

figure and the number of legs raised by the horse on which he sits in his memorial statue. On a smaller scale, the Bay Area Reference Center (BARC) in San Francisco issues calls for help with difficult questions, and publishes responses and thanks, in each issue of *BARC Notes*, which circulates to member libraries of the BARC network.

The use of electronic mail and computer conferencing techniques to speed up this process is an obvious improvement, and one that is being applied informally, and with much less fanfare than EIES, in many libraries as they hook into electronic networks. But let's call it what it is—a particular type of information exchange network, not a "tailored and structured "journal"" to use Turoff and Hiltz's terminology.

Anne B. Piternick
School of Librarianship
2075 Wesbrook Mall
University of British Columbia
Vancouver, BC V6T 1W5 Canada

1. Turoff, M.; Hiltz, R. S. "The Electronic Journal: A Progress Report." *Journal of the American Society for Information Science*. 33(4):195-202; 1982.

Author's Reply

Professor Piternick's letter presents an opportunity to clarify a few points that may have not been emphasized strongly enough in this particular article.

The first observation is that if you take a human communication structure and convert it to the electronic medium of communications via computer, the individuals and groups using the structure will not behave in the same manner as with that structure implemented in a different communication medium such as printed journals. The idea of people making inquiries and getting responses from others is not new and unique as an idea and has many representations in face-to-face, telephone services, letters, and printed media.

TOPICS on EIES did not function in isolation from messaging, conferencing, and notebook capabilities and a response might very well generate message exchanges or conference discussions. The increase in the speed of response, the ability of current responders to view responses to that moment, the private message exchanges resulting from responses, follow-up conference discussion, and most importantly, the ability to view how many relative inquiries and responses any member has made, all make the behavior of individuals and groups in this environment very different than what can be generated in any sort of print medium. The important point here is that an electronic automation of an established process does not have the same impact as its original analog. We regret if this point was not made strongly enough in the current article. Even "electronic mail" systems which are sold as automations of internal mail have dramatic long term impacts on organizations that the internal mail never had.

In the normal print media the idea of "inquiry-responses" has had only very limited appeal to very selective groups and is a drop in the bucket when compared to standard journal formats and procedures. It is our point that TOPIC like systems may be far more appealing to wider audiences in the electronic form

than standard copies of the current journal formats and procedures. In other words, what sells in one medium may not be the capability that sells in another.

While EIES has only tested groups in the area of 20 to 100 members for capabilities like TOPICS, a new implementation of the TOPICS design is now being offered commercially on the SOURCE (e.g., PARTICIPATE) with potential user groups in the thousands and organizations like INC Incorporated sponsoring TOPIC like exchanges. EIES has no tie to this commercialization of the TOPICS design but in terms of EIES serving as a vehicle to prototype, evaluate, and evolve designs that work in this new medium, we find the movement to commercial versions of systems developed on EIES a number of years earlier gratifying and supportive of our observations and conclusions.

Murray Turoff
Starr Roxanne Hiltz
Computerized Conferencing and Communications Center
New Jersey Institute of Technology
323 High Street
Newark, NJ 07102

File

Citation Indexes in Information Retrieval

Sir:

The article by Susan Bonzi in JASIS [1] demonstrates that there are still serious misconceptions about the role of citation indexes in information retrieval. Using the undocumented assertions of a single author made over 15 years ago [2], Bonzi builds a fragile hypothesis. On what evidence can one claim that "the dilemma of the user of the citation index is that he knows from experience that only a fraction of the references which cite his starting reference will be relevant to his search requirement?" And continuing this unwarranted premise, "yet he cannot find out which references are relevant and which irrelevant without conducting a time-consuming, hit-or-miss library search." Undoubtedly, there must be instances when her claim is true, but an index of such size and scope, designed for use by a wide spectrum of users, cannot be perfect for every user and remain cost-effective.

My own personal, almost daily, experience in the use of citation indexes for information retrieval is relevant. Rarely have I experienced any of the problems Bonzi describes. When I encounter a highly cited paper, there is usually a review paper in the list of citing papers. That will be my first choice depending upon my need. But in any case, it is the title of the citing paper which provides the discrimination I need in determining which papers will be relevant.

There is a good reason why Cranfield-type evaluations of citation based retrieval do not work. The results of a citation search will vary depending upon the needs of the user. There is no one ideal response, except perhaps—"here is the actual list of papers that cited your starting reference."

Of course, if you start with the most popular work in your field, you can expect more hits than if you start with the most obscure. But do not underestimate the value of the latter type of query. Recently, a scientist at the British Museum [3] was pleasantly surprised to learn how often a particular obscure work had been cited. He had believed that it had been totally ignored by other scholars [4].

While we really do not know what the typical situation is, the number of citing articles one finds in an average search is not so large as to present a problem. But even in the atypical situation, one does have a list of the citing *titles*. Thus, Bonzi omitted the one most important piece of discriminating information in the typical use of citation indexes.

