

How do people find information on a familiar website?

Roy A. Ruddle

School of Computing, University of Leeds, Leeds, UK.

r.a.ruddle@leeds.ac.uk

ABSTRACT

Previous research has investigated how people either navigate the web as a whole, or find information on websites of which they have little previous knowledge. However, it is now common for people to make frequent use of one site (e.g., their employer's intranet). This paper reports how participants recalled and navigated a familiar website they had used for 8-20 months. Sketch maps showed that participants' memory for the site's content and structure was very limited in extent, but generally accurate. Navigation data showed that participants had much more difficulty finding the region of the site that contained a piece of information, than then finding the information itself. These data highlight the need for directly accessed pages to be given greater prominence in browser history mechanisms and designers to make information regions memorable. Finally, two navigational path metrics (stratum and percentage of revisit actions) that correlated with participants' performance were identified.

Categories and Subject Descriptors

H.3.3 Information Search and Retrieval: Search process. H.5.4 Hypertext/Hypermedia: Navigation.

General Terms

Experimentation, Human Factors.

Keywords

Navigation, information space, familiarity, sketch map, stratum, user study.

1. INTRODUCTION

There is a long history of studies into how people navigate the real world, which has both informed our basic understanding of the cognitive processes that are involved in navigation and provided designers with practical benefits. For example, it is generally accepted that our cognitive map contains different categories of spatial information (e.g., landmark, route and survey knowledge; [1,2,3]), landmark-action pairs are the most important component of effective route descriptions [4], best practice has been established for the design of maps that are used as navigation aids [5,6], and guidelines and models exist for the design of "legible" cities [7,8].

Similarly, researchers have carried out studies of hypertext and

web navigation. Some of this has been to understand how people access information (e.g., [9,10,11,12,13]), and other research has investigated ways in which the tools that people use may be improved (e.g., [14,15,16]).

One of the most common places people look for information is on their employer's intranet. However, there is a distinct lack of research into how people remember where information is in familiar websites, or go about finding it. This paper reports how people found information on a website they had been using for 8-20 months. Participants were first asked to recall the structure of the site by drawing a sketch map, and then navigate the site to find the answers to 12 questions. The results show how people go about the everyday task of finding information in a familiar place, and highlight ways in which the process of finding could be made much easier.

2. INFORMATION SPACE NAVIGATION

Early formal studies of information space navigation were conducted using hypertexts that contained up to 50 screens of information [17,18], and showed how the structure of a hypertext affected the accuracy with which, and the way it was recalled. Although the small size of these hypertexts means that the studies have limited relevance to the day-to-day problems people now encounter when navigating in large websites, the studies do highlight the utility of sketch maps for quantifying people's knowledge of information spaces.

More recent research has used web-scale spaces and tended to focus on: (a) how people navigate, (b) the techniques people use to remember and return to particular information, or (c) evaluation of new browser functionality.

2.1 How do we Navigate Web-scale Spaces?

Major studies of the way people navigate have been conducted in both the early days of the web [10,19] and recent times [12,13]. Through the use of instrumented browsers, these studies provide rich data about how people navigate the web as a whole in an everyday setting.

At current rates, a person will make 1.3 million visits to web pages during their lifetime, approximately half of which will be returns to pages that have been visited before [12]. Analysis of people's navigational actions shows that the percentage of occasions that URLs are accessed via a hyperlink (43%) or directly (e.g., typing or a bookmark; 9%) has remained stable over time, but use of the back button has declined substantially to be replaced by a corresponding increase in the submission of forms (including search engine queries) and the use of tabbed browsing ([10,12,19]. Analysis of the paths people follow highlights large differences between individuals. At one extreme, navigators tend to follow sequential paths and revisit many pages, whereas at the other, browsers' paths contain frequent branches [13].

Ruddle, R. A. (2009). How do people find information on a familiar website? *Proceedings of the 23rd BCS Conference on Human-Computer Interaction (HCI'09)*, 262-268.

Ten years ago, people only visited a small number of pages frequently [10]. However, this is likely to have changed because many organizations now make widespread use of intranets to provide information to their employees. These are information spaces with which those people become very familiar, so the way they access them may change just as our knowledge of the physical environments in which we live and work undergoes qualitative and quantitative changes over time [1,3,8].

2.2 Remembering/Returning to Information

Remembering where information is and being able quickly to return to it is essential if people are to conduct their everyday lives in an efficient manner. Most people make frequent use of a web browser's back button, but this is only useful for returning to information in one's current navigational session. The two mechanisms that browsers provide to aid return visits at a later date are bookmarks and the history list (usually accessed via the navigation bar).

