3-1-1994

Target, Freestyle, WIN: Searching Takes on a New Look

Carol Tenopir
University of Tennessee - Knoxville

ONLINE DATABASES

BY CAROL TENOPIR

Target, Freestyle, WIN . . . Searching Takes on a New Look

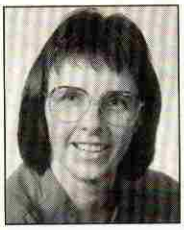
TARGET, FREESTYLE, WIN—are these the names of some new TV quiz shows? Or perhaps some new Las Vegas card games? No, they all are new search techniques available on major online systems. DIALOG, Mead, and West, respectively, are each wagering that these new non-Boolean, more natural language–like search methods will pay off for them in new users and increased usage.

Reacting to limitations

Although each of the three systems is different, they have some commonalities. DIALOG's Target, Mead's Freestyle, and West's WIN are each a reaction to the limitations of the standard Boolean logic search engines.

There are major limitations with Boolean systems. The first of these is also often a strength in reducing large set size, i.e., all concepts linked with AND must be present in a record. If a searcher enters tourism AND hurricanes AND Florida, an important article on the effect of hurricanes on tourism elsewhere will not be retrieved. Many marginal, potentially relevant items are lost.

The second limitation has to do with record display. Retrieved records are normally displayed in reverse chronological order or in a user-specified sorted order, such as alphabetically by author. Reverse chronological order is good for current event searching or updating an old search, and sorting is useful in providing an organized look at search results. Obvious false drops are mixed in randomly with highly relevant items, however, and the display order provides no clue to the potential relevance of any item.



Carol Tenopir is Professor at the School of Library and Information Studies, University of Hawaii at Manoa, Honolulu

Third, users must be conversant with Boolean operators and also, often, with commands and other "unnatural" input conventions. Boolean logic is counter-intuitive for most people ("Why does describing my topic more fully by adding something with 'AND' retrieve *fewer* items?"). Commands require training, practice, and perusal of documentation. All of these make online searching difficult, or at the least stifling and unappealing for many end users.

The new systems do not replace their Boolean precursors. Instead, they offer an alternative for both novice users and experienced searchers to locate relevant items that may be missed by a Boolean search. They are all proprietary programs, but they are also "associative" or "relevance" retrieval systems that find records most likely to be relevant to a search and display the items in order of likely relevance.

What the new systems offer

In the October 1, 1993 Online Databases ("The New Generation of Online Search Software") and that of November 1, 1993 ("Natural Language Searching with WIN"), some relevance systems were described in greater detail. In general, they search for all terms a user inputs (hurricanes, tourism, Florida in a sort of OR relationship) and retrieve records that have any of the terms.

To avoid the many false drops that "ORing" the terms together would cause, the systems employ some fairly sophisticated processing. Some of them do automatic stemming, stop-word checking, and phrase identification. All of them use a formula that counts how many of these terms occur in each document, how many times each term occurs in each document, each document's length, and how often each term occurs in the entire database compared with how many times it occurs in each document. Some give bound phrases or terms more weight in a system thesaurus than free-text words.

Records are displayed in order of likely relevance. As a user moves through the displayed list of items, the more peripheral items will be displayed. To make the process practical, each system has a cutoff number of items displayed, normally a maximum of 25–100.

Relevance search engines mean users don't have to remember Boolean operators, and some go one step further toward ease of use. Some are paired with natural-language interfaces that allow users to input any search statement (e.g., "What long-term effects do hurricanes have on tourism in Florida?") for machine deciphering.

