Web Users Information Retrieval Methods and Skills

Carol S Bond
Senior Lecturer, Informatics for Health and Social Care
Institute of Health and Community Studies, Bournemouth University
Bournemouth House, Christchurch Road
Bournemouth,
Dorset,
England.
BH1 1LH
Tel 01202 504356

Email cbond@bournemouth.ac.uk

Abstract

When trying to locate information on the WWW people are faced with a variety of options. This research reviewed how a group of health related workers approached the task of finding a named document. Most were eventually successful, but the majority encountered problems in their search techniques. Even experienced web users had problems when working with a different interface to normal, and without access to their favourites. No relationship was found between the number of years experience web users had and the efficiency of their searching strategy.

Introduction

The Internet can be an excellent resource, and is increasingly being used by nurses to both help meet their own information needs, and to help patients locate reliable health information. The major disadvantage of using Internet based resources is the problem of locating good quality information from within the vast amount of information available.

The majority of Internet based information used by nurses is found on the WWW (World Wide Web), it is however very easy to forget that the web is still relatively new, having only been available to the public for the last decade. It is developing rapidly, search tools change, new ones are developed, and the pool of information available is growing more rapidly than anyone would have imagined when it was first developed.

It is perhaps then not surprising that many people, not just nurses, feel overwhelmed by the amount of information available. A study carried out with nursing students (Bond 2003) found that almost two thirds complained about finding too much irrelevant information when looking for something on the WWW.

Finding Information on the WWW

There have been a variety of studies looking at how users navigate the WWW, but few review the initial choice users make when faced with a specific information retrieval task, and with the user having the a free choice of approach.

The use of links has been explored, for example by Reed et al (2000) who researched how users internet experience affected the way they followed links, finding that more experienced users tended to follow more linear links than inexperienced users, who tend to make more use of links with non linear relationships to the preceding link.

The effect of page layout and content on the user is another area that has been reviewed. Diaper and Waelend (2000) found that the use of graphics on a page did not affect the users' ability to find information. The effect of using frames as a navigational tool was considered by van Schaik & Ling (2001) who found that placing the navigation frame at the top or the left

hand side of the page was equated with better user performance in terms of speed and accuracy.

The usability contributed (or constrained) by the browser used was reviewed by Cockburn and Jones (1996) who concluded that web users were frequently not aware of where they were in 'WWW Subspace', and that they frequently did not interpret web navigation and use the web browsers' capabilities in the way that web site designers anticipated.

A lot of studies (for example Muylle et al 1999, Battleson et al 2000, Diaper & Waelend 2000) follow users information retrieval skills or experience on a given site, but do not include a review of how the user finds the site in the first place. Greenburg (2000) found that 85% of web users turn to a search engine when they are trying to find information.

The research

Method

This research project was designed to investigate how people approached the task of finding information on the web. The participants were staff of a nursing department in an English university. The participants had a range of skill and experience in using the Internet. The aim of the research was to observe their normal techniques for finding information. The sample was purposive, with people invited to participate according to the author's personal knowledge of their IT skills to try to ensure that the sample included people with a variety of skill levels and experience.

The method was reviewed for any ethical issues that needed to be considered. One possible issue was that all of the participants were, to some extent, known to, and colleagues of, the researcher. The researcher also explained in the invitation to participate that the research was part of her doctoral studies. This could have meant that there was a danger that the participants would try to either impress or to help the researcher by 'getting it right'. In view of this care was taken in the wording and presentation of the task to make sure no suggestion of a 'right' approach was made. To further ensure that no bias was introduced the participants weren't told exactly what the aim of the research was before they undertook the task. All participants were however offered a full debriefing immediately afterwards.

Participants were asked to carry out a simple task, finding a document that was available on the web. All the participants used the same computer for the task, set up with Internet Explorer (V5), without any favourites, and with a neutral home page, (the academic departments' home page). The browser cache and history were cleared before each user to ensure that they were not influenced by previously followed links. The wording of the request to find the information was carefully planned to avoid using any terms that might influence the approach to the task. Participants were asked to try to find the document referred to, using whatever methods they would normally use. Potentially leading words or phrases such as 'search for the document' were avoided.