Bookmarks are a mechanism by which people explicitly choose to store the location of a piece of information. An observational study of managers, librarians and researchers highlighted five other methods that achieve the same aim: storing the URL in an email, on a personal website or in a document, and storing the webpage itself either electronically or on paper [20].

Each of these explicit methods has advantages (bookmarks are integrated with one's browser, but the other methods are more portable). However, storage of a large number of items brings with it the problem of finding the one that you wish to use at a particular moment [21], and the act of storage has a time cost associated with it which inhibits usage. The result is that, even though the majority of people make some use of bookmarks, they are only relevant to a modest number of the pieces of information that people revisit [9].

A browser's history list automatically stores the address of pages that are visited, without requiring any input from a user. Studies show that people very rarely access the history list directly [10,12], but the information it contains also provides prompts when people type URLs into a browser's address bar. Thus, history list data are frequently used to help people recall the address of information to which they wish to return.

Studies conducted in an everyday setting show that, on the majority of occasions, people know (or think they know!) how to find a piece of information and locate it using an "orienting" strategy. This involves making a large step to the general region that contains the information and then using local cues to narrow in on its precise location [11].

Research that specifically investigated revisiting shows how people make this first step to an information region. Participants were interviewed to determine which websites in their history list they expected to return to within the next 12 months, and in a follow up session 3-6 months later the participants were asked to revisit some of those sites. Participants were successful at the first attempt on 93% of occasions, directly entering the URL in 42% of these first attempts, and using a bookmark (18%), search engine (18%) or approaching via another website (16%) for most of the remainder [9]. Participants took twice as long to find sites they only visited once or twice a year than those they visited daily, and relied on recalling their own memory (e.g., to type in the URL)

rather than information they had purposefully stored (e.g., bookmarks).

2.3 Browser Functionality

The previous section shows that people make little use of the functionality that current browsers provide to help revisit information. This has led to research into a number of types of new functionality.

First, several studies have highlighted the structural differences between how people navigate and the stack used with a browser's back button [10,14]. This led to the development of a "SmartBack" feature, which provided people with direct access to URLs that they used as hubs and substantially reduced conventional usage of the back button [22].

Second, other studies have investigated ways of incorporating information scent [23] when pages are displayed by a browser to guide people toward the information they seek. This helped participants to find information quicker than when conventional searching or browsing was performed [24].

Third, a number of researchers have developed browser add-ons that display an overview diagram of the pages which people have visited [14,15,16,25,26]. Evaluations indicate that these overviews help people to return to information, but further research is required to determine how to make this functionality scale to the large quantity of places that people visit over an extended period of time.

3. EXPERIMENT

We know of no previous research that has made detailed investigations of how people mentally represent and navigate familiar websites. In the following experiment, participants who had made frequent use of the School of Computing's website for 8-20 months were first asked to recall the site's structure and content, by drawing a sketch map, and then asked to navigate the site to answer 12 questions. The results help us understand people's memory for familiar websites, how they access information in such sites and how this process could be facilitated by improvements in browser functionality.

3.1 Method

3.1.1 Participants

Twenty School of Computing undergraduate students (18 men and 3 women) took part in the experiment. Their mean age was 22.1 years (SD = 4.5). All the participants volunteered for the experiment, gave their informed consent, and were paid an honorarium for their participation.

The experiment took place during June 2007, when 12 of the students had just completed the first year of their degree and the other eight had just completed their second year. The academic year starts at the end of September, so these sets of students had used the site for 8 and 20 months, respectively. Subsequent analysis of weblogs for the password protected parts of the site showed that the participants had accessed it an average of two days out of every five (including weekends and holidays) during the 38 weeks that led up to the start of the experiment.

Ruddle, R. A. (2009). How do people find information on a familiar website? *Proceedings of the 23rd BCS Conference on Human-Computer Interaction (HCI'09)*, 262-268.

3.1.2 Website

The School of Computing's website is divided into two main parts, one containing information for external people and an intranet that staff and students login to access. A crawl of the site conducted using WebSphinx [27] found 17999 different pages, excluding the School's wiki which is a large information space in its own right.

The external site contained 7596 pages, the majority of which concerned either research (2897 pages, describing the School's research groups and projects) or people (4257 pages; mostly staff and PhD students' personal webpages). The intranet contained 10403 pages, of which 4872 were on websites for the modules taught by staff in the School and 2823 were from SIS (School Information System), a web front end to a database that contained organizational (e.g., committees, minutes and room bookings) and personal information (e.g., a particular student's timetable and marks).

