

## Viewpoint

## What's In a Space? FIND Out Now

*Phrase searching in different databases has always been confusing*

by Péter Jacsó

[Editor's Note: Péter Jacsó, a man with an opinion, is taking the first crack at writing for IT's new "Viewpoint" space. We invite others among you in the library and information professions who want to voice an opinion, address an issue, or put forth a commentary to query IT with a proposal. See editorial contact information in the masthead, p. 5.]

Phrase searching or compound term searching in online and CD-ROM databases has been enigmatic for many casual searchers. Carol Tenopir<sup>1</sup> and Marydee Ojala<sup>2</sup> have written about the differences of various systems in interpreting search phrases. There have been two extremes in interpreting user queries that consisted of two or more words separated by a space.

#### The Ultrarestrictive Interpretation

DIALOG represented one extreme. It interpreted the query "select INFORMATION STORAGE" as the equivalent of saying "select all those records from the database where INFORMATION STORAGE is the exact descriptor or identifier."

It did not find records where the *title* or *abstract* or *full text* included the words "information storage" next to each other in this order. The words had to be in the descriptor or in the identifier field (in databases that had an identifier field). Not only that, but the descriptor or identifier had to be exactly in this format, i.e., descriptors/identifiers such as INFORMATION STORAGE TECHNIQUES, or INFORMATION STORAGE & RETRIEVAL did not qualify as a match. In order to retrieve records that include this compound term from the title, abstract, and full-text fields, or as a part of a longer descriptor/identifier, the user had to enter "information (w) storage." This is counterintuitive and illogical even for seasoned searchers, as compound terms such as drug abuse, United States, deficit reduction, New Mexico, or even attention deficit disorder just flow from the fingers of someone who is making a search.

In a library that has databases sporting various search software programs, this is especially confusing and illogical for the patrons. Using a database with the UMI software, the query "information storage" retrieves records where the two terms are next to each other in this order in fields that are used to create the basic index. In SilverPlatter databases, the software retrieves records where the component words are in any order. While this latter may backfire by retrieving both school library and library school, or surgeon general and general surgeon, these lenient approaches are much better than the ultrarestrictive interpretation of DIALOG, especially in indexing or abstracting-and-indexing databases. In databases running with the EBSCO software, it retrieves

records where the terms are not more than 20 words apart, and the proximity can be changed. Database-dependent stopwords in some software further mystify the issue for the casual users.

Now, finally, DIALOG has changed the excessively restrictive interpretation of the space character by the introduction of the FIND command. It can do all the tricks that the SELECT command can

