

SEARCH

FATIGUE

Finding a cure for the database blues

by Jeffrey Beall

It is a feeling of frustration and dissatisfaction experienced by searchers trying unsuccessfully to find desired information in a database. It results when searchers cannot find what they are looking for and when they repeatedly get results that do not match their information needs. This is search fatigue.



You've probably experienced search fatigue yourself: You try several searches to find information you think ought to be present in a database, but no matter how many different ways you enter your search, you fail to find what you're looking for.

As online databases grow in size and as the simple search box, such as the one popularized by Google, becomes the norm, search fatigue will become an increasing problem. Fortunately, librarians are well-positioned to help database searchers overcome search fatigue by designing and implementing databases

and search systems that rely on value-added features that provide searchers satisfying and comprehensive search results.

The chief cause of search fatigue is a reliance on keyword searching. A number of inherent flaws plague keyword, or full-text, searching. One major flaw in keyword searching is that of synonyms. For example, a searcher looking for information on false teeth will probably miss all the resources that use the term *dentures*.

A more extreme example of synonyms is the term *Atlantic cod*. There are at least 60 different terms for

this species of fish, including *codling*, *Newfoundland fish*, *schrod*, *shoal fish*, and *winter fish*—all different names for the same fish. But very few resources likely use all 60 terms; in fact, most probably use just a single term. So any keyword search on a single name will likely miss all the other resources that use different names. Users who refer to Atlantic cod by one of its less-common names will probably find little information on the topic. In this way, keyword searching shortchanges the minority who use the less common term and favors those who use more common terms for a given topic; it also yields incomplete search results.

Another major weakness of keyword searching is its inability to deal well with homonyms. One example is *leaks*. There are at least two major meanings of the word: One refers to an unintentional hole that allows something to escape, such as water from a pipe or air from a tire; but people also use the term to refer to supposedly secret information that has been divulged to the mass media. A keyword search on leaks is going to pull up resources without distinguishing between pipes and politics. Searchers will have to wade through the results and determine which documents match their needs—a time-consuming process that results in search fatigue.

Keyword searching also functions poorly in searches that use common terms or names, since these retrieve many results and are difficult for the search software to rank by relevance. For example, searching for information about Los Angeles or a common name such as Mike Wilson will retrieve abundant results in most systems, and many of the search results won't have anything to do with what the searcher is looking for. Recently I needed to find information about someone

named Michael Ensign. But because there is an actor (a different person) with that name, most of my search results in Google were about the actor, since those were ranked highest by Google. This ranking caused me search fatigue because it required me to look through many results, and ultimately I was unable to find the information I needed.

Another weakness of keyword searching is its inability to effectively search vague terms and concepts. It's difficult to get good search results for searches about *life* or *health* because these terms are so imprecise. Searching such terms generally yields very large result sets, sets that are often too large to sort through. Large result sets are one of the chief causes of search fatigue.

Keyword searching also generally fails to pull up documents in languages other than that of the original search. For example, if you search for something using a French term, most of the results will be in that language. The exceptions include documents written in both French and English and documents that contain cognates (words spelled exactly the same) in both languages. But generally, keyword searching is monolingual; this can be a source of search fatigue by eliminating relevant documents. A salient example is Brazil: In Portuguese, the national language of Brazil, the country's name is spelled *Brasil*. So a keyword search for *Brazil* will probably exclude most of the documents that originate from the country itself.

Relevancy ranking itself can be another cause of search fatigue. Relevancy is a computer's way of ranking what it thinks are the most relevant search results, listed in order from the top of the retrieval display. But it's difficult for a computer to know what is most relevant. Moreover, different search systems use different



algorithms to determine relevancy, so what appears at the top in one system may not rank that high in others.

The whole idea of relevancy started with keyword search engines. Before keyword search engines,

metadata is that it compensates for all the weaknesses of keyword searching. A controlled vocabulary provides consistency for subject headings, so the person searching for information about false

A searcher **cannot find something if it isn't there**, but it may take the searcher a **fatigue-filled hour to come to this conclusion.**

people searched metadata-enabled search engines and had their results ranked alphabetically. Alphabetical sorting is about as natural an order as one can get, because it's easy and we're accustomed to it. But keyword search engines cannot sort results alphabetically because they don't know what elements to base the sort on. Instead, they use relevancy ranking, which is a mysterious, inconsistent, and unnatural means of sorting search results, and a source of perpetual search fatigue.

Some search interfaces are so poorly designed or so confusing to use that the search interface itself can be a cause of search fatigue. Some search engines default to the Boolean "or"; others default to the Boolean "and." Moreover, poor data quality in a database, such as spelling and typographical errors, contributes to search fatigue because it can cause some resources not to appear in the search results list, rendering them virtually unfindable. Data in a database is also often missing or incomplete. A searcher cannot find something if it isn't there, but it may take the searcher a fatigue-filled hour to come to this conclusion.