It may be reasonable to assume that the papers cited most often will be looked up more often than less frequently cited papers, but we do not yet have any reliable data. Even if we assume there is some proportionality in use and frequency of citation, the actual citation frequency shows why most users are not inundated with citing papers. Can we speak about the need for subdivision without establishing the ideal user hit threshold?

Frequency data for the 1981 *Science Citation Index* (SCI) is shown in Table 1. Less than 1% of all the items are cited over 17 or more times in one year. In the more than 18 years we have operated our citation-based SDI service called *Automatic Subject Citation Alert*, I have never once seen a complaint from a user stating that a cited reference question was producing too many hits. By contrast, the number of hits on certain title words is often beyond the needs of the research worker. Who but the author can judge whether any current paper which cites his or her work is not relevant? Of course there are times when the title of the citing paper does not make clear why one of my papers is cited. But I have never been disappointed to learn of citations to my work. This is especially true for journals I never see or publish in. In this way, I often learn about authors who are writing on subjects which are not at all seemingly relevant to citation indexing or information science. And that is just the point. The user may find most interesting that citing paper which Cranfield-type logic or evaluation would say is not relevant.

Don't misunderstand this comment. I am not saying we should not use such methods for measuring or estimating relevance. In the design of ISI's on-line systems we do obtain a measure of "relevance" to a research front by counting the number of "core" papers cited by each retrieved item. But this is classification at an entirely different level. Even there, the actual user may find most interesting the paper which cites only one of the core papers for completely unexpected reasons. Julian Smith once called this "systematic serendipity" [5]. But if the purpose of the search is to find the most "relevant" paper on that topic, then the system displays first the review papers that cite the most core papers on that topic.

We should not lose track of the statistical realities of information discovery or retrieval. If the number of papers retrieved is small, then there is generally no need to subdivide. If the number of papers retrievable is large, then subdivision is desirable. But it

is precisely because the literature of a research front has reached a critical mass that we are able to subdivide algorithmically.

The Bonzi assumption implicitly denies the reality of the level of specificity achieved in citation indexing. Can you imagine a subject heading authority list with ten million terms? But the equivalent and more is built into a five-year SCI cumulation. Each of these unique "subjects" is symbolized by an article, book, or whatever. Naturally, many of them concern "similar" subjects but rarely, except in duplicate publication, the identical ones. While Bonzi and I are writing on the "same" subject, each of our papers is different; and in the future I will be interested to learn of any other papers that cite her paper or mine or both. If we provoke enough discussion in the subsequent literature, then maybe a co-citation pair will be established that will define a new subdivision of the growing literature on citation indexing.

Perhaps the most relevant concluding observation on this issue is a statement by M. E. Maron in the September 1982 issue of JASIS [6]—"The function of a document retrieval system cannot be to retrieve all and only the relevant documents (or the citations to those documents). I say this because the system does not know which are the relevant documents—that information simply does not exist . . . But only after the searcher himself has examined the corresponding documents (or possibly the records themselves) can he say which are the relevant ones and which are not relevant."

Eugene Garfield

Institute for Scientific Information
3501 Market Street
Philadelphia, PA 19104

1. Bonzi, S. "Characteristics of a Literature as Predictors of Relatedness Between Cited and Citing Works." *Journal of the American Society for Information Science*. 33(4):208-216; 1982.
2. Lipetz, B. A. "Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators." *American Documentation*. 16:81-82; 1965.
3. Patterson, C. Personal Communication. August 13, 1982; July 20, 1982.
4. Patterson, C. "Cladistics and classification." *New Scientist*. 94(1303):303-306; April 29, 1982.
5. Smith, J. F. "Systematic serendipity." *Chemical Engineering News*. 42(35):55-56; 1964.
6. Maron, M. E. "Associative Search Techniques versus Probabilistic Retrieval Models." *Journal of the American Society for Information Science*. 33(5):308-310; 1982.

TABLE 1. Citation frequency distribution data for 1981 *Science Citation Index*® (SCI®).

Times Cited	Number of Items	% of File
101-Over	469	.1
51-100	1,671	.1
26-50	8,279	.2
17-25	17,110	.4
10-16	58,206	1
5-9	235,456	5
2-4	1,008,420	24
1	2,943,816	69
Total	4,273,427	100%

.8%

Thematic Analysis in Information Science

Sir:

T. D. C. Kuch, in his article "Thematic Analysis in Information Science: The Example of 'Literature Obsolescence'" [1] illustrates his first model (M_1) of the process of *change of state of knowledge* by referring to the Garden of Eden story and the Socratic Method. Both illustrations seem singularly inappropriate. The state of innocence exemplified by the Garden of Eden, that is to be regained by humanity, requires casting aside the long standing tradition of sin and corruption, which began with Eve's revision of a prior instruction. This is essentially a revolutionary process involving a break with past behavior, and