It is not known how the size of the site compares with other organisations' intranets. We expect that the School's intranet would be considered "mid-sized" and, in any case, no person is likely to have or need exhaustive knowledge of large intranets.

3.1.3 Procedure

The experiment was divided into three parts. First, a participant was asked to draw a diagram of all the parts of the School's website that they had visited, naming each page so the experimenter could identify it, and showing all links to other pages. Each participant spent between 30 and 45 minutes doing this.

Second, the participant was asked to provide some biographic information. This was performed after the sketch map to provide a mental break before the participant tackled the last part of the experiment.

Third, the participant was asked to answer 12 questions, which each required the participant to navigate to a particular page on the School's website, and took up to 30 minutes to complete. For this part of the experiment, the participant had a choice of using a Windows or Linux PC and logged in using their normal username, so they could use their preferred web browser (Internet Explorer, Firefox, or Opera) together with their own profile, bookmarks and history list.

Participants' navigational actions were recorded using webserver logfiles and by videoing the mouse, keyboard and screen. The logfiles captured requests made to the School's server, and the video allowed requests for cached pages to be noted (e.g., those that arose from use of the back button), the tab used to view each page to be determined, and contextual information such as when a participant abandoned one question and moved onto the next to be recorded.

Each question just required participants to navigate to a particular page on the School's website and look up the answer. Participants may not have known the exact location of each answer but, provided they knew where in the region of the site to look, then finding the answer itself was expected to be straightforward. The questions were:

- 1) On what date did induction week start in 2006?

- 2) How many textbooks are listed as "Background Reading" for the Introduction to Programming (SE15) module?
- 3) On the Introduction to Programming (SE15) module, for how many courseworks are "Extra hints and tips" provided?
- 4) What was the topic of the last lecture of the Computer Operating Systems (SY21) module?
- 5) On the Object-Oriented Software Engineering (SE20) module, when were the instructions for coursework 2 last revised?
- 6) For which years are past exam papers available on the Object-Oriented Software Engineering (SE20) module's website?
- 7) One of the industrial placement opportunities for 2007/8 is with Astute Ltd. In which town are they based?
- 8) How many sample exam questions are provided on the Professional Development 2 (PD31) module's website?
- 9) How many different groups were there for the coursework on the Professional Development 2 (PD31) module?
- 10) In which year was the Wren prize first awarded?
- 11) How many MSc/MRes taught masters degrees are offered by the School?
- 12) Research activity in the School is organised into three areas. How many research groups are there within the Multidisciplinary Informatics area?

These questions ran chronologically through a student's time at university, starting with induction (question 1), then modules studied in year 1 (questions 2 & 3) and year 2 (questions 4-6), an industrial placement (question 7), modules studies during a student's final year (questions 8 & 9), graduation prizes (question 10), and finally options for postgraduate study (questions 11 & 12).

At the end of each question (except 2, 5 & 8), the participant was asked when they had last visited the pages containing the answer (this week vs. within the last month vs. some time in 2007 vs. 2006 (or earlier) vs. never) and how frequently they had visited at that time (several times per day vs. several times per week vs. once or twice vs. never). These data were not requested for questions 2, 5 and 8 because those involved the same region of pages as the subsequent question.

Participants were predicted to use the website's link structure to navigate to the region containing eight of the answers, but for questions 2, 4, 5 and 8 to enter the URL directly. This was because those questions referred to specific modules, which each had a homepage in the form <http://www.comp.leeds.ac.uk/xxxx>, where 'xxxx' was a four letter code that was used throughout the School and was well known to staff and students (e.g., 'se15' for the Introduction to Programming module).

3.2 Results

The results are divided into three parts. The sketch map data provide information about participants' ability to recall the content and structure of the website. Logfiles and video data from

Ruddle, R. A. (2009). How do people find information on a familiar website? *Proceedings of the 23rd BCS Conference on Human-Computer Interaction (HCI'09)*, 262-268.

when participants answered the questions showed how participants navigated to the answers when they had access to all the cues they would have in an everyday setting. Metrics of navigational success/lostness were then investigated.

3.2.1 Sketch Maps

Each sketch map was transcribed into a machine-readable form by identifying each page that was drawn and the links between pages. Only 0.6% of the pages that participants drew on the sketch maps could not be identified. Another 5.6% of the pages did not exist, which mostly occurred when a participant had drawn several module websites with the same structure when they actually contained small differences. A total of 7.6% of the pages were either within the School's wiki, which participants had been instructed to omit from the sketch map, or on faculty/university websites not the School's. A further 2.1% of the pages were generic (e.g., labelled "lectures" when, in fact, there was a link to each individual lecture direct from the module's homepage) and reflected the fact that participants only included a certain level of detail on the sketch maps.

