



11-1-1998

Proving Your Point with Word Searches

Carol Tenopir
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_infosciepubs



Part of the [Library and Information Science Commons](#)

Recommended Citation

Tenopir, Carol, "Proving Your Point with Word Searches" (1998). *School of Information Sciences -- Faculty Publications and Other Works*.
https://trace.tennessee.edu/utk_infosciepubs/407

This Article is brought to you for free and open access by the School of Information Sciences at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

Disclaimer: This is a machine generated PDF of selected content from our databases. This functionality is provided solely for your convenience and is in no way intended to replace original scanned PDF. Neither Cengage Learning nor its licensors make any representations or warranties with respect to the machine generated PDF. The PDF is automatically generated "AS IS" and "AS AVAILABLE" and are not retained in our systems. CENGAGE LEARNING AND ITS LICENSORS SPECIFICALLY DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING WITHOUT LIMITATION, ANY WARRANTIES FOR AVAILABILITY, ACCURACY, TIMELINESS, COMPLETENESS, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Your use of the machine generated PDF is subject to all use restrictions contained in The Cengage Learning Subscription and License Agreement and/or the Gale Academic OneFile Terms and Conditions and by using the machine generated PDF functionality you agree to forgo any and all claims against Cengage Learning or its licensors for your use of the machine generated PDF functionality and any output derived therefrom.

Proving your point with word searches

Author: Carol Tenopir

Date: Nov. 1, 1998

From: Library Journal(Vol. 123, Issue 18)

Publisher: Library Journals, LLC

Document Type: Article

Length: 1,573 words

Abstract:

Online search services may be used to trace word usage which in turn may be used to determine the growth of an industry or to help in budgetary planning. Dialog and Lexis-Nexis are effective online services for use in analyzing word usage.

Full Text:

IF YOUR MORNING RADIO, like mine, is set to National Public Radio (NPR)'s "Morning Edition," you probably notice every time you hear "online services provided by Dialog" or "by Lexis-Nexis," or, in the past, even "by DataTimes." Journalists use online databases for research, to read competitors' stories, and to locate experts. They also use online sources to trace word usage over time, thus creating stories or proving a point about the growth of a concept or a technology.

A popular type of newspaper "filler" item compares how many times a word is used now with how it has been used in the past. For example, to demonstrate the growing popularity of the Internet, the article may say that "Internet" was mentioned only 68 times six years ago in U.S. newspapers, as compared to nearly 25,000 mentions in the first eight months of this year. Although there are limits to the conclusions drawn from such analysis, librarians, too, may find that using online sources can help in budget documents or in justifying new services.

Tracing web words

Although the Internet was started in the 1960s by the U.S. military and researchers, even by the early 1990s popular newspapers wrote little about it. The term "Internet" appeared in fewer than 100 stories in each year from 1989 to 1992 in 16 major newspapers; it wasn't until 1995 that mainstream discussions of the Internet really took off. Not coincidentally this was just two years after the birth of the World Wide Web. In the first eight months of 1998 "Internet" had already appeared in almost 25,000 stories in 16 major newspapers.

The World Wide Web has spawned a growing interest in all applications and facets of the Internet. The web itself generates tremendous numbers of stories, although not nearly as many as the Internet alone and, surprisingly, fewer stories in 1997 than in 1996 and perhaps even fewer in 1998. Through August 1998 the number of mentions was just over 3800. (Note that I searched only the phrase "World Wide Web" or "www" since the word "web" alone is much too ambiguous. Maybe by 1997, people were so familiar with the "web," the papers referred to it by its shorter nickname. Or to many readers the term "Internet" is now synonymous with the "World Wide Web.")

To find out how many newspaper stories provided a web address, I searched "URL or HTTP" (and removed false drops from 1990 to 1992). Obviously, the practice of referring readers to a web site is gaining steam quickly.

Seeking old jargon

To trace the most common terms, search for jargon. The terms "information superhighway," "way," information highway," or "infohighway" reached their peaks in 1994. Their popularity is definitely on the wane and might peg the user as out-of-date. The term "National Information Infrastructure" has also peaked.

I tried out several other words to see what conclusions I could draw. The term "America Online" (or "AOL"), for example, didn't crack triple-digit mentions until 1993 but has been mentioned in more stories than the World Wide Web in nearly every year since then. The terms "online" or "on-line" have been mentioned more and more throughout the decade. I didn't try "CD-ROM" or "libraries" or "information industry," but those and many others might prove interesting.

Dialog and Lexis-Nexis are best

Dialog and Lexis-Nexis are the best online services to use for such analyses because they have the most extensive collections of U.S. newspapers and news magazines. Coverage of many major newspapers on these services (such as the Washington Post, Boston Globe, Chicago Tribune, or USA Today) goes back ten or more years, so a reasonable comparison over time can be made. Dialog's "major papers" OneSearch category allows 16 papers to be searched at the same time, while both Lexis-Nexis and Dialog allow users to specify which newspapers they would like to search together.

Using classic Dialog

Several features of the "classic" Dialog search system make it the best system to research word stories. Dialindex, Dialog's master index for all of its databases, is a fast and inexpensive way to find how many times a word has been used in any of the newspapers on Dialog. The cost of Dialindex is low since it is mostly a finding tool, but it is ideally suited to determining how many times specified words appear in selected databases or groups of databases, such as major U.S. newspapers.

