Earning online trust

Seth A. Maislin

The problems of online searching, why it so frequently ends in failure, and what could be done to improve the situation.

When we conduct online searches, our burden is one of trust, and there is little precedent for awarding it. Effective searching requires the user to know search logic and information design, to recognize and use interactivity, to grasp context and scope instantly – and never to give up. Nevertheless, there is a gap between what happens when users search and what users think happened. For example, why do most users often equate bad results with search failure, and not bad content?

Consider how we get information without computers. Sometimes we ask people. If the answer is 'I do not know', we accept this as truth and ask someone different. As an alternative, we might look for answers in textbooks. If an adequate answer is not there, we assume the book lacks the answer we want.

We generally trust people, and we generally trust books. With computers, blame abounds. We mistype words and second-guess our spelling, the search engine does not handle synonyms or suffixes, web pages are outdated, browsers are incompatible, and most authors are amateurs in the medium. We are surrounded by failure and, consequently, the simplest possibility – that there really is nothing to find – is either totally unrecognizable or else completely dismissed.

Popular computer culture teaches us that everybody can publish online, and that everyone does. Unlike a book with a finite number of pages, we view the World Wide Web as an 'infinite book', so huge that even the search engines miss 5 out of 6 pages. No wonder people get a little confused. The problem lies in a grave misunderstanding of search scope. We expect web search engines to search 100 percent of the world’s online content, and we expect every website to map all relevant content. This is a foolish assumption, and yet even the most experienced web users make it. Ask yourself a few questions. When searching, how do you respond when you see the 'No results found' message? Do you search again? Do you search with a completely different term, or with only a slight change? Do you try another website or use a different search engine? How many searches will you perform before you surrender to asking a human – two? ten?

Fixing the user’s mistake

After an unsuccessful search, many users forgo trying something significantly different and instead refine their original query only slightly. For example, a user who searches for 'web sites' may then search for the concatenated 'websites' or the singular 'web site'. In the extreme, users actually retype the original query, with an expectation of improved results. (This fact is based on actual query data from America Online and Lycos.) Yet more alarming is that some misspellings are more common than the correct spellings: 'millenium' is typed twice as often as the correct 'millennium'.

This behaviour can be described by defining ‘language clumps’, which are collections of text strings that are similar in fundamental ways. Clumps can include alternate spellings and close (or common) misspellings; various conjugations and tenses of the same verbs; synonyms, including dialect-based variations; and so on. Few search engines incorporate adequate clumping. Even spell-checking, which can be accomplished with a surprisingly simple algorithm, is rarely performed by search engines. (Worse, many web catalogues and sites themselves include misspellings!)

There is no magic prevention against misspellings, but some search engines do attempt to intervene. Yahoo http://www.yahoo.com suggests alternative word spellings, although only if the original query retrieves a nil result. Lycos http://www.lycos.com and America Online include common misspellings in their database infrastructures, hoping invisibly to connect the user to desired content. Consistency of language, while admired and often necessary in writing, is the enemy of searches: it prevents any natural language clumping. As an extreme example, consider again how often 'millenium' is misspelled. Spelling this correctly 100 percent of the time fails 67 percent of your users.

Handling misspellings is only one small part of clumping, however. Consider users who type ‘directories’ when they mean ‘folders’, ‘medicine’ instead of ‘medical’, ‘theater’ instead of ‘theatre’, or ‘cd-rom’ instead of ‘CD-ROM’. The more strictly consistent your documentation, the greater the burden when it comes to effective searching. Each item in the author’s style sheet is a reason to apply an additional language clump.

Handling ambiguity

While documentation is often too consistent, the queries themselves are frequently too general or ambiguous for obtaining effective results. A good book index can provide both options for refining a search (sub-entries) and synonyms (see-type cross-references), but most search engines just output all matching results.
Consider a user interested in Rochester. Of course, there are several places named Rochester in the world. The index to a printed US atlas, for example, would make this clear by providing state names. (Formally, the state names serve as disambiguating modifiers.) Online, however, a search for the word Rochester would not categorize results. Further, those results would sort alphabetically, not geographically, and would be undifferentiated from any literary or historical characters named Rochester.

Instead, writers working with a limited set of documentation (such as a CD-ROM product or a single website) can improve results by providing intermediate results. For example, consider an electronic atlas of the United States. A search for Rochester should call up a single Rochester page, from which the user can choose one particular Rochester, or all of them. This is known as iterative searching. Writers are familiar with ambiguities within their fields of expertise, so creating such pages would be a natural extension of their abilities.

Scope is another source of ambiguity. Users rarely know how much they are searching. Studies demonstrate that users intuitively believe that searches performed from the home page of a website have much greater scope than searches performed from deeper inside the site. For example, users expect different results from the search at http://www.microsoft.com than from that at http://www.microsoft.com/windows95/whyupgrade/top10.asp. This rarely happens.

Indexing what isn’t there

Search technologies are still in their infancy. With the growing quantities of information available today on the Web, there are no guaranteed solutions. Further, every project has its own requirements. Until online trust is developed, whether by improved tools or from a greater understanding of the online environment, information specialists should remain aware of the distinction between printed and online/CD-ROM documentation: the need to index what isn’t there.

A good analogy can be found in the descriptions for cookbook recipes. If a recipe is missing a particular ingredient, that absence is worth noting: meat (vegetarian), milk (non-dairy), sugar (low-calorie). This is a subjective decision, too; I have never seen the index entry ‘turnips, absence of’.

When users expect results, ‘Query not found’ and ‘No results available’ messages are unsatisfying. Results should be invented that handle any reasonable queries, even if the answers are outside the product's scope. For example, a software product that converts text to hypertext may not handle animated graphics. In its help system, advantage should be taken of queries like ‘animation’, ‘animated’, ‘movies’ and ‘QuickTime’ (an Apple-copyrighted movie format) to promote other products or give advice: ‘For animation help, visit the OptiAnimator web page’.

Another missed opportunity is the bookstore search engine partnered to an online portal, which unthinkingly drops users’ queries into a search form: ‘Buy Books About Football East Liverpool Schedule’. Imagine the power of language discrimination and clumping here!

In summary, online users make two assumptions: (1) that everything is always available, and (2) that the online medium is saturated with human and computer errors. It is the responsibility of writers and indexers to stop those assumptions from diminishing the user experience, and to earn the users’ trust.

Seth Maislin is a directory manager at Lycos, Inc. and sole proprietor of Focus Information Services. He provides indexing, information architecture and consulting services to public and private audiences, regularly presents workshops at conferences, and indexes books and web documents. He is an elected director of the American Society of Indexers (ASI), chairs the ASI Indexing Technologies Committee, and serves on the Boston STC Council. Some of Seth’s ambitious projects include indexing America Online, building taxonomies for Lycos, and cataloguing clip art for Mattel. Fax: +1-781-641-1981; email: smaislin@world.std.com