The remaining 84.1% were distinct pages from the School's website. Breaking these down by topic (see Table 1) showed that the majority came from either the websites of the modules that the students studied, or SIS (the School Information System). However, it was surprising how infrequently the sketch maps included pages from staff who taught the students or their research.

Table 1. Mean number of pages drawn on each participant's sketch map, divided by topic. The number found in the crawl of the site is shown for comparison.

Topic	Sketch map	Webcrawl
Home page	1.0	1
Research	0.8	2897
People	0.5	4257
Other external pages	2.0	441
Modules	21.2	4872
SIS	8.7	2823
Other intranet pages	2.8	2708
TOTAL	37.0	17999

Previous research showed that people were able to return to frequently visited websites twice as quickly as those that were only visited occasionally [9], and increases in recency and frequency are well known to have a positive effect on people's ability to recall information in general memory tasks. All the participants were 1st/2nd year students. The first six navigation tasks involved navigating to regions on the School's website that were primarily relevant to these students, so the data about the last time/how often participants visited these pages provides information about how recency and frequency affected participants' ability to recall details of the School's website for inclusion on the sketch maps.

The induction pages had been visited by 15 participants but were only included on one sketch map, the SE15 pages had been visited by all the participants but were only included on half the

sketch maps, and the SY21 and SE20 pages had been visited by all the 2nd year participants but were only included on two and three sketch maps, respectively. Overall, the likelihood of a participant recalling a region for inclusion on their sketch map increased with the recency of their last visit but, apart from visits

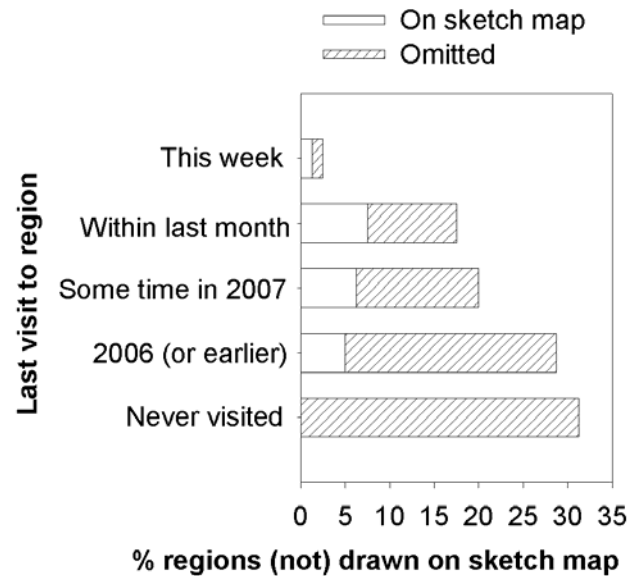


Figure 1. Percentage of times the induction, SE15, SY21 and SE20 pages were drawn on sketch maps, categorized by participants' self-reports of how recently they had visited that region of the website.

made within the previous week, the proportion who recalled a region were still a minority (see Figure 1). There was no discernible trend for visit frequency because, for 63% of regions that participants had visited, they had done so at least several times per week.

Participants drew an average of 43.1 links on their sketch maps (links explicitly marked as two-way were counted twice), of which 73% were correct. Another 18% were marked as directly connecting two pages when, in fact, a third page lay in between.

As previously stated, there was (inevitably) a limit to the amount of detail that participants drew on their sketch map. A sketch map either displayed a page as a branch (one or more outgoing links) or a leaf (no outgoing links). Analyzing only the branches on each sketch map, a comparison with the webcrawl data showed that participants only drew 17% of the links that actually existed.

To investigate what types of link were recalled, each webpage that appeared as a branch node was manually inspected and the links coded according to where on the page they appeared: header, menu (horizontal or vertical), the body of the page, or footer. If a link appeared more than once then the proportion that appeared in each region was noted, so the total for every link was always 1.0.

The manual inspection showed that 70% of links appeared within the body of a page, but only 15% of these were recalled and drawn on participants' sketch maps. Menus accounted for 27% of links of the pages, but participants only recalled 22% of these. The header and footer had few links, but participants recalled less than 7%.