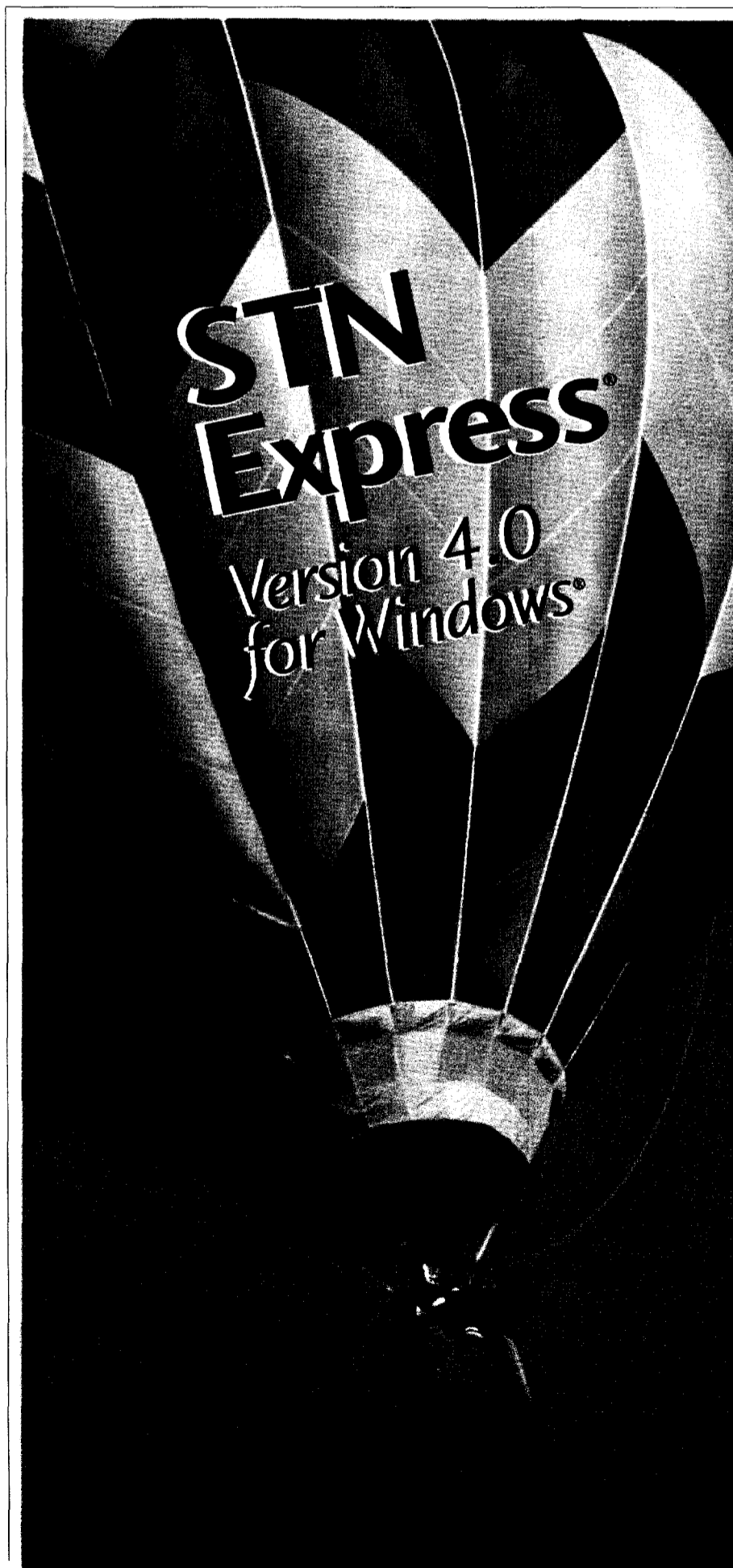
(but not SuperSelect), and it interprets such queries in a common-sense way: words next to each other in this order, i.e., find INFORMATION STORAGE will retrieve records where this pair of words occurs in this order in *any* of the fields that are used to create the basic index, including descriptors and identifiers that have this compound term as part of longer ones. This is intuitive. This is reasonable.

I had been begging for it for 20 years privately and publicly, and I had given up on it as Columbus had given up on finding India after being marooned on the shores of Jamaica. Now this feature arrived without fanfare, to be explored by yours truly on Columbus Day.

#### The Ultralenient Interpretation

Interestingly, the other extreme interpretation—still in existence, by the way—comes from DataStar, Knight-Ridder Information's other online service. Using DataStar, if you enter the term INFORMATION STORAGE, it will retrieve every record that includes either

(continued on page 54)



## Soar to New Heights!

Announcing Version 4.0—

Now is the perfect time for you to rise to the occasion and discover how STN Express can help you do your online searching of scientific technical databases—better, faster, and with more confidence.

#### STN Express

- provides flexible autologon to major vendors
- saves time (and money) with offline query preparation
- includes structure and reaction query drawing with easy-to-use menus and tools
- enhances TIFF image handling
- offers Internet access to online vendors via Windows sockets
- supports fast communication speeds
- permits network sharing

And now—for ONLY \$195 (new order; single-user license); \$49 (single-user upgrade from any previous version)!