Participants were reassured that they were not being judged, or graded in any way, and that their success or otherwise in finding the document was immaterial to the success of the research. They were not, however briefed on the purpose of the exercise beforehand. This was to avoid saying anything that might influence how they approached the task.

Each participant was given written instructions, to ensure both consistency, and that the full and accurate name of the document was available to them at all times.

The exact wording of the task was:

> *The UK Department of Health are developing a National Service Framework for diabetes.*
>
> *As part of this they undertook a consultation with users, and produced a report entitled Listening to Diabetes Service Users. This is available on their website.*

This document was chosen for several reasons. The topic was one that many of the participants would be familiar with, and there were a wide range of options available for locating it. It was significant enough to be listed by major search engines (this was checked before this document was selected for the task). It could be located through the Department of Health (DoH) web site, which has several ways of locating information in the site on its home page. There is a search option, a side category based navigation bar, and a link to an A-Z topic listing.

Participants were not limited in either the number of attempts or methods that they used to try and find the document.

Results

In total 18 people participated in the research, all but three finding the information, eventually. People were asked for their age, the range spanned from 19 to 63 (see chart 1). Although there is a popular idea that young people are more comfortable with new technology this study did not find any evidence that they had more experience, spent more hours on line, or had better searching skills than older participants.

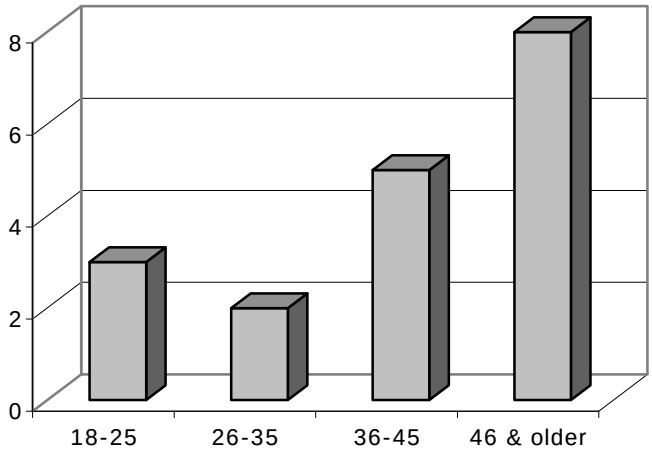

Chart 1

Age Range

The amount of time that people had been using the Internet ranged from 1 to 7 years, the average being 4.2 years. Respondents' estimation of their current usage ranged from half an hours use per month, to 90 hours online per month, with an average of 24 hours. This is shown in chart 2, where the number of years experience in using the internet that the participants had, and the amount of time they estimate that they currently spend online each month are shown.