3.2.2 Navigation

Unlike drawing a sketch map, where participants had to recall the website's content and structure, when answering the 12 questions participants navigated the website with all the cues that were present in an everyday setting. In this section, only data from participants' first attempt at each question are reported. For each participant, data from the School's weblogs were combined with information from the videos to produce a machine-readable transcript of the whole navigation session which showed every action that took place (e.g., create new tab, select link, or select back button), the page and tab being viewed, and identified when a participant finished one question and moved onto the next. The transcript was used for the analyses reported in this section, including determining the browser's history list at the end of each question, taking into account all the tabs that were open, so the shortest path to the next question could be calculated.

First, the overall pattern of user behaviour was broadly similar to previous studies (e.g., [12]). Link navigation accounted for 52% of participants' actions, back actions for 21%, direct access (10%) occurred almost wholly when participants typed in a URL, and 9% of actions were search operations. The remaining 8% of actions were to open, select and close tabs/windows.

A number of researchers have highlighted the need to separate the initial stage of navigation, where a person makes a large step to the general locality of their information goal, from subsequent movement to find the information's exact position [11,22,28]. Once in an appropriate region, there are generally sufficient contextual cues ("scent" [23]) to allow the person to narrow in on their target in a series of small steps. In the present study, questions 3, 6 and 9 proved to be trivial to answer because a participant only had to navigate within a given region (the SE15, SE20 and PD31 module websites, respectively). For all the other questions, if a participant found the correct region, then finding the answer itself was generally straightforward.

The remainder of this section only considers how participants navigated to the region that contained each question's answer. Data are not included for the 10 (out of 180) cases where a participant either wrote down the wrong answer (e.g., the number of MSc modules, not MSc/MRes degrees) or found the correct answer in an unintended place (e.g., on Astute's website not the School's). Initial analysis of the navigation data showed that the questions should be divided into three categories, according to the manner in which participants navigated to the region that contained the answer: direct (questions 2, 4, 5 & 8), link (questions 11 & 12) and other (questions 1, 7 & 10). A summary of the actions that participants performed is shown in Figure 2.

The four direct questions were the first question about each module. Participants knew that the answers to these questions were likely to be contained on the relevant module's website, and on 81% of occasions accessed this directly by typing in the URL of the module's homepage (e.g., <http://www.comp.leeds.ac.uk/se15>). This required one action, compared with the shortest path using hyperlinks/back button, which required 3-9 actions ($M = 4.2$).

The link questions were the two that, on 82% of occasions, participants accessed by using just the back button and hyperlinks. Of these, 88% followed a route that either was the shortest, or involved only one or two additional actions.

The other three questions were those that participants often found difficult to answer. In 42 out of 60 cases participants eventually provided the correct answer, though in five of these cases the answer was found in an unintended place (e.g., for question 7, on Astute's website rather than the School of Computing's). A factor that these questions shared was that pages that seemed relevant to the answer appeared in two distinct regions of the site. General information about induction and industrial placements was provided on the external pages for students who were considering coming to our university to study, and details about induction and placements were provided on the intranet for current students. Details of prizes were on the intranet but, by coincidence, the person after whom the Wren prize was named had just been awarded an honorary degree by the university and was featured in publicity material on the external pages. Search engine queries tended to either rank external pages higher than those on the intranet, or only list external pages because those on the intranet were not accessible to crawler agents that the search engine used for indexing.

The following section uses data from the 37 cases where

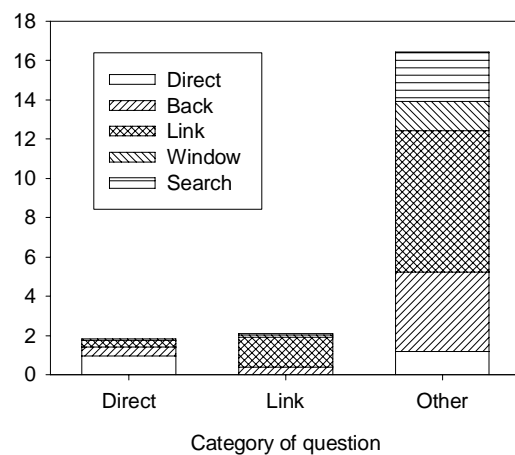


Figure 2. Mean number of each type of action that participants performed while navigating to the region of the website that contained the answer to question categorized as direct (Q2, Q4, Q5 & Q8), link (Q11 & Q12) and other (Q1, Q7 & Q10).

participants found the correct answer for the "other" questions in the intended place, to compare metrics that can quantify navigational success and the extent to which a participant was "lost".