CALL TODAY TO PLACE YOUR ORDER

1-800-753-4227 or  
1-614-447-3731  
Internet <http://www.cas.org>



A division of the American Chemical Society®

# Calendar

## 1997

### January

26-28: Fort Lauderdale (Marriott Marina). The **Special Libraries Association's 1997 Winter Education Conference**, titled "Riding the Technology Tidal Wave," includes a variety of half-day workshops, full-day courses, technology-related exhibits, and demonstrations of the latest products and services designed to enhance the success of special libraries and information centers. Information professionals can take the opportunity to learn the latest technology trends from experts in the library and information profession. Contact: SLA's Professional Development Department, 202/234-4700, ext. 649; e-mail: gianna@sla.org.

### February

14-20: Washington, DC (Washington Convention Center). The **ALA Mid-winter Meeting** is expected to attract some 8,000 librarians and leaders in the library and information industry. Winners of the "Academy Awards" of children's literature—the Newberry and Caldecott Medals—and other awards will be announced. Topics to be covered include intellectual freedom and legislative concerns. Exhibitors will showcase the latest in books, videos, computers, and other materials for libraries and their patrons. Contact: ALA Public Information Office, 50 East Huron Street, Chicago, IL 60611, 312/944-6780; 800/545-2433.

19-21: Kissimmee, Florida (Hyatt Orlando Hotel). **Orlando Multimedia '97** is the 15th conference and exposition on interactive instruction delivery and distance learning systems. This event will focus on distance learning and telecommunications, Internet applications, interactive instruction materials development, electronic performance support systems, industry and management training, public and university education, health-care science education and training, legal issues in multimedia programs, and innovative technology for multimedia/distance learning systems. Contact: Society for Applied Learning Technology, 50 Culpeper Street, Warrenton, VA 22186, 540/347-0055; Fax: 540/349-3169.

### March

9-13: Arlington, Virginia (Hyatt Regency Crystal City). **Computers in**

**Libraries '97**, North America's largest technology conference and exhibition for librarians and information managers, is held in cooperation with the Special Libraries Association. It features more than 100 speakers participating in three simultaneous tracks covering all aspects of library technology, an extensive exhibit hall, and pre- and post-conference workshops. Tracks focusing on the Internet and the World Wide Web provide practical examples of how this new technology is being integrated into library platforms. Case studies and creative applications of multimedia, CD-ROMs, information delivery, digital and virtual libraries, and library systems provide many ideas to employ in your own organization. The education and technology track looks at the use of technology in schools and learning environments. Contact: Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055, 609/654-6266; Fax: 609/654-4309; <http://www.info.today.com/cil.html>.

16-18: New Orleans (Royal Sonesta Hotel). The **ASIDIC 1997 Spring Meeting**, to be chaired by Mike Tansey, president of the Institute for Scientific Information, will focus on the theme "Strategic Perspectives on Change in the Information Industry." Speakers representing primary publishers, secondary publishers, distributors, and customers will examine the impact of change on the strategic direction of general reference, business information, and the scientific and technical information industry. Contact: Jeanette Webb, ASIDIC Secretariat, P.O. Box 8105, Athens, GA 30603, 706/542-6820; Fax: 706/542-0349; [jwebb@uga.co.uga.edu](mailto:jwebb@uga.co.uga.edu).

### April

11-14: Nashville, TN (Nashville Convention Center). The **Association of College and Research Libraries**, the largest division of ALA with 11,000 members, is holding its eighth national convention. The conference theme is "Choosing Our Futures." The three-day conference, which will feature more than 200 exhibits and 175 program sessions, is expected to draw 2,500 academic librarians. Registration materials will be mailed to all ACRL members before the end of 1996 and can also be found at <http://www.ala.org/acrl.html>. Contact: ACRL, 50 East Huron Street, Chicago, IL 60611, 800/545-2433, ext. 2519; e-mail: [ddavis@ala.org](mailto:ddavis@ala.org).

## UMI

(continued from page 1)

us with a first-generation tool for providing much information through the Business Analyst feature of the EyeQ and DataTimes online service."

"Rather than investing in the development of two systems," Riner said, "we decided to concentrate our efforts and resources on further developing Business Analyst. The next generation of Business Analyst is already under development, which will allow for greater flexibility in retrieving information and expanding access to a more comprehensive content base."