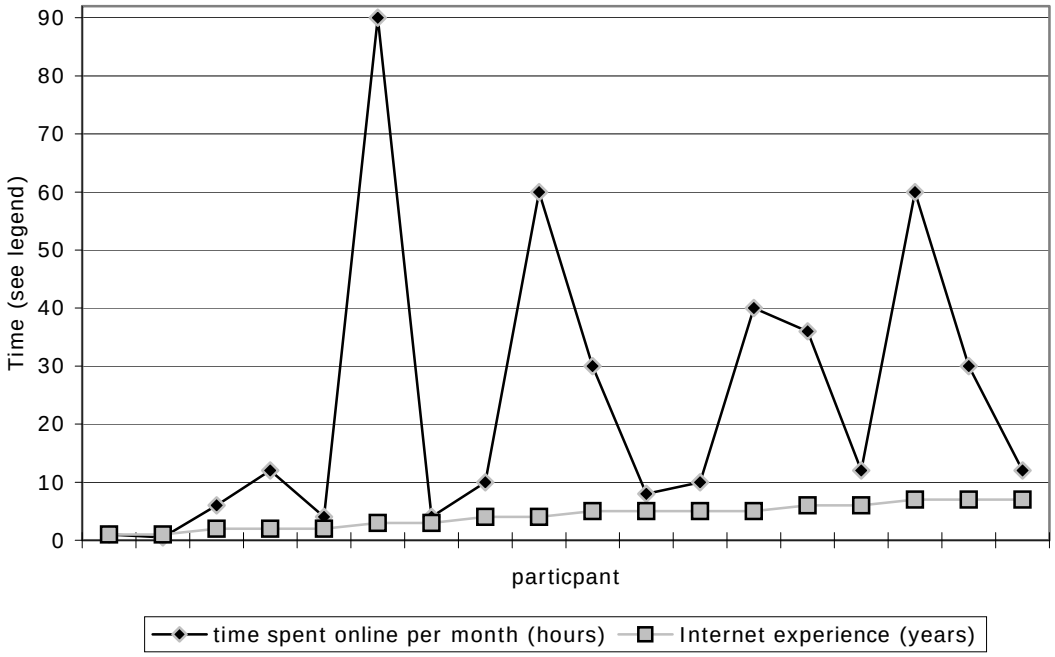

Chart 2

Years of Experience & Monthly Use per Participant

As can be seen there is no relationship between the number of years experience of the user and the amount of time spent online each month. This suggests that neither measure can be used in isolation as an indicator of the amount of experience that a user has.

In all but one case time online included both work and leisure related use, undertaken in the workplace and at home. The other respondent, although in an academic job, only used the internet at home for leisure, and was the respondent who used the internet least. Three of the more experienced users multi-tasked, opening one browser window to go directly to the DoH web site (by entering the address), and one with a search engine.

In analysing the results it emerged there were three ways that the respondents had first tried to find the information. Some tried going directly to the DoH web site by enter the web address; some used the default search option available within the browser to try to find the Department of Health web site; and some selected their own preferred search engine to find either the Department of Health or the document itself (chart 3 below).

Of those who initially searched by one means or another all but two searched for the Department of Health in preference to the named document.

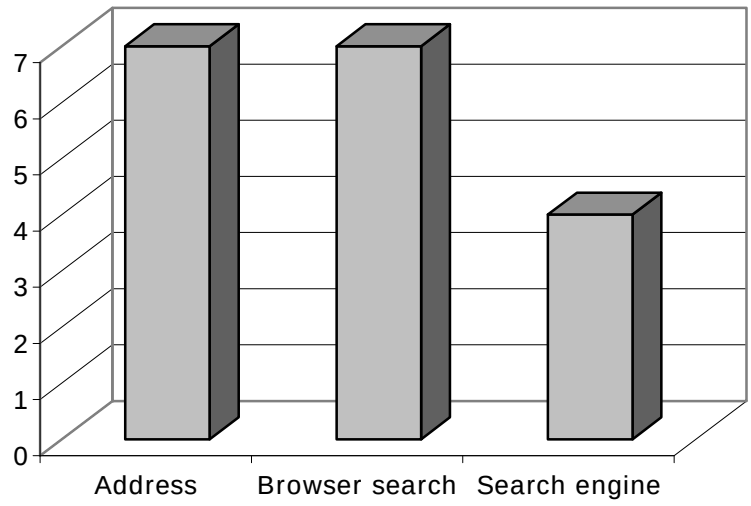

Chart 3
Initial search approach

Thirteen people reached the Department of Health home page, either through entering the address in the address bar, or through searching for the Department of Health web site. Just over 50% (7/13) then chose the site's own search option as their first method for finding the document, with a further five trying it when their preferred route failed. Of these only four used the document title as their initial search term, and one more tried this when their first attempt failed. The word diabetes was the preferred search term for five participants, three added more detail to this by specifying the diabetes national service framework, and one looked for national services frameworks. The three who tried the A-Z index all chose D as the first letter to check, but found no reference to diabetes in the listing. Four initially tried selecting options from the side navigation bar. No-one successfully found the document by following either the A-Z list or the side navigation bar.

