information input. Otherwise, the service becomes a filter which reduces the probability of creative action and, furthermore, will be rejected by the individual.

We must also recognize that our assumption might be naive. The creative individual could very well have refined his selection process to such a degree that the increased literature available actually increases his probability of creative action, even though he never sees a large part of it. In this case, it is unlikely that any interposed system would improve the probability or, for that matter, be acceptable to the individual.

A further consideration is the balance between a centralized information service and a dispersed system, where the service is performed by individuals who are an integral part of the research teams. Here, too, the effects on the probability of creative action of the two approaches must be a most important factor in the evaluation.

The course of this discussion leads on to the conclusion that a qualitative understanding of the role of an information service activity can be developed on a rather general fundamental basis. However, it is clear that we do not yet have nearly enough detailed and quantitative data to enable management to make a reliable evaluation of an information service activity in a specific laboratory, particularly when that laboratory is a heterogeneous one doing exploratory research.

The discussion also suggests rather strongly that an empirical approach involves the danger of introducing unsuspected constraints which would reduce the probabilities of creative action.

The most obvious positive conclusion is that we must undertake further studies to develop a more complete understanding of the relationship between information flow and creativity.

* Presented before the Division of Chemical Literature, ACS National Meeting, Chicago, Ill., Sept. 6, 1961.

(1) J. Hillier, "A Theory of Communications in a Research Laboratory," *Research Management*, Vol. III, Number 4, pp. 255-270, Winter, 1960.
(2) J.G. Miller, *Am. J. Psychiatry*, **116**, 695 (1960).

Keeping Research in Contact with the Literature: Citation Indices and Beyond*

By JOHN W. TUKEY
Bell Telephone Laboratories, Murray Hill, N. J. and Princeton University, Princeton, N. J.
Received October 2, 1961

Varieties of Aim.—Those who write about information systems appear to believe that scientists go to libraries to get "information." This undoubtedly does occur, quite possibly much more frequently for scientists concerned with development. In many cases, however, especially in research, he goes to the library to *interact with the literature*. This is an active process, one in which the scientist's understood aims change steadily, both as a result of what he has found, and as a result of what he has accidentally noticed. The antithesis of the development man in search of specific information is the researcher who has come to browse.

It is not surprising that librarians and information systems proponents have concentrated upon the quest for specific information, for this is the library function which can be most completely transferred to another person, or even to a machine. Browsing, which can be least well transferred, is of key importance to keep research men alert and stimulated. Continual browsing can contribute much to the variety of the fields in which such a man can become expert on short notice. And with the exponential growth of science forcing more specialization on everyone, we should all be concerned with making it as easy as possible to be less specialized. Insofar as literature interaction for research scientists is concerned, we have to emphasize both easy browsing, and easy, rapid contact with the literature of a new sub-area. With these goals in mind, let us turn to some of the other aspects of the problem.

* Presented before the Division of Chemical Literature, ACS National Meeting, Chicago, Ill., Sept. 6, 1961.

Exponential Growth.—Today the most crucial fact about science is its continuing exponential growth (2, 7). Number of articles, number of journals, number of men and women, number of dollars, number of students, all are growing exponentially, doubling every few years, the exact doubling time varying somewhat from field to field and measure to measure. While exponential growth cannot continue indefinitely, and while the eventual slowdown will result in many new problems, it seems likely that research, the part of science most directed toward writing articles, will continue to grow exponentially for the longest time, as will the number of articles published.

Growth has already strained individual abstracting and indexing service to varying degrees. Many steps to help them are being considered, but the extent to which the services can surive is not yet determined. It is clear that we must not only try to help them, but must learn both how to supplement them and, perhaps, how to replace them. We need to keep as much of what we have had as we can afford, but we need to have new things as well. Both sheer volume and delayed indexing have made detailed searches using what we have today noticeably less effective. And the situation is going to get worse, either steadily or catastrophically. The needs are clear: (1) a way to dig into narrow fields; (2) a way to browse effectively; (3) a way to make exponential growth not a handicap, and; (4) a result that does not go out of data.

Non-obsolescence and the Literature Network.—If our means of focused access and easy browsing is to be slow to go out of date, it should depend upon relationship rather than upon classification. Times change, and the

useful way of classifying a field changes surprisingly rapidly, but the relationship of one article to another changes relatively little. The most ancient and familiar relationship between articles is provided by the references which each author makes to earlier articles. We are all accustomed to using these links to trace the development of a subject, to pull together a bibliography of references on a special topic, or to try to find out what is known about some particular question. We all know the main deficiency of using these links when organized into lists of reference: it is only possible to trace our way backward in time.

