Analysis of Retrieval Performance in Four Cross-Disciplinary Databases: Article1st, Faxon Finder, UnCover, and a Locally Mounted Database

Scott Stebelman

As an increasing number of cross-disciplinary databases become accessible over the Internet, librarians are presented with the dilemma of which to choose to support patron research. Several factors, such as cost, retrospective coverage, and document delivery, are usually considered in making a decision. However, one key factor—citation retrieval performance—is often overlooked because comparative data have been unavailable. A study of four cross-disciplinary databases was undertaken to provide those data. In addition to citation frequency distribution, two other variables were examined: percentage of unique periodicals cited per search and relevancy of citations to stated search topic. An analysis of the data is provided, with its implication for database selection.

he advent of commercial cross-disciplinary databases that can be searched on the Internet has been welcomed and enthusiastically promoted by librarians. These databases are seen as important supplements to traditional printed indexes, to specialized CD-ROM databases, and to other more expensive commercial systems, such as those produced by DIALOG and BRS. In some cases searches are free, while in others the library (or user) either pays an annual subscription fee or a fee for each search statement. If databases are accessed over the Internet, telecommunication charges are negligible.

Several articles have been written about the merits and user reaction of one system versus another, but no article has been published that compares citation retrieval rates for the different systems.[1] This factor, however, is important to many researchers, who often need to locate as much literature as possible germane to their topics.

To make such an assessment, three popular and extensively marketed databases were chosen: Article1st (a database on OCLC FirstSearch), UnCover, and Faxon Finder. A locally mounted consortium database, called GENL, was also included in the assessment. This database is comprised of six Wilson databases: Readers' Guide to Periodical Literature, Business Periodicals Index, Humanities Index, Social Sciences Index, General Science Index, and Index to Legal Periodical Literature.

Scott Stebelman is a Humanities/Social Sciences Librarian at the Gelman Library, George Washington University, Washington, DC 20052; e-mail: scottlib@gwuvm.gwu.edu. The author wishes to thank the following librarians who assisted in the citation relevancy assessment: Daniel Barthell, Shmuel Ben-Gad, Deborah Bezanson, W. Chris Filstrup, Elizabeth Harter, Rebecca Jackson, James Kaser, Patricia Kelley, James Kelly, Caroline Long, and Virginia MacEwen. Of course any error in data analysis is attributable to the author.

METHODOLOGY

Thirty subjects spanning a variety of disciplines were searched. These subjects were chosen because they have been discussed frequently in the media, or because they have been common research topics for students and faculty at the author's institution. The searches were conducted during a five-day period in January 1994. Because GENL and UnCover include references predating 1990, but Article1st and Faxon Finder do not, the search period was restricted to 1990–93. UnCover does not index book, motion picture, or music reviews, so these also were excluded. Newspaper articles and duplicate citations appearing in the same search were also left out. To be consistent, searches were entered in the same manner for all databases; this provided advantages for Article1st, UnCover, and Faxon Finder, which automatically "and" terms in bound phrases.

RESULTS

Table 1 indicates the citation frequencies for the thirty searches. Figure 1 illustrates the differences. The range in citation frequency distribution is considerable, with the best performing database outpacing the worst by a magnitude of 3.3 to 1.

To place the citation counts in perspective, it is necessary to indicate the approximate numbers of periodicals indexed by each database at the time searches were conducted.[2]

Article1st	8,500
UnCover	14,000
Faxon Finder	10,000
GENL	2,200

Hence, even though Article1st retrieved only 57 percent of the number retrieved by UnCover, it indexes only 61 percent of the latter's journal number. In some cases the disparity in retrieval count can be largely explained by the disparity in numbers of journals indexed by each database; however, this correlation breaks down when GENL's figures are examined. The number of periodicals it indexes represents 6 percent of the total, yet it retrieved 45 percent of the total number of citations. Its nearest rival, UnCover, which indexes 40 per-
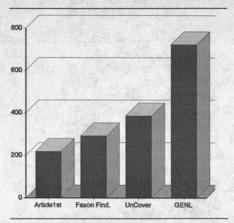

FIGURE 1
Citation Frequencies

cent of the total journals, retrieved only 24 percent of the total citations. The explanation for GENL's superior performance probably lies within its subject indexing, a feature lacking in the other three databases, and the frequency with which abstracts are included with citations.[3] This last feature is also included in Article1st, and to a lesser degree in UnCover, but is totally absent in Faxon Finder. It should be noted that keyword searches in Article1st omit the journal title field; this would reduce its retrieval capacity vis-à-vis the other databases.