3.2.3 Metrics of Navigational Success and Lostness

There are three forms of being lost: (a) Not knowing where to go, (b) Knowing where to go but not how to get there, and (c) Not knowing where one is [29]. In web navigation (c) is of limited importance because, unlike the real world, it is always trivial to return to a known place such as one's home page or the root page of the website one is currently navigating.

Metrics of lostness (and its opposite, success) can be divided into three main categories: time, distance and the path followed (for a summary, see [30]). Time is the best overall measure of

performance because it takes account of time taken to perform different actions. For example, it may be much faster for a participant to press 'back' a multitude of times to return to a known location than to scroll down a page and select a particular link that leads to the same location. Distance in web navigation is best expressed as the number of actions taken to travel from one page to another, because this takes account of every step a participant makes (e.g., directly enter a URL, select tab, search, back, or select hyperlink). Path metrics include counting the number of pages that are revisited, the amount of backtracking, and measures of the overall shape of a path such as stratum and compactness. These time, distance and path metrics are similar to those used to analyze the navigation in Euclidean (non-hyperlinked) spaces such as the real world and virtual reality.

In the present study, the time that participants took to navigate to the regions of the website that contained the answer to the "other" questions (Q1, Q7 & Q10) was determined from the videos, and compared with distance and path metric data. As would be expected, there was a significant correlation between time taken and the total number of actions performed ($r = .83, p < .01$). However, there was no correlation between time taken and the shortest path (the minimum number of actions necessary; $r = .06, p > .05$), which was indicative of the fact that participants often had difficulty finding information that was only a few steps away.

Three path metrics were used: number of iterations, percentage of revisit actions, and stratum. An iteration was defined as finding a start point (directly typing in a URL, or performing any number of back, search, close tab or select tab actions) followed by any number of steps "forward" (selecting hyperlinks). There was a significant correlation between time taken and the number of iterations ($r = .72, p < .01$). On average the iterations only contained 3.7 actions ($SD = 1.7$), with participants navigating in short iterations irrespective of whether the answer was proved straightforward or difficult to find.

There was a significant correlation between time taken and the percentage of actions that caused a revisit ($r = .64, p < .01$). Further investigation showed that 40% of revisits occurred when participants selected a hyperlink and found themselves on a page that had already been visited while trying to answer the current question. This was indicative of participants unintentionally, but repeatedly, navigating through the same parts of the site instead of looking somewhere new.

Stratum and compactness are two related metrics that were originally developed to analyze the structure of hypertext [31] and have subsequently been used to analyze navigation paths [30]. Stratum and compactness are normalized to take account of the size of a hypertext/path (any linear graph has a stratum = 1.0, and any fully connected graph has a stratum = 0.0). The present study used stratum because it was designed to capture the linear ordering of a hypertext [31], and a linear path is indicative of a participant knowing where they want to go and how to get there. A disadvantage of the original definition of stratum is that it only considers the path network and not the number of times each link in that network was traversed, but the present study overcame this by using the number of node visits, rather than number of nodes, to normalize the value of stratum that was calculated. There was a significant correlation between time taken and the stratum of participants' paths ($r = -.65, p < .01$), and between stratum and the

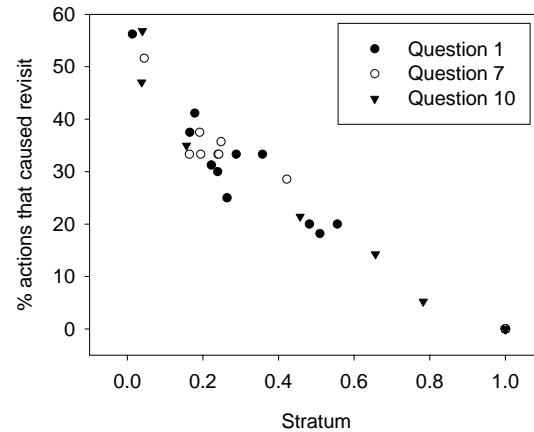


Figure 3. Correlation of stratum with the percentage of actions that caused a revisit. The data are for the 37 cases where participants found the answer to the "other" questions (Q1, Q7 & Q10) in the intended place

percentage of actions that caused a revisit ($r = -.97, p < .01$; see Figure 3).

4. CONCLUSIONS

This is the first study that has performed a detailed investigation into how people find information on a large website with which they are familiar. Previous research has either investigated how people navigate the web as a whole, not differentiating between familiar and unfamiliar sites (e.g., [10,11,12,13]), or studied how people found information on a site of which they had little or no previous knowledge (e.g., [24,25,29]). A homogenous group of participants (School of Computing students) were recruited, so that navigating in one familiar website could be studied in detail. While these participants are likely to have a greater general understanding than the population as a whole of the internal structure of websites, there is no reason to expect this understanding to have fundamentally affected participants' ability to navigate, just as when Computing students have taken part in virtual reality (cf. 3D computer games) navigation studies the difference between individuals' performances have been large.