The Citation Index.—These same links can be reassembled so that topics can be traced forward in time. A *citation index*, when entered with a particular article, reveals a list of articles which have referred to the particular article (*1, 3, 4, 5, 6*). The ability to *go both forward and backward in time* which is provided by *combined use* of citation index and lists of references, and the ability of the user to assess which paths are closer to his goal, or more likely to approach it, combine to make the literature network into a very effective tool for browsing, for covering a limited area in depth, or for keeping up with one's own special interest.

The necessary judgments are already being made, the links are routinely forged. We have only to bring the citations together, re-sort them, and make the resulting citation index available. This has been done for lawyers for many decades by "Shepard," a reference tool the legal profession has long regarded as essential.

Meeting Exponential Growth.—If we are to face exponential growth, the scientific judgment involved must: (i) be made about individual items, (ii) not require a broad knowledge of who is who, and (iii) involve a number of people making judgments which increases exponentially at the same rate as the literature. (Exponential dollar costs can be met for a while, particularly if they begin low.) All the scientific decisions, all the decisions about content, which are required to make a citation index operative are made on a dispersed basis. Each author as he writes his article, each referee or editor as he considers it, is already responsible for exhibiting, through references to earlier work, the proper relationship of each article to the earlier literature. The numbers of authors, referees, and editors probably will grow exponentially at the same rate as the literature itself. Each person will need to make about the same number of decisions as to "what is related to what" as he does now.

Scope, Preparation and Speed.—There are those who think that there should be a single citation index for all of science. There are others, and I believe that I am one, who feel the separate citation indices for individual fields will be more convenient, efficient enough, and easier to start. After all, Shepard does not cover *all* U. S. law in one index. Compared to the importance (and to the inevitability) of the citation index as a tool for literature-interaction, these differences are minor.

Today we are on the verge of citation indexing. Eugene Garfield, with the support of some far-seeing geneticists and the National Science Foundation, is about to start on a three-year-long experiment of citation-indexing genetics. Developing from early trials by Joseph Hodges and others, some of us are cooperatively assembling a citation index for statistical theory and methodology. Soon I expect to hear of others.

I have used an incomplete citation index in my own research and browsing I know I would not want to be without it. (For a sample see the appendix to this paper.) There is a fair amount of routine labor involved in assembling the material, but this is labor that has by-products. During the past summer I have done a fair amount of such routine coding myself, just to get back in touch with some of the literature. I am sure that the gains to graduate students, both in browsing and in learning about the shape of the literature, are so great that the job of covering the accumulated literature in statistics can be well done by them to their own profit. Every field, roughly, has graduate students in proportion to its literature, and can use them to cover its own.

Once the idea of a citation index is adopted in a field, that field will adopt some version of the pattern now current in the law: (1) a new integrated edition each decade, (2) a new supplementary volume each year, (3) a temporary supplement each month.

Once the idea is fully appreciated, the material for the temporary supplements will be made up in the editorial (redacting) offices of the journals, and will be completed at the page proof stage. Thus appearance in the citation index can be nearly coincident with appearance of the article itself.

THE IMMEDIATE TASK. Citation indices are inevitable. Our concern must be to see: (i) that they come soon enough (or almost so), (ii) that their usefulness is understood, and (iii) that further progress is not neglected.

To the first end, I can only encourage each of you to think of some small field where a trial can be made easily, and then to make it. To the second end, I can urge you to make your trials cooperative and to include graduate students, beginning scientists, beginning engineers, and like among your cooperators. To the third end, I can mention a dream or two of my own as samples of what lies beyond the simple citation index.

A SEQUENCE OF DREAMS, ALL ATTAINABLE. My short-range dream for the literature of statistical theory and methodology will surely come to pass. It is, of course, the citation index itself, which needs to be used in connection with a smaller or larger library. One needs to turn backwards in time with the aid of the reference lists given by articles as well forward in time with the aid of the citation index itself, and one needs to refer frequently to the actual articles to keep on or near the trail, whether the aim be quite specific or merely browsing.

My medium-range dream is of what we may call a RECAP, which combines: (1) *RE*ference lists from articles, (2) a *C*itation index. (3) an *A*uthor index giving titles, and (4) a *P*ermuted title index. Here (3) and (4) are of greatest use when first starting while (1) and (2) are used alternately during the main tracing, with assistance from (3) in cutting branches likely to be unproductive.