A chi-square analysis was made to determine whether the statistical differences among the databases were significant. When GENL is included in the analysis, $p < .01$, $df = 3$. Because GENL is a unique composite database, reflecting the idiosyncratic choices of a local consortium, and because its subject indexing provides it an intrinsic advantage over the other databases, a separate chi-square analysis was made which omitted GENL. A significance level of .05 was established, but the differences among the three databases did not meet this level. The null hypothesis—that the frequency distributions are attributable to chance—cannot be rejected.

UNIQUE PERIODICAL CITATION COUNTS

In addition to citation frequency counts, unique periodical citation counts are

TABLE 1
CITATION COUNT BY SEARCH TOPIC

Subject	Article1st	Faxon	UnCover	GENL
Abortion and sex education	1	0	1	13
AIDS and Asia	6	6	11	19
Arms control and China	3	5	4	18
Art and psychoanalysis	3	6	5	23
Autobiography and women	3	11	1	31
Capital punishment and juveniles	1	1	0	5
Copyright and piracy	9	6	8	56
Epic poetry	4	4	9	27
Fellini	14	18	10	12
France and terrorism	0	1	4	7
Frank Lloyd Wright	29	44	52	40
Free trade, protectionism, and Mexico	0	1	1	19
Gene therapy and ethics	6	6	3	11
Humor and 19th century	0	0	3	2
Hypertext and literature	0	0	3	2
Ishmael Reed	1	6	3	8
Islam and fundamentalism	5	5	3	40
Jackson Pollock	7	12	7	16
Leadership training	48	25	92	19
New historicism	15	35	57	15
Nuclear plants and Russia	2	0	2	0
Ontological argument	8	12	14	8
Ozone layer and Antarctica	1	3	1	13
Poetry and San Francisco	0	2	2	2
Pornography and the First Amendment	0	2	5	21
Poverty and health	40	63	54	181
Rap music and violence	2	1	2	22
Suicide and drugs	7	6	8	56
Transcendentalism	3	11	16	32
Vietnam War fiction	1	1	5	6
Total	219	293	386	724

provided. *Unique periodical citations* is defined as the number of individual periodicals cited in a given search; for example, on the topic of "Copyright and Piracy," Article1st retrieved nine citations, seven of which were to different periodicals. Unique periodical citation count is viewed as an important factor in the assessment, because if a database cited more magazines then its competitors, it would have an advantage (magazines are published more frequently than journals). Second, a popular, and sometimes incorrect inference, is that a database that cites a high number of references on a subject is manifestly superior. This may not be the case if the majority of citations come from a few sources. Conversely, a database that has a lower citation frequency count nonetheless may be a valuable resource to scholars because it retrieves citations from a greater variety of publications.

Table 2 displays the data. Chi-square analysis established the differences to be significant at $p = <.01$. The anomaly previously mentioned is disclosed in these statistics. Although GENL had the highest citation frequency and the highest number of journals cited, it had the

TABLE 2
UNIQUE PERIODICAL CITATIONS

Database	Total Cites	Unique Journals Cited	%
Article1st	219	189	86
Faxon Finder	293	239	82
UnCover	386	302	78
GENL	724	467	65

TABLE 3
CITATION RELEVANCY FIGURES

Database	Total Number of Cities	Total Number Relevant	% Relevant
Article1st	219	136	62
Faxon Finder	293	199	68
UnCover	386	215	56
GENL	724	327	45

lowest percentage of unique periodicals within its searches. That means its subject descriptors are retrieving more citations but to fewer individual titles. Moreover, those databases retrieving the fewest number of citations—Article1st and Faxon Finder—have the highest ratio of unique periodicals to total citations retrieved. Finally, if GENL is omitted in the database comparison, an inverse correlation exists between the number of journals indexed by a database and the percentage of unique periodicals cited. Article1st, indexing 8,500 journals, has the highest percentage, while UnCover, indexing 14,000, has the lowest. What these data suggest is that the numbers of journals covered do not necessarily predict unique journal citation strength.

One might argue that this analysis is beside the point, given that the raw figures indicate that the databases indexing the highest number of journals retrieved the highest number of unique periodical citations. However, defining database superiority is not so simple: if Database A, which indexes 10,000 journals, retrieves 200 unique journal citations, and Database B, which indexes 7,000 journals, retrieves 180 unique journal citations, can one necessarily assume that Database A outperformed Database B?

Libraries may be more impressed with Database A for the sheer number of unique journals cited, but in terms of measuring the inherent tenacity of a database's retrieval performance (expressed as a ratio between the number of unique journals covered and the number of unique periodical citations retrieved), a cogent case could be made for Database B. In spite of this perspective, however, the highest raw numbers will probably be compelling to most users, whose need for journal variety is often a paramount consideration.

Unique periodical citation count is viewed as an important factor in the assessment, because if a database cited more magazines then its competitors, it would have an advantage.