4.1 General Findings

In keeping with the findings of previous research [11,28] analysis of the actions that our participants performed indicated that navigation should be divided into two parts: (a) reaching the general region that contained a piece of information, and (b) finding the information's exact location.

If navigation was problematic then the difficulties almost always occurred in (a), and participants tended to navigate in short iterations, each of which only contained a few actions. However, rather than each iteration looking somewhere new, a substantial number of pages were revisited and 40% of this occurred when participants selected a hyperlink (rather than back, forward or another tab) indicating that the revisiting may have been unintentional.

Ruddle, R. A. (2009). How do people find information on a familiar website? *Proceedings of the 23rd BCS Conference on Human-Computer Interaction (HCI'09)*, 262-268.

Two metrics that take account of the amount of revisiting in a navigation path and can be normalized to allow comparison of different websites and information retrieval tasks are: (a) the percentage of actions performed that cause a revisit, and (b) the stratum [31]. The advantage of stratum is it takes account of the connectivity of a navigation path, and so can distinguish between a path that contains: (i) linear branches caused by deliberate backtracking, or (ii) cycles/loops caused by unintentionally revisiting a page. However, stratum requires much more computation than the percentage of revisit actions, which is difficult to justify because the actions metric correlated as strongly as stratum with participants' performance in the present study.

If participants did know which region contained the answer then they either followed a linear path that revisited few, if any, pages, or accessed the region directly by typing the URL into the address bar. This indicates two improvements that would be beneficial for web systems. First, directly accessed pages should be given greater prominence/weighting in browsers' history mechanisms. Second, persistent identifiers [32] are well-suited for storage by computer applications but not for recall by humans. Websites would be more navigable if regions were named using memorable persistent identifiers even if individual items of information continued to have abstract identifiers (e.g., 10.1016/0169-7552(95)00043-7 [19]). This is the case on our School's website, where the regions associated with modules happen to be named with a four character code that staff and students find easy to recall.

4.2 Recall vs. Recognition

The sketch maps showed that participants were able to recall only a small amount of the content and structure of the School's website, despite having made frequent use of it for 8-20 months. Some might say "so what", arguing that websites are navigated by recognition and recall is not relevant, but this misses the point. In the real world, recall plays a critical role helping us remember parts of an environment that lie outside our immediate view, plan where to travel and give other people directions. Websites are more difficult to navigate than the real world, but the situation would be substantially improved if recall played a more prominent role.

Web browsers could promote recall by presenting a diagrammatic overview based on the structure of the website. Overviews have proven benefits [15,16,26], and presenting one with a world-referenced, rather than ego-referenced, perspective [5] would provide consistency from one site visit to the next.

Finally, such overviews could be annotated with a person's path or automatically generate "ScentTrails" [24] based on the person's most recent searched queries. The path would highlight the repetitious nature of their travels during iterative navigation, making it more likely that the person would then channel their energy into looking at new places. The optimum depth for scent would depend on the structure of the site, but three links would have been sufficient in the present study.

5. REFERENCES

- [1] Ishikawa, T. and Montello, D. R. 2006. Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric

knowledge and the integration of separately learned places. *Cognitive Psychology* 52, 93-129.

