Findability and Information Seeking Behavior

Findability is how easy an object is to find, or how easy a system or environment is to navigate (Morville, 4). Information seeking is the process through which people gather all the available information and sort through it to find information which is relevant to their current information needs. This process involves not just fact finding but also problem solving (Huston, 229). Information retrieval is a concentrated effort to retrieve as much of the relevant information and as little of the non-relevant information as possible (Borlund, 1).

Bates states that humans gather most of our information passively, and that we arrange our environments so that the need for active information seeking is minimized (Information Seeking and Searching, ¶ 5-10). We tend to put forth as little effort as possible and often settle for information that is only good enough, instead of expending the effort required to find the best or most correct information. Morville points out that Web users tend to operate on the Principle of Least Effort, that "each individual will adopt a course of action that will involve the expenditure of the probably least average of his work (by definition, least effort)" (55). Web search queries tend to use as few words as possible, and advanced search features are almost never used. In fact, only 35% of queries applied to search engines appeared to be of the type search engines traditionally supported (Keller et al, 1001).

What this means for Web sites is that they should be as findable as possible, because Web designers can't expect users to expend a lot of effort in looking for their site, or in navigating the site. Users don't want to spend a lot of time on active information seeking

and information retrieval, so if a site is not found quickly, it is likely to not be found at all.

Keller et al found that the majority of activity on the Web consists of repeated tasks, many of these monitoring tasks or transactions such as checking e-mail (1009). We are creatures of habit who don't often break out of our established habits. However, most information systems are designed for active information seeking. The site architecture usually requires to user to put forth a significant amount of time and energy to learn the structure and how to navigate it. If the structure is too dense or too confusing, the user is likely to give up and go to an easier site.

This means that system navigation should be as transparent—easy to understand and use—as possible, because Web users are not likely to spend much time at a site that is confusing or difficult to use. Users return to what they know, even if it is not the best tool for the job; for example the large number of navigational queries that show people use search engines as navigation mechanisms, even though this is not what search engines are designed for and browsers have much better ways for navigating built in (Keller et al, 1001). Once users get used to doing something a certain way, they will usually return to it over and over again, even if better or even easier ways exist, because finding and learning these ways takes effort.

Morville also points out that many users will not start with the home page to a specific site, so making sure the site is easy to navigate from the homepage is not the only concern. Another question that needs to be considered is whether users can find what they need from wherever they are (8). Consistent navigation that appears on every page, such as header or footer or side links, is one way of accomplishing this. Breadcrumbs or

some other way to let the user know at a glance where in the hierarchy they are is also useful. For large sites, a site map, easy to get to from anywhere on the site, is a good idea. What must also be considered is how easy different sections of the site are to find from outside the site. The more points of access a site has, the more likely a user is to find it.

Information seeking and information retrieval are major concerns a designer must consider when constructing a Web site, but they bring up a third concern, that of findability. An understanding of information seeking behavior can help designers build sites that easy to find and easy to navigate.

References

Bates, M. (2002). Toward an integrated model of information seeking and searching.
*The fourth international conference on information needs, seeking and use in
different contexts*. Retrieved August 30, 2007, from
http://www.gseis.ucla.edu/faculty/bates/articles/info_SeekSearch-i-030329.html

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American society for
information science and technology*, August 2003, 913-925. Retrieved August 30,
2007.

Choo, C., Detlor, B., & Turnbull, D. (2000). *Information seeking on the Web: an
integrated model of browsing and searching*. Retrieved August 30, 2007, from
http://firstmonday.org/issues/issue5_2/choo/index.html.

Kellar, M., Watters, C., & Shepherd, M. (2007). A Field study characterizing Web-based
information-seeking tasks. *Journal of the American society for information science
and technology*, 58 (7), 999-1018. Retrieved August 30, 2007 from Wiley
InterScience.

Morville, P. (2005). *Ambient Findability*. Sebastopol, CA. O'Reilly Media, Inc.