Introduced in May, Business Analyst bundles an array of investment-oriented data from such financial information providers as Standard & Poor's, Disclosure, Investext, Tradeline, and others. Information in Business Analyst includes company reports, comparisons, price history, corporate news, financial statements, and corporate profiles.

### Benefits to Users of Business Analyst

"Business Analyst," Riner said, "offers several advantages that were unavailable with the reports from Avenue Technologies. The first is that information is

available more immediately online. Previously, end users had to wait several hours to obtain information. Any wait was unacceptable. And when they received the information, it was provided only on paper."

Another advantage of Business Analyst is customization. "With the predefined reports," Riner said, "customers could not select the 'nuggets' of information they wanted. Basically, it was an all-or-nothing proposition. But with Business Analyst, when you ask specific questions, you will receive specific answers to those questions."

A third advantage Riner cited is pricing. "With Business Analyst," he said, "cost control is directly in the hands of the end user. If you know what information you're looking for, you pay only for that information instead of a complete report."

"As a result of the many advantages offered by Business Analyst and with the growing trend in our markets for delivery of information immediately and online," Riner said, "we reluctantly came to the conclusion to cease producing *Executive Reports* and to close Avenue Technologies on November 14."

Source: UMI, Ann Arbor, MI, 800/521-0600, 313/761-4700; <http://www.umi.com>.

## Viewpoint

(continued from page 9)

the word "information" OR the word "storage" in any of the fields used for creating the basic index. Yes, Boolean AND and proximity operators like ADJ are available in DataStar, but the default interpretation is counterintuitive.

I had the same problem with *Compton's Encyclopedia* before, and I am delighted to see that it has been changed in the 1997 edition, at least as an exclusive interpretation.

### A Change for the Better

Prior to the 1997 edition of the *Compton's Encyclopedia*, the space between words was also interpreted as an OR operator. True, the articles that had all the words got higher rankings, but completely irrelevant articles were also retrieved. If there were no articles that had both (or all) the words present, only irrelevant articles were thrown at the user.

In my review of the 1996 edition of *Compton's Encyclopedia* in the April 1996 issue of *Computers in Libraries*, I illustrated how the ill-named SmarTrieve software retrieved dozens of irrelevant articles in response to the search "David Robinson." It did not have an article about David Robinson but listed the articles about Johan Wyss (who wrote *The Swiss Family Robinson*, in case you were wondering what the connection was), Camp David, Star of David, and—most puzzling of all—the national anthem of Swaziland (because its composer's name, which appears in the caption, is David Rycoft).

One of the developers responded to my inquiries, assuring me that in the next edi-

tion there would be Boolean, proximity, and positional operators as well. And there they are, indeed, in addition to the traditional solution. I am happy, but not perfectly happy. I hope that in the 1998 edition, the software will be smart enough to turn the priority around and retrieve first those articles where search terms are adjacent to each other in the given sequence, then—if nothing is found—in any sequence, then with one or two intervening words, then—if still nothing is found—with an AND logic. Finally, only if all else fails, it should switch to the weighted OR operation.

This is not idle fantasizing. It is possible to implement, as the software is now capable of all these operations. The developers just have to put the horse before the cart. I am open to compromise; I don't want to spoil the party for those who rave about natural language. As I put it two years ago in *IT* in response to an article that dismissed the need for Boolean operators and favored an exclusive natural language approach: Don't kiss Boolean goodbye; it's AND NOT OR let alone XOR. Now we've almost made a deal.

*Péter Jacsó is associate professor of library and information studies at the University of Hawaii. He writes for this and other professional magazines, speaks at professional conferences, and regularly offers his online/CD-ROM workshop series. His e-mail address is [jacso@hawaii.edu](mailto:jacso@hawaii.edu).*

### References

1. Tenopir, Carol. "The Black Box Syndrome." *Library Journal* (June 1, 1994).
2. Ojala, Marydee. "Who's Hosting This Search?" *Online* 19(4): 88-92 (July/August 1995).