- [2] Kitchin, R. M. 1994. Cognitive maps: What are they and why study them?. *Journal of Environmental Psychology* 14, 1-19.
- [3] Thorndyke, P. W. and Hayes-Roth, B. 1982. Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology* 14, 560-589.
- [4] Denis, M. 1997. The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition* 16, 409-458.
- [5] Aretz, A. J. 1991. The design of electronic map displays. *Human Factors* 33, 85-101.
- [6] Levine, M., Marchon, I. and Hanley, G. 1984. The placement and misplacement of you-are-here maps. *Environment and Behavior* 16, 139-157.
- [7] Hillier, B., Penn, A., Hanson, J., Grajewski, T. and Xu, J. 1993. Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design* 20, 29-66.
- [8] Lynch, K. 1960. *The image of the city*. MIT Press, Cambridge, MA, USA.
- [9] Bruce, H., Jones, W. and Dumais, S. 2004. Information behavior that keeps found things found. *Information Research* 10(1) paper 207, available at <http://InformationR.net/ir/10-1/paper207.html>.
- [10] Tauscher, L. and Greenberg, S. 1997. How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies* 47, 97-137.
- [11] Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R. 2004. The perfect search engine is not enough: A study of orienteering behavior in directed search. *Proc. CHI 2004*, ACM Press, New York, NY, 415-422.
- [12] Weinreich, H., Obendorf, H., Herder, E. and Mayer, M. 2006. Off the beaten tracks: Exploring three aspects of web navigation. *Proc. WWW 2006*, ACM Press, New York, NY, 133-142.
- [13] White, R. W. and Drucker, S. M. 2007. Investigating behavioral variability in web search. *Proc WWW 2007*, ACM Press, New York, NY, 21-30.
- [14] Cockburn, A. and Jones, S. 1996. Which way now? Analysing and easing inadequacies in WWW navigation. *International Journal of Human-Computer Studies* 45, 105-129.
- [15] Milic-Frayling, N. and Sommerer, R. 2003. WebScout: Support for revisitation of web pages within a navigation session. *Proc. IEEE/WIC Web Intelligence 2003*, IEEE Press, 689.
- [16] Wexelblat, A. and Maes, P. 1999. Footprints: History-rich tools for information foraging. *Proc. CHI 1999*, ACM Press, New York, NY, 270-277.
- [17] Edwards, D. M. and Hardman, L. 1989. "Lost in hyperspace": cognitive mapping and navigation in a

Ruddle, R. A. (2009). How do people find information on a familiar website? *Proceedings of the 23rd BCS Conference on Human-Computer Interaction (HCI'09)*, 262-268.

- hypertext environment. In R. McAleese (Ed.) *Hypertext: Theory into Practice*, Intellect, Oxford, UK, 105-125.
- [18] Simpson, A. and McKnight, C. 1990. Navigation in hypertext: structural cues and mental maps. In R. McAleese and C. Green (Ed.) *Hypertext: State of the Art*, Intellect, Oxford, UK, 73-83.
- [19] Catledge, L. D. and Pitkow, J. E. 1995. Characterizing Browsing Strategies in the World-Wide Web. *Proc. WWW 1995*, DOI: 10.1016/0169-7552(95)00043-7.
- [20] Jones, W., Bruce, H. and Dumais, S. 2001. Keeping found things found on the web. *Proc. CIKM 2001*, ACM Press, New York, NY, 119-126.
- [21] Abrams, D., Baecker, R., and Chignell, M., 1998. Information archiving with bookmarks: Personal web space construction and organization. *Proc. CHI 1998*, ACM Press, New York, NY, 41-48.
- [22] Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A. and Sommerer, R. 2004. SmartBack: Supporting users in back navigation. *Proc. WWW 2004*, ACM Press, New York, NY, 63-71.
- [23] Furnas, G. W. 1997. Effective view navigation. *Proc. CHI 1997*, ACM Press, New York, NY, 367-374.
- [24] Olston, C. and Chi, E. H. 2003. ScentTrails: Integrating browsing and searching on the web. *ACM Transactions on Computer-Human Interaction* 10, 177-197.
- [25] Hightower, R. R., Ring, L. T., Helfman, J. I., Bederson, B. B. and Hollan, J. D. 1998. Graphical multiscale web histories: A study of PadPrints. *Proc. Hypertext 1998*, ACM Press, New York, NY, 58-65.
- [26] Jhaveri, N and R  ih  , K. 2005. The advantages of a cross-session web workspace. *Proc. CHI 2005*, ACM Press, New York, NY, 1949-1952.
- [27] Miller, R. C. and Bharat, K. 1998. SPHINX: A framework for creating personal, site-specific web crawlers. *Computer Network and ISDN Systems* 30, 119-130.
- [28] Pandit, S and Olston, C. 2007. Navigation-aided retrieval. *Proc WWW 2007*, ACM Press, New York, NY, 391-400.
- [29] Otter, M. and Johnson, H. 2000. Lost in hyperspace: Metrics and mental models. *Interacting with Computers* 13, 1-40.
- [30] Gwizdka, J., and Spence, I., 2007. Implicit measures of lostness and success in web navigation. *Interacting with Computers* 19, 357-369.
- [31] Botagogo, R., Rivlin, E., and Shneiderman, B., 1992. Structural analysis of hypertext: identifying hierarchies and useful metrics. *ACM Transactions on Information Systems* 10, 142-180.
- [32] Tonkin, E. 2008. Persistent Identifiers: Considering the Options. *Ariadne issue 56*, available at <http://www.ariadne.ac.uk/issue56/tonkin/>.