Architecture of the Getty Web Site

There are two navigational constants on the Getty site, the tabs across the top and the shortcut links across the bottom. This means that every part of the site can be accessed fairly quickly from anywhere else on the site. The tabs also act as a sort of "you are here" map by highlighting the section of the site currently being viewed. The subsections are listed in a space beneath the tabs, with the current subsection highlighted. The subsections which were divided further have navigation links along the left side, with breadcrumbs listed beneath the tabs to indicate where in the web site the page is located. In addition, there is a search function, and a site map that lists each section and its major subsections. The search bar included is on every page, and some of the individual sections have their own search features.

Overall the site is easy to navigate and consistently provides users with indicators of where they are. The only major drawback is the layout of the site is designed for a wide browser window. The search bar and the last few tabs, as well as the content of some of the pages, were usually hidden off the side of the window. Having to scroll sideways could discourage users from using the search function, and without scrolling the user wouldn't see the last few tabs.

One of the most prominent parts of the site is information about visiting the Getty Center and the Getty Villa, including links to event calendars, exhibitions, and resources for teachers and students on the front page. This last implies that students and school groups are among their most common visitors. Otherwise, the groupings of information shown by the tabs seem to assume that a visitor to the site would be looking for information that is scientific in nature. Two of the tabs are labeled "research institute"

and "conservation institute," and two of the subcategories under the tab labeled "museum" are "education" and "research and conservation." The majority of the information is written for the layperson, which means the site is designed for the general public. The first three links under the "conservation institute" tab are "science," "field projects," and "education," and likewise the links under the "research institute" tab are "exhibitions," "conducting research," and "scholarly activities." Much of it is aimed at people who might be interested in participating in the research or conservation efforts, but are not yet involved. However a frequent user of the site could easily bypass the information aimed at newcomers and go straight to other parts of the site, such as the newsletter, events calendar, or bookstore.

One of the most frequent links is to the "events calendar." In addition to being in the list of "useful links" at the bottom of most pages, it is also prominently displayed on the front page and is the second link under the "visit" tab. Likewise, the bookstore is linked from the front page as well as under the "museum" tab. Many links are repeated like this. The design seems to anticipate the most frequently used areas and display them prominently, as well as put them in some sort of logical structure using the tab navigation. The structure seems to anticipate that visiting one section of the site might prompt users to visit other sections, and so makes it easy to access those sections without having to return to the front page.

Newspaper Search Engines

I decided to search for articles on bipolar disorder in teenagers. In the New York Times search, the query "bipolar teens" returned 13 documents, three of which were relevant. The query "bipolar youths" returned too many documents for me to manually evaluate (85), but I found four relevant documents in the first 20 that were not returned in the first search. Based on this, both the recall and precision of the first search seemed quite low.

The LA Times database returned no documents for "bipolar teens" or "bipolar youths," and five for "bipolar children," three of which were relevant. A search on "bipolar" alone returned seven. Since no additional relevant articles were returned on the search for "bipolar," I would say that the recall of the search on "bipolar children" was high, while the relevance was fairly good, since over half the returned documents were relevant.

The biggest flaw in both of these searches is that they're full-text searches, with no way to narrow the search by proximity, and no option to search on subjects or keywords. Many of the articles in the NY Times search mentioned bipolar in passing but were actually about something else. The advanced search offers a title search, but this would only be useful in increasing precision and would actually severely decrease recall, because most of the relevant articles I found did not mention the search terms in the title and had no consistent wording between them.

The best way I can see to increase relevance and precision would be to include some manual indexing of the articles, by associating them with some subject headings or keywords. This way an article that happens to mention bipolar but is actually about

schizophrenia would not be returned during a subject search for "bipolar." The drawbacks to this are the increased cost for the database and a possible delay in getting the articles available, and that users would either have to guess at the subjects or have to scroll through a list. Also, for large databases such as these, the cost of manually indexing every article might be prohibitive.

A more economic solution would be to enhance the search function to include word proximity and perhaps word frequency features, or a way to limit the search to the first paragraphs or first few paragraphs, since news articles usually mention the main subject early on. Being able to specify a proximity of, say, five or ten words, means that an article that mentions "bipolar" in one paragraph and "teen" two paragraphs later, which is not likely to be relevant, would not be returned. Frequency would mean that an article that mentions "bipolar" in passing is not as likely to be returned as an article that is about bipolar.