Discussion

The majority of participants were successful in locating the document, even if not on their first attempt. The three who failed to do so failed for different reasons. One was struggling to find the document when she arrived at a related page (about the National Diabetes Service Framework) on the DoH

web site that said the report she had been asked to find was due for publication after the date she was doing the search. She concluded that it wasn't actually available, and decided to stop looking any further. One tried to go to the Department of Health web site, but in spite of trying several variations on the name couldn't guess that that it was www.doh.gov.uk and gave up. The third, used an international search engine to try to locate the 'Department of Health' without including that it was the UK Department of Health that was wanted, and was confused by the number of different Health Department sites in the results. When he failed to quickly locate the right one he gave up.

Nearly half of the participants were to some extent thrown by being faced with a browser with a different appearance and features to the one they had set up on their usual computer. These people normally interacted with the Internet through a combination of their pre-stored 'favourites' and a default search engine, usually the one provided by their Internet Service Provider (ISP). Some were able to cope with a different set up, but some didn't know the address of, or how to locate, the search engine they usually used.

Whilst the sample size was not large enough to look for statistical significance in the results, the number of years experience of using the web did not have any link with peoples reliance on a familiar set up. Both of the participants who had only been using the internet for a year found it difficult to work without their usual set up, however one of the three users who had seven years experience couldn't find their usual search engine without access to their favourites. There was a stronger relationship with the hours of use per month and reliance on the usual interface. No-one who used the internet for more than 30 hours a month had problems with the different interface or lack of their favourites.

Internet Explorer v5 defaults to the MSN search engine if the address entered in the address bar cannot be found; five participants used this feature when they couldn't access their usual search engine. Two were unable to progress when they couldn't access their usual browser set up, and were prompted that they could do a search using a default search tool in this way, rather than exclude them from the research at this early stage.

The majority of participants chose to try to find the document through the Department of Health web site, rather than by carrying out a search for the

document directly. Only two people carried out searches using an chosen search engine and looking directly for the document. Both were successful, finding the document quickly.

The 16 people who went to the Department of Health site had varying experiences. The search strategies of the people who tried this method fell into two main groups. The larger group searched for a combination of diabetes and national service framework. The smaller group tried searching for the document by name.

The results list from the Department of Health site search function highlight the search terms in the result listing, which by default are listed in order of relevance. Searching for the words diabetes national service framework results in almost 32,000 results being returned. There were 10 results listed on each page, and the document required did not appear in the first 3 pages of listings. Searching for the document by name returned a link on the first page of results. Participants did not however find it very clear that it was a link to the required document, reading the information carefully, and in some instances scrolling down to check other items on the list before returning to try the first link.

Several participants commented on how slow the search engine was. Two of the people who multi-tasked said that they started an alternative method because of the speed problems with the Department of Health search engine.

No-one went to the search help page to read about the search syntax that worked for the particular search tool used by the Department of Health site. If they had they would have discovered that unless the words are put in quotation marks they are not searched for as a phrase.

The people who tried to find the document by using the A-Z or side navigation bar found the experience frustrating and slow. The three who tried following the A-Z listing all tried D (diabetes) as their first option, and drew a blank in the D section. Two then went to the search page, the other tried the side navigation bar. In total four people tried the side navigation bar at some stage, all of these also tried following the publications link, some persisted in returning to the navigation bar several times trying different links and following them looking for possible links to the document.

One user became hopelessly lost having followed a link out of the Department of Health site without realising this.

The people who had the most success, i.e. found the document quickly, were those who searched for the document by the full title, either in a search engine, or through the Department of Health search page. Of the eight people that used an external search tool, (i.e. not the Department of Health site search function), five used Google and three used Yahoo.

The quickest results were obtained by the people who entered the Google web address in the browser address bar to reach the search engine, and then searched for the document by name. These participants found it at the top of the list on the first page of results. The slowest, and most frustrating experience for the participants, was using either the A-Z or side navigation bar options at the Department of Health site, no-one found the document using either of these two routes.

