 Yahoo! Category Matches link
523 Yahoo! Web Site Matches link
524 Yahoo! Web Page Matches link
525 Yahoo! Sponsored Matches link
526 Yahoo! More Sponsor Matches link
527 Yahoo! Site Listings
528 Other Yahoo! results
529 Yahoo! "More sites about:" category
530 MSN sidebar Featured Site link
531 MSN sidebar Sponsored Links link
532 MSN sidebar Web Directory Sites link
533 MSN sidebar results link
534 MSN "Top ten most popular for"
540 Metacrawler "Featured search results"

Directory categories

60 Major portal category directory/web guide
61 Site category directory
62 Category selection specific to person/task
63 AOL channel
64 AOL channel category directory

65 AOL top pull-down menu directory category
66 Main menu item link

Advertisements

70 Click on banner or other graphical ad
71 Fill in banner ad form
72 Click on pop-up ad
73 Click on commercial "sponsored" link from search engine results page (explicit ad)
74 Close pop-up ad
75 Pop-up ad appears
76 Pop-up window (non-ad) appears
761 Close pop-up window (non-ad)
77 Click on OK to close a window
78 Confusion, can't leave site
79 Reset

Links

80 Click link—to offsite
81 Click link—to within site
810 Click AOL Channel link—within AOL
82 Click link—to within frame
83 Click link—to within page
84 Click on graphic button—to offsite
85 Click on graphic button—to within site
86 Click on graphic button—to within frame
87 Click on graphic button—to within page
88 Click on multimedia link
89 Click link to within frame but different site

Miscellaneous

90 Browser add-on (e.g., What's Related)
901 Set homepage
902 My Netscape button
91 AOL top browser directory
92 Automatic redirect
93 Close window
930 Close two windows quickly in a row
931 Close three or more windows quickly in a row
932 Application quits (window automatically closes)
933 File → Exit or Close
94 Select other open window
941 Select other open window by accidentally clicking on it
95 Restart browser
96 Maximize window
961 Minimize window
97 Miscellaneous (e.g., have to enter name/age for amusement purposes only)
971 Error
98 Close pop-up application/program
981 Application pops up
99 Username/password form

Scrolling

0 No scrolling
1 Scroll down
2 Scroll down and up
3 Scroll down and up and down
4 Scroll down, up, down, up

5	Scroll up
6	Scroll left or right or both
7	Scroll all over

NOTE: → indicates scrolling through a drop-down or pop-up menu. URL = uniform resource locator.

NOTES

1 One of the users counted as having typed in the URL directly in the location bar used AOL's browser feature, which opens a box that has "www." written in it and requires the user to type in the rest of the URL.

2. The coding scheme makes it easy to search for respondents who had used the word *not* or the minus sign in their queries by searching for cases where *not* or the minus sign is part of the column that includes the content of the query (see the Search Query column in Table 2).

3. In this study, we know what users were looking for, so it is possible to deduce this from their queries.

REFERENCES

Anigbogu, J., & Rice, R. E. (2001). Expectations and experiences of seeking infertility information via the Internet and the telephone directory. In R. E. Rice & J. E. Katz (Eds.), *The Internet and health communication: Experiences and expectations* (p. 459). Thousand Oaks, CA: Sage.

Bell, H., & Tang, N. K. H. (1998). The effectiveness of commercial Internet web sites: A user's perspective. *Internet Research*, *8*(3), 219-228.

Catledge, L. D., & Pitkow, J. E. (1995, April 10-14). *Characterizing browsing strategies in the World Wide Web*. Paper presented at the third International World Wide Web Conference, Darmstadt, Germany.

Cockburn, A., Greenberg, S., Jones, S., McKenzie, B., & Moyle, M. (2003). Improving web page revisitation: Analysis, design and evaluation. *IT and Society*, *1*(3), 159-183.

Goldfarb, A. (2002). Analyzing website choice using clickstream data. In M. Baye (Ed.), *Advances in applied microeconomics v.11: The economics of the Internet and e-commerce* (pp. 209-230). New York: Elsevier.

Greenberg, S., & Cockburn, C. (1999, June 3). *Getting back to back: Alternate behaviors for a web browser's Back button*. Paper presented at the Proceedings of the fifth annual Human Factors and the Web Conference, NIST, Gaithersburg, MD.

Hargittai, E. (2002). Beyond logs and surveys: In-depth measures of people's web-use skills. *Journal of the American Society for Information Science and Technology Perspectives*, *53*(14), 1239-1244.

Hargittai, E. (2003a). The digital divide and what to do about it. In D. C. Jones (Ed.), *The new economy handbook*. San Diego, CA: Academic Press.

Hargittai, E. (2003b). *How wide a web? Inequalities in accessing information online*. Unpublished doctoral dissertation, Princeton University, Princeton, NJ.

Hargittai, E.(2004). The changing online landscape: From free-for-all to commercial gatekeeping. In P. Day & D. S. Day (Eds.), *Community practice in the network society: Local action/global interaction*. Boston: Routledge Kegan Paul.

Howard, P. E. N., & Jones, S. (2003). *Society online*. Thousand Oaks, CA. Sage.

Huberman, B. A., Pirolli, P. L., Pitkow, J. E., & Lukose, R. M. (1998). Strong regularities in World Wide Web surfing. *Science*, *280*(5360), 94-97.

Jarvenpaa, S. L., & Todd, P. A. (1996). Consumer reactions to electronic shopping on the World Wide Web. *International Journal of Electronic Commerce*, *1*(2), 59-88.

Katz, J. E., & Rice, R. E. (2002). *Social consequences of Internet use: Access, involvement and interaction*. Cambridge, MA: MIT Press.

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

Montgomery, A. (2003). *Predicting online purchase conversion using web path analysis* (GSIA Working Paper no. 2003-E26). Pittsburgh, PA: Carnegie Mellon University.

National Telecommunications and Information Administration. (2002). *A nation online*. Washington, DC: Government Printing Office.

Novak, T. P., Hoffman, D. L., & Yung, Y. -F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, *19*(1), 22-42.

Rice, R. E., McCreadie, M., & Chang, S. -J. L. (2001). *Accessing and browsing information and communications*. Cambridge, MA: MIT Press.

Shneiderman, B., Lazar, J., & Melody, I. (2003). Introduction: Web navigation. *IT & Society*, *1*(3), i-vii.
Shum, S. B., & McKnight, C. (1997). World Wide Web usability: Introduction to this special issue. *International Journal of Human Computer Studies*, *47*(1), 1-4.
Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer*, *35*(3), 107-109.
Spink, A., Wolfram, D., & Jansen, M. B. J. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, *52*(3), 226-234.
Wang, P., Hawk, W. B., & Tenopir, C. (2000). Users' interaction with World Wide Web resources: An exploratory study using a holistic approach. *Information Processing and Management*, *36*, 229-251.
Wellman, B., & Haythornthwaite, C. (2002). *The Internet in everyday life*. Oxford, UK: Blackwell.
Wildemuth, B. M. (2002). Effective methods for studying information seeking and use. *Journal of the American Society for Information Science and Technology*, *53*(12), 1218-1222.

Eszter Hargittai is an assistant professor in the Department of Communication Studies and the Department of Sociology (by courtesy) and faculty associate of the Institute for Policy Research at Northwestern University. She holds a Ph.D. in sociology from Princeton University, where she was a Woodrow Wilson Scholar. She may be reached by e-mail at `research@eszter.com`.