Target

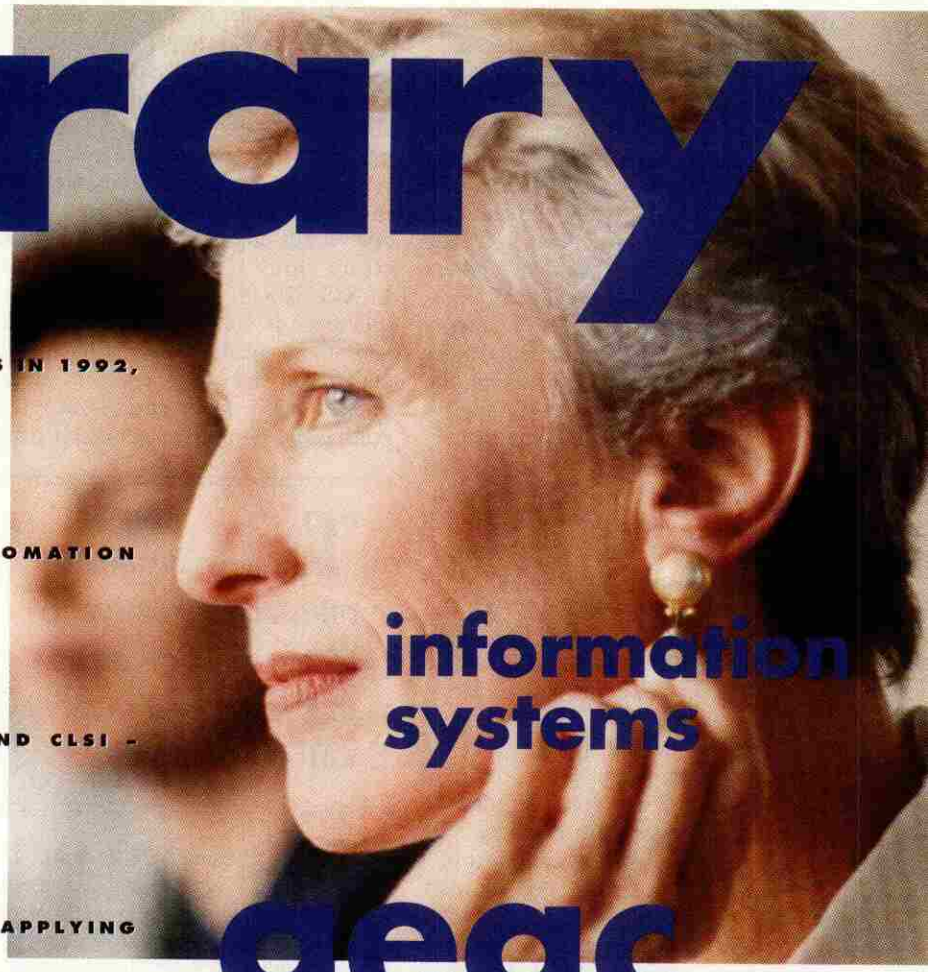
Target will work on a single file or on a OneSearch grouping. It works best on full-text databases, because searching reverts to the basic index (typically titles, abstracts, descriptors, and full text) and because the relevance formula depends on a reasonable number of words in each record. Relevance ranking may be needed most in full-text searching anyway.

After beginning in a database, simply enter "Target" to receive a prompted search screen. Alternatively, experienced users can input their search statement on the same line as the Target command. A Target search may look like this: Target dog cat food.

Target does not use commands or operators, but it is not a truly natural-language system either. You must enter only the terms you want to search, rather than a sentence describing the search. Also, consistent with DIALOG's long-standing philosophy, searchers control most of the search process rather than having the system control it with automatic features.

Several "enhancements" or refining tools help the searcher control results. Truncation must be specified to search plurals or other word stems (e.g., dog?), phrases are identified with single quotes (e.g., 'dog food'), and synonyms can be grouped with parentheses [e.g., (dog canine)]. Especially useful is the ability to flag critical terms that must be present for an item to be re-

library information systems



SINCE JOINING FORCES IN 1992,

TWO LIBRARY AUTOMATION

Advanced Systems

PIONEERS - GEAC AND CLSI -

Superior Service

HAVE BEEN QUIETLY APPLYING

Designed for Tomorrow

THEIR COLLECTIVE EXPERIENCE

Available Today

AND EXPERTISE IN DEVELOPING

LIBRARY SYSTEMS AND SERVICES

FOR THE 21ST CENTURY.

Now, by listening and responding to one of the world's largest group of library system users, Geac offers two remarkable new systems which far exceed the first-generation systems most libraries use today.

With a $10 million investment in research and testing, Geac now delivers two flexible client-server library systems with total open networking capabilities to meet every library's needs. Both are backed by Geac's uncompromising, comprehensive service and support program — ServicePlus.

Geac

The right time. The right choice.

320 Nevada Street
Newtonville, MA 02160