The failure of everyone who tried to use the navigation routes offered, as opposed to the search engine available, suggests that there were design faults in the sites navigation structure. The participants in this study were not thinking in the same way as the web designers, as indicated by the selection of 'D' in the A-Z expecting to find information on Diabetes grouped in some way.

Of particular interest was the finding that even when people searched there was a tendency not to search for the title of the document that they had been asked to find. People who initially used a search engine were much more likely to search for the document by it's full name. Half of those using this approach searched in this way, compared to only a third of those who searched from the Department of Health site.

Conclusions

The majority of participants first choice when faced with the task of finding a specific document, produced by a well known organisation, was to go to (or try to go to) the organisations' site, and to locate the document from there. The most successful way of finding the document however was by the efficient use of a search engine. Very few participants made informed

choices about the search engine they used (if they used one) with a high proportion normally relying on the default provided by their ISP.

The literature does not seem to contain much information on peoples' initial choice of search technique, or on how they chose search engines or search terms. This was only intended as a small scale research study to see what initial search choices people made. It would be interesting to repeat this experiment with a larger number of participants, and to include a post activity semi structured interview to discuss why they had made the choices they did.

The heavy reliance by some participants on a familiar home page and the availability of favourites show the importance of familiarity to the users. Along with comments received in the debriefings it also suggest that a lack of understanding of the advantages of developing search strategies, and poor searching skills are the main underlying factors for the problems people encountered.

Many people never learn about the Internet or the web, they just set up an internet service, and self learn what they think they need to know. As this research has shown this approach doesn't develop effective searching or problem solving skills, and relies heavily on the availability of familiar starting points and favourites. There was also a lack of understanding of search tools, and of even quite basic searching skills.

Some participants did not have an understanding of the structure of web addresses which meant that they were unable to guess the DoH web site address if they did not know it, and did not have access to their favourites.

If people are to be able to use the web quickly and efficiently as an effective information retrieval tool, as opposed to a recreational tool to surf the internet, they need to have both an understanding of the medium and the tools, and the skills to use them effectively.

References

Abels, E; Domas White, M & Hahn, K.	1997	Identifying user-based criteria for Web pages Internet Research: Electronic Networking Applications and Policy 7 (4) pp 252-262
Battleson, B; Booth, A & Weintrop, J	2001	Usability Testing of an Academic Library Web Site: A Case Study. The Journal of Academic Librarianship 27(3) pp 188-198
Bond, C	2003	Nursing Informatics for All. 8th International Congress on Nursing Informatics. IMIA. Rio De Janeiro
Cockburn, A & Jones, S	1996	Which way now? Analysing and easing inadequacies in WWW navigation International Journal of Human-Computer Studies 45 pp 105-129
Cockburn, A & McKenzie, B	2001	What do web users do? An empirical analysis of web use. International Journal of Human-Computer Studies 54 pp 903-922
Davis, E & Hantula, D	2001	The effects of download delay on performance and end user satisfaction in an Internet tutorial Computers in Human Behaviour 17 pp 249-268
Diaper, D & Waelend, P	2000	World Wide Web working whilst ignoring graphics; good news for web page designers Interacting with Computers 13 (2000) pp 163-181
Greenberg, K	2000	Search Patterns Media Week 10 (35) pp 72-75
Light, A & Wakeman, I	2001	Beyond the interface; users' perceptions of interaction and audience on websites Interacting with Computers 13 (2001) pp 325-351
Muylle, S; Moenaert, R & Despontin, M	1999	A grounded theory of World Wide Web search behaviour. Journal of Marketing Communications 5 pp143-155
Reed, W et al	2000	Computer Experience, learning style and hypermedia navigation. Computers in Human Behaviour 16 pp 609-628
van Schaik, P & Ling L.	2001	The effects of frame layout and differential background contrast on visual search performance in web pages. Interacting with Computers 13 (2001) pp513-525

Zakon, R 2001 Hobbes' Internet Timeline v5.4
http://www.zakon.org/robert/internet/timeline/
Accessed online 2/11/2001