When Bad Searches Happen to Good Searchers

Think of all the times you just knew you had the perfect search strategy, only to find out that it didn’t work. The answer that you thought would be so easy to locate turns exceedingly elusive on you. The database you were sure covered the topic doesn’t. The search engine you’ve relied on in the past for similar queries drops brick walls in front of you. The Web site you’ve used before has disappeared. The specific page of the Web site can no longer be accessed without a subscription.

Sometimes bad searches happen, not because of technological glitches or editorial policy decisions, but because the underlying assumptions or basic data you’re working with are flawed. You’ve got the wrong company, the wrong country, the wrong industry, or the wrong person. You’ve been assured that the company trades on a public stock exchange. It doesn’t. Your client is positive an event happened in Taiwan. It didn’t. Your patron swears he’s interested in the fast food industry, but really it’s snack food. The marketing director whose name you’ve been given left the company last year.

The one we generally hate to admit is bad searches due to operator error. We mistype (I’ve lost track of the times I’ve typed dialgo instead of dialog or but instead of buy), choose an improper database, truncate in a search engine that doesn’t recognize truncation symbols, or enter a URL in a search box instead of the address box. As professional business researchers, we like to feel superior to the general public. We think we’re above the fray when it comes to performing quality searches. Most of the time that’s true. However, we’re not always perfect and might as well admit it.

With all the things that can go wrong—think Murphy’s Law here—we should really be congratulating ourselves that we can get good answers from online searches. Instead, we beat ourselves up over the ones that got away.

BUSTED BY BOOLEAN

Long-time searchers were weaned on Boolean logic. These are the people who automatically transform research requests, at least in their heads if not on paper, into Venn diagrams. Then there’s a whole new generation of searchers who are learning Boolean as Web search engines portray them. These folks automatically add plus and minus signs to search terms. The use of AND and OR in daily language sometimes reflects a true Boolean relationship, but not always. “Pass the salt and pepper” means you want both handed down the table to you. If your bank tells you it has checking and savings accounts, it means they can offer you both. However, you may not want both. You may want to choose checking OR savings.

When it comes to business research, there are terms that automatically go together as a phrase. That doesn’t mean the phrase is true in a Boolean sense. Take “import and export” as an example. If someone wants to know about Canadian import and export of timber, that person isn’t using AND in its Boolean sense. Most likely, the request is for the import of timber into Canada.
or the export of timber out of Canada. On the other hand, it might be they want to know about timber exported by Canada or imported into Canada. Either way, it's a Boolean OR, not an AND. If a market research report says it covers the “Banking and Finance Industries,” it contains information on the banking industry along with other types of financial institutions. Retrieving information from this report is an OR relationship. It will discuss banks OR other financial institutions, since it may devote a chapter to each. Likewise, the person who asks about sales and selling of toys is asking about (sales or selling) and toys. The ANDed concept is the toys; the sales concept is an OR relationship. One would hope that “sales and service” would be true AND, but it’s more than possible to get the sales without the service. That, in Boolean-speak, is sales NOT service.

NOT can be a knotty problem, occasionally even naughty. It's undoubtedly the most misused Boolean operator. Some search engines render it as a simple NOT, others insist upon AND NOT, such as AltaVista. More importantly, the indiscriminate use of NOT will frequently eliminate results that you actually want. If you're looking for Google but not AltaVista, NOT'ing out AltaVista will also throw out Web sites that state, “Google and AltaVista are search rivals, but this site only discusses Google.” In traditional online searching, it throws out articles that include a sentence similar to that. The best use of NOT is to disambiguate search terms. You're searching SLA the library association rather than SLA the Symbionese Liberation Army or SLA the Service Level Agreement. You can NOT out the two phrases that clearly aren't connected with the Special Libraries Association. You can also use NOT to make sure you don't see things you've already looked at. This works particularly well with search engines that let you build sets, such as Dialog.

**SOURCING THE STONE**

An early adage of online searching was, “Don't enter a common word into a specialty database.” In other words, if it's American Banker, don't enter a one-word search on banking. If you want to know about recent information technology trends in the banking industry, use information technology and trends (ORed with likely synonyms). Don't include banks or banking in the search strategy—the banking concept is assumed.

This is true whether you're searching on a free Web site or using a fee-based premium service. The financial implications are different, since the pricing structure of fee-based services frequently penalizes the searcher who uses unnecessary terms. Web search engines also penalize you, but it's the time sink rather than out-of-pocket expenses that will do so. By entering common words, you dilute results to the point that you need to drill down through pages and pages of search results to identify relevant sites.

**COUNTING CROWS**

It's long been a competitive talking point. My database has more records than yours. My search engine spiders more Web sites than yours. What is too often overlooked is that quality and quantity are not synonymous terms. Information professionals are accustomed to pointing out to producers and, now, to Web search engine representatives, that 15 bad answers are not better than one really good, correct answer. Why do these companies insist upon continuing to equate quantity with quality?

The answer is actually pretty simple. It's easy to count; it's hard to evaluate quality.