The searcher himself can also be a source of search fatigue. A searcher may consistently misspell a search term, turning up only resources that contain the wrongly spelled term. The searcher may also be unfamiliar with keyword searching and not know how to effectively use even the most simple search interface. A common error among novice searchers is to enter too broad a search term, such as *art* when they really want information about, for example, 19th-century French art.

Keyword vs. metadata enabled

Sometimes keyword searching performs well. For instance, if you're searching for a rare word in a large database, a keyword search is probably going to be a quick and easy way to find that term. A metadata-enabled search engine is one that searches metadata rather than full text to generate search results, such as an online catalog. The great advantage of

teeth is referred to dentures. And every document that contains information about false teeth or dentures in any language or by any other name is assigned the subject heading "Dentures" so that they all will be retrieved in a search on this topic. In this way, the search is comprehensive, and no relevant information is excluded from the results.

To better understand the strengths and weaknesses of keyword and metadata-enabled searching, it helps to divide searching into casual information-seeking and serious information-seeking. Keyword searching can be



Enrich Your Work with
Active Older Adults
at the Second Annual

Lifelong Access
Libraries™
Leadership
Institute

Late Summer 2007 (dates TBD)
University of North Carolina, Chapel Hill

Learn from examples of excellence and innovation drawn from the national Lifelong Access Libraries Initiative, a program of Americans for Libraries Council. This intensive, four-day workshop will build knowledge, enhance skills, and inspire new ideas.

Mid-career librarians with a demonstrated interest in reaching out to older adults are encouraged to apply.

Application Deadline: April 13, 2007

Applications available at: www.lifelonglibraries.org

 AMERICANS FOR LIBRARIES COUNCIL

The great **advantage** of metadata is that it **compensates** for all the **weaknesses** of keyword searching and **helps** eliminate search fatigue.

adequate when a complete search result is not needed; when one or two resources, regardless of their quality, are sufficient; and when the information isn't a crucial need for the searcher. Keyword searching functions poorly, however, for serious information-seeking, which requires highly relevant and precise results. It involves searches that relate to scholarship in medicine, business, and other fields where exhaustive search results are needed that are not polluted with irrelevant data.

Gresham's Law

The shift that the library world is now going through from metadata-enabled searching to keyword searching is a case of Gresham's Law in action. Gresham's Law was named for Sir Thomas Gresham, a 16th-century economist. In those days, people would sometimes cut or scrape off some of the metal from coins, and Gresham observed that when different coins with the same face value are in circulation, people hoarded the better coins—that is, the ones with a higher metal content—and the less preferable coins with the lower metal content became far more common. Although all the coins had the same face value, Gresham found that people kept the good ones and used the bad ones for buying and selling. Another way of stating Gresham's Law is, "The bad drives out the good."

Many people think all types of searches have the same "face value." Keyword searching is becoming extremely popular and is essentially beginning the process of replacing metadata-enabled searching, such as online catalogs. If this process continues, metadata-enabled searching will become a high-priced specialty service, one that is not generally available. Keyword searching, with all its flaws and weaknesses, will dominate and become the only type of search available. We are observing Gresham's Law firsthand: Cheap and abundant keyword searching is beginning to replace metadata-enabled searching. The bad is driving out the good.

Because keyword searching is so prevalent, librarians can help searchers make the best of keyword searching by helping them learn how to maximize this type of search. We should teach patrons that keyword searching, despite its many flaws, does have some uses and can sometimes be an effective tool for information discovery and retrieval, especially in casual information-seeking.

However, search fatigue will certainly become more

common as keyword searching becomes the main means of information discovery, as metadata-enabled search engines become fewer and fewer, and as full-text databases start to be measured in terabytes and petabytes rather than megabytes and gigabytes. Librarians can work to preserve the high-quality searching that metadata and controlled vocabularies help provide. We can continue to devote resources to metadata creation and to metadata-enabled search engines, both of which will be crucial for information discovery in enormous databases. But the most valuable work that librarians can perform is to explain to searchers the great value of metadata and metadata-enabled search engines. Perhaps by doing this we can save metadata-enabled searching from the extinction to which it is now heading. ■



E-Z Drop

A NEW name in book returns!

Model EZ4646

- Easy to use drive up or walk up
- No pull down doors—easier and safer to use for AV or print needs
- 216 Marine Grade Stainless Steel
- Modern curved top—rotatable and sheds the weather
- All carts padded with disposable bottom tray
- Replace individual panels if damaged
- Push-button lock—no key!

Model EZ Drop Model 51

- One piece construction—easy assembly
- Inside chute can be adjusted up or down
- May be RFID compatible

EZ Drop FROM THE BIRCHARD COMPANY
(877) 4EZ-DROP
fax (909) 625-2208
P O Box 1438,
Claremont, CA 91711
*Will not fail due to corrosion
www.ezdrop.biz
ron@birchard.biz

Visit us at booth 4100 ALA D.C.