Another statistic judged to be useful was the number of unique journals cited in one database search and not cited by the other databases. This was thought valuable because libraries and users might want to know how rich a particular database might be in covering periodicals not indexed by its competitors. If all four databases had the same searchable fields, such an analysis could be undertaken; unfortunately, the search field discrepancies already noted precluded this.

RELEVANCE

Although a database might be successful in retrieving high numbers of citations on a topic, it was uncertain how many of these were relevant. To make such a determination, twelve subject specialists at George Washington University evaluated the searches most closely congruent with their subject responsibilities. For example, the subject specialist in biology assessed citations retrieved from the "Gene Therapy and Ethics" search, and the subject specialist for art assessed those retrieved from the "Art and Psychoanalysis" search. As previously mentioned, those databases

including subject descriptors and abstracts have an advantage over those databases that do not. To control for these differences, subject specialists were instructed to base their decisions exclusively on the citation and to ignore additional fields. Judgments of relevancy were determined by only one criterion: Was the subject of the citation germane to the search topic?

It must be stressed that because of the large number of search topics and citations retrieved, interscorer reliability was not established for the data. Given the inherently subjective nature of these judgments, the results must be viewed as suggestive rather than conclusive.

Table 3 displays the data analysis of variance, which established the differences to be significant at $p = <.01$. The pattern of data parallels that of table 2. GENL, highest in retrieval frequencies, is also highest in the number of relevant articles retrieved. However, the inferences that can be drawn from these numbers are equivocal: while a large number of relevant articles were retrieved, this number—as a ratio of the total number of citations retrieved—was lowest among the four databases. This suggests that in spite of the subject descriptors, some other field in the database is producing false drops. It might be assumed it is the abstract field: abstracts can generate higher numbers of irrelevant citations, because key words within an abstract may be separated by any number of sentences.[4] However, Article1st has the second-highest relevancy rate; yet it also includes abstracts. More research is needed to explain this negative correlation.

CONCLUSION

The results of this study demonstrate that a database that includes subject descriptors and a large number of abstracts has the ability to retrieve more citations than a database that restricts searches to the basic citation fields. Particularly noteworthy was the fact that GENL outperformed its competitors on this measure, even though it indexed far fewer periodicals, and that the citations it re-

trieved yielded the highest number of unique periodicals. Although GENL retrieved the highest number of citations, the percentage of its citations that were judged relevant to the search topics was lowest among the databases. Ironically, those features that provided GENL with a retrieval advantage—descriptors and abstracts—may also have reduced precision. Statistical analysis of multiperformance measures reveals that none of these databases is clearly superior.

Although GENL retrieved the highest number of citations, the percentage of its citations that were judged relevant to the search topics was lowest among the databases.

In determining which database would be most advantageous for its patrons, libraries will probably consider other factors in addition to retrieval performance. For example, a database that yields a lower number of citations than its competitors might be more user-friendly, and this factor may be weighed more heavily than others in making a final decision. A database might also be offered as part of a package by the vendor that includes auxiliary databases critical to one's clientele. Cost, of course, is another factor: a database that can be searched freely over the Internet, such as UnCover, and that indexes unique periodicals not covered by the others services, will be inherently attractive. Finally, the document delivery features of a service will probably be an important criterion for selection: those services that provide a multitude of suppliers, or that allow orders to be transmitted directly to the interlibrary borrowing unit, will be more competitive than those that do not.

As this study indicates, the identification of a superior database is not always an easy process. Performance has many measures, yet statistics can play a large part in determining which database is an appropriate institutional choice.

REFERENCES AND NOTES

1. See Candace R. Benefiel and Steven Smith. "FirstSearch: A Survey of End-Users," *OCLC Micron* 7 (Dec. 1991): 16–18; Katherine Fuller McKenzie, "FirstSearch in Virginia Libraries," *Virginia Librarian* 39 (Apr./June 1993): 21–23; Susan M. Riehm, "A First Look at FirstSearch," *Online* 16 (May 1992): 42–53; and Karen R Snure, "The FirstSearch Experience at the Ohio State University," *Library Hi-Tech* 9, no. 4 (1991): 25–52.
2. The GENL figure was derived by adding the periodicals listed in the Wilson paper editions. Figures for the other databases were given at the time of database logon or were indicated in the vendor's literature.
3. A useful review of the strengths and weaknesses of free text versus controlled vocabulary searches can be found in C. P. R. Dubois, "Free Text vs. Controlled Vocabulary: A Reassessment," *Online Review* 11, no. 4 (1987): 243–53.
4. The increased recall and false drops that occur when the abstract field is searched has been noted by Carol Tenopir, "Searching by Controlled Vocabulary or Free Text?," *Library Journal* 112 (Nov. 15, 1987): 58.