If the command SET DETAIL ON is used, Dialindex will show how many times each word occurs in each of the newspapers. With SET DETAIL OFF, the newspapers are treated as an aggregate. To find how word appearance has changed over time, each publication year in turn is "anded" with the word.

Since there are no records in Dialindex, the searcher can't know if there are any false drops in the numbers displayed. Searches may be saved in Dialindex and rerun in a specified database or group of databases to check for false drops. I was suspicious when "URL" had a few hits in 1990, 1991, and 1992--before the World Wide Web and its concept of Uniform Resource Locators existed. A check of Dialog's KWIC (key words in context) display format for those records showed they were all false drops: URL was the name of a recalled pharmaceutical product in the early 1990s. The "OneSearch" feature allowed me to search all of the major newspapers at the same time; KWIC display allowed me to check for false drops. Of course, it is impractical to check for false drops in most word searches because the number of records is so large.

Try a RANK command

Although I didn't use it, Dialog's RANK command is also helpful. According to Dialog, "the RANK command provides the ability to perform trend or statistical analysis on an existing search set." For example, you could check what subject headings are used the most in articles that include a phrase (e.g., "information superhighway") in the title. After creating the title-word set, the searcher inputs RANK to perform a rank command on the set, specifying that the descriptor field is to be ranked. The rank listing will show which descriptors occur the most times in this set.

Rank works in most databases and is free in all but approximately 20 Dialog databases. A maximum of 50,000 terms can be ranked, selected from most phrase-indexed or numeric additional index fields or in phrase-indexed descriptor or identifier fields.

Research limitations

This general approach to research has some major limitations. First, it shows only how many articles in the databases selected included the specified words or phrases in any year. It does not show how many times in each article the words appeared or how the words were used in the articles. The term "Information Superhighway" might be mentioned only in passing in many articles or in a pejorative sense. These numbers draw no distinction.

Some newspapers are highly selective in the articles they pass on to commercial online services. For example, syndicated columns or wire service stories that appear in print versions are generally omitted from the online versions. Typically, fewer articles are found online than appear in a newspaper's print version.

New newspapers are added regularly by both Dialog and Lexis-Nexis. Perhaps the great increases in word occurrence over time may reflect an increase in the number of sources searched. For example, Dialog's coverage of the Los Angeles Times begins with 1993; coverage of the Denver Post begins in 1994. It only contains the last 90 days of the New York Times. Figures for 1992 thus include two fewer newspapers than those from 1995. Only 1998 figures would include anything from the New York Times. Lexis-Nexis, by contrast, has much deeper coverage of the New York Times.

Why should libraries care?

It's mostly newspaper journalists who try to prove a point with online word counts. Why should librarians care? First, simply to keep up with what our constituency reads in the popular press. If the "National Information Infrastructure" is no longer in the news or "information superhighway" has become a cliché, patrons may not understand or appreciate its use in library materials.

More specifically, librarians can adopt this technique as a way to strengthen budget documents or justify new or improved services. Those who demonstrate growing awareness of concepts such as the "World Wide Web" can supplement the typical charts and graphs that show the growth of web servers or web sites. It shows another reason why a library should increase its web presence or add more web training classes.

More than a parlor trick

Still, such a tactic is inexact at best and should be used only for general indications of growth, not absolutes. While it is not a science, it is certainly more than a parlor trick and can provide useful supplementary oomph when you try to make a point. The next time you see an "amazing" word story in the newspaper, look at it critically and then decide how you might use your online searching skills to

create this type of information.

TABLE 1 Year Mentions of Internet 1989 42 1990 98 1991 44 1992 68 1993 922 1994 4,205 1995 12,312 1996 22,415 1997 27,568
TABLE 2 Year Mentions of Mentions of World Wide Web URL or HTTP 1989 0 0 1990 0 0 1991 0 0 1992 1 0 1993 4 0 1994 277 198
1995 3,853 2,976 1996 7,531 9,664 1997 6,542 12,631 TABLE 3 Mentions of Mentions of National Information Information Year
Superhighway Infastructure 1989 2 1 1990 5 0 1991 16 0 1992 20 3 1993 699 55 1994 3,216 77 1995 2,019 45 1996 1,015 48
1997 526 6

Carol Tenopir is Professor at the School of Library and Information Science, University of Tennessee at Knoxville. Her E-mail address is tenopir@utkux.utk.edu

Copyright: COPYRIGHT 1998 Library Journals, LLC. A wholly owned subsidiary of Media Source, Inc. No redistribution permitted.
<http://www.libraryjournal.com.proxy.lib.utk.edu:90/>

Source Citation (MLA 8th Edition)

Tenopir, Carol. "Proving your point with word searches." *Library Journal*, 1 Nov. 1998, p. 32+. *Gale Academic Onefile*, https://link-gale-com.proxy.lib.utk.edu/apps/doc/A21256080/AONE?u=tel_a_utl&sid=AONE&xid=9ad79c8e. Accessed 20 Dec. 2019.

Gale Document Number: GALE|A21256080