**ASSIGNING BLAME**

When online searching started, there were companies responsible for producing information for sale. These highly structured databases, boon that they were to business researchers, were not always constructed to the exacting standards of information professionals. They boasted of their indexing and categorization schemes, but information professionals were suspicious. The attitude was, “I'm the searcher, I'm OK; you're the producer, you're not OK.” Any misstep, mis-categorization, typographical error, or mis-entered term resulted in the database producer being castigated by customers demanding that the error be fixed. Every long-time business researcher will be happy to share some of the more egregious things encountered in searches of established databases. The fact is, however, that those errors were few, considering the large number of records in these databases.

Reva Basch wrote an article on “The Seven Deadly Sins of Full Text Searching” in the August 1989 issue of DATABASE. Those sins were duplicity, verbosity, wimpiness, irrelevance, sloppiness, hyperbole, and obfuscation. In July 1992, she followed up on the sinfulness theme by writing “The Seven Deadly Sins of Online Services,” published in ONLINE. This time around, she identified avarice, inconsistency, aloofness, uncommunicativeness, pride, inadequacy, and ignorance.

When it comes to Web searching, life changes. Web pages found by search engines are not “products.” They have no “database producers.” They're not for sale. They're not designed with classic information retrieval policies and practices in mind. In the Web search world, there is no blame. There may be sins and sinners, but forgiveness is more common. Web searchers are not as concerned about duplicate Web sites showing up in search results, probably because they're not paying for them. Professional searchers are more likely than end-users to avoid the operator error of not scrolling down through several pages of results, knowing that the answer may not be in the top group of search results. Irrelevant results aren't so onerous anymore, either, because it's cost effective to do multiple searches using several Web search engines, something that fee-based services may obviate against.
There’s still some “pridefulness” apparent in Web search engines. Google, for example, if you misspell a word, will suggest, “Did you mean – correctly spelled word?” If your misspelled word retrieved no hits, Google goes a step further and re-does your search using the correctly spelled word. Uncommunicativeness remains an issue, if not a sin, as most documentation is either missing or inadequate for Web search engines.

**SOLVING BAD SEARCH**

Are there answers to searcher-induced mistakes? One possibility is templating the search. This is close to impossible in general search engines. I’m amused that the same types of people who, 20-odd years ago, criticized Dialog’s user interface for its stark question mark as an indication that you should type in a search query, now applaud Google’s minimalist search box as the ultimate in search technology. Both invite bad searches by not providing guidance. Both have few alternatives, searching against a highly diverse, heterogeneous body of knowledge.

Database producers and electronic publishers which put their information directly on the Web have a real advantage in that they can capitalize on file structure to guide searchers to a good answer. Take the Kompass company directory databases (www.kompass.com). The site provides search boxes for company name, city, postal code, country, trademark, number of employees, gross revenues, share value, year established, executive names, and job titles.

Even on Web sites with more diverse information, it’s possible to provide guidance to users. Standard & Poor’s NetAdvantage provides an example. It includes 12 databases, each somewhat different, all published by S&P, each with its own search template, based upon the structure of that particular database. S&P assumes searchers have some knowledge of the data sitting behind various search boxes, such as ticker symbol, CUSIP number, or beta. However, if you don’t, there’s an Information button you can click to go to a glossary that explains what you’re searching on. What it doesn’t do is save you from entering the wrong ticker symbol, CUSIP number, or beta. That’s still up to the searcher to do correctly.

**HAVING A BAD SEARCH DAY**

There are certainly days when it seems no matter what you do, how many Web search engines you use, which databases you access, what search strategies you try, you can’t find the answer. This is probably the time to take a deep breath, leave your computer for a while, and do something else. Come back fresh to the search problem and try alternative approaches. First, think about what you might have done wrong. Spelling? Boolean? Synonyms? Different ways of phrasing the search strategy? Trying to use functionality that’s not implemented in the search engine you’re using? Not capitalizing a Boolean operator if that’s required by the search engine? Entering a search string in the wrong order so that the Boolean operators are combining terms in ways you didn’t intend?

Second, try running the search in something like Dialog’s DialIndex (File 411) to suggest alternative databases in topic areas you hadn’t realized covered the subject. If you’re addicted to Web search engines, run the search in an engine that helps disambiguate the phrases, such as Northern Light or Teoma. These create folders that suggest other ways of looking at the topic.

Third, re-examine the underpinnings of the search problem. Are they correct? If you’re starting from a faulty premise, you’re not likely to come up with an answer. If a company is headquartered in Spain and you’re looking at a Japanese directory, you’re not going to find the needed information.

The Web has made it easier to make mistakes in searching. It has also made it easier to correct mistakes and to try many different approaches to a search problem. Traditional online search engines and databases have raised our expectations. Both are necessary for comprehensive business research.

Maryde Ojala (maryde@xmission.com) is the editor of ONLINE and a long-time business researcher.

Comments? E-mail letters to the editor to marylde@xmission.com.