If a reasonably compact code is used for (1) and (2), and if (4) is compressed carefully, it is possible that a single, very thick volume will accommodate the whole of this material for a field the size of statistical theory and methodology. If this were so, and the volume were at hand, one could go quite far in pulling together a lists of articles for examination in a particular connection

TABLE I

Sample page from a citation-index out of about 30 volumes of AMSX (the Annals of Mathematical Statistics)

Cited	Citing
31PRN 23BMTA 11	34TRR 5AMSX 324
31PRN 23BMTA 23	58SKE 29AMSX 60
31PRN 23BMTA 114	39MCY 10AMSX 337
" " "	52BRY 23AMSX 103
" " "	56SCE 27AMSX 251
" " "	57SKE 28AMSX 188
31LRX 23BMTA 134	40BKR 11AMSX 219
31HJO 23BMTA 315	49HWL 20AMSX 305
" " "	50PLI 21AMSX 100
" " "	55CHU 26AMSX 112
" " "	55CHU 26AMSX 593
31PRN 23BMTA 361	55CHU 25AMSX 593
31PRN 23BMTA 364	43GML 14AMSX 163
" " "	55CHU 26AMSX 593
" " "	58CLK 29AMSX 862
31RTZ 23BMTA 424	36HTG 7AMSX 29
" " "	44CRE 15AMSX 102
32SGU 24BMTA 65	37DWR 8AMSX 21
" " "	38DWR 9AMSX 97
" " "	40PRE 11AMSX 311
32PRN 24BMTA 203	43GML 14AMSX 163
" " "	54GML 25AMSX 76
" " "	58CLK 29AMSX 862
32RTZ 24BMTA 288	32LRG 3AMSX 126
" " "	33FSR 4AMSX 103
" " "	44CRE 15AMSX 102
32PRN 24BMTA 290	33FSR 4AMSX 103
32PRN 24BMTA 292	34KLK 5AMSX 263
" " "	36KLK 7AMSX 51
" " "	41ARN 12AMSX 429
" " "	42SCE 13AMSX 371
32PRN 24BMTA 404	50GRS 21AMSX 27
" " "	52MSN 23AMSX 126
" " "	59HRR 30AMSX9800
32FLR 24BMTA 428	36RTZ 7AMSX 144
" " "	41CRS 12AMSX 409
" " "	42PLN 13AMSX 233
32WST 24BMTA 441	44LHR 15AMSX 388
" " "	57SCN 28AMSX 902
32WLS 24BMTA 471	34KLK 5AMSX 263
" " "	35KLK 6AMSX 202
" " "	36FRG 7AMSX 113
" " "	38HSU 9AMSX 231
" " "	39GRK 10AMSX 203
" " "	39LNL 10AMSX 365
" " "	40DLY 11AMSX 1
" " "	40MDW 11AMSX 125
" " "	40MCY 11AMSX 204
" " "	41WLD 12AMSX 137
" " "	41HSU 12AMSX 279
" " "	42CMP 13AMSX 62
" " "	42CRG 13AMSX 74
" " "	43HTG 14AMSX 1
" " "	43CCN 14AMSX 205
" " "	46TKY 17AMSX 318
" " "	46ANN 17AMSX 409
" " "	47BRN 18AMSX 514
" " "	48VTW 19AMSX 447
" " "	51MOD 22AMSX 266
" " "	55PLI 26AMSX 117
" " "	56KLK 27AMSX 122
" " "	57WJN 28AMSX 414
" " "	58ROY 29AMSX 491
" " "	58ROY 29AMSX1177
33WST 25BMTA 52	37DWR 8AMSX 21
" " "	38DWR 9AMSX 86
33WCL 25BMTA 121	34TRR 5AMSX 324
33PRN 25BMTA 158	54GRN 25AMSX 671

without using any other source. It would still be vitally important to be able to get at a fairly large number of the articles in question, but *back-and-forth access* between literature aid and articles would not be anywhere nearly as essential as it is for a citation index.

Notice that *all* the portions of a RECAP can be produced routinely. Once the title, author(s), coded location and coded references are assembled for each article, the remainder of the task can even be fully mechanized if desired. Except for dollars, further exponential growth offers no preparation problem. RECAPs thus seem entirely feasible.

Notice, however, that there are implications of a scale. A RECAP for all of science is conceivable, but would

have to be in an impractically large number of volumes. If we are to have a RECAP that an individual user will find effective, we must come down to a field of modest size, one characterized by a world-wide membership (in all closely related technical groups) of perhaps ten to thirty thousand persons—of somewhat less than the number that tempts a publisher to bring out an annual "Advances in . . . ".

But the RECAP need not be the end. There is a place for something that goes much further, something probably not quite for the individual worker, but something which could be wisely provided for each small group of workers (say, for *each* of the half-dozen or so groups of statisticians in the Bell Telephone Laboratories).

Here my long-range dream is for an INFORMATION LEDGER, consisting of one page per article, one volume per year (or maybe two; again there is a natural limit on scope of coverage). The one page would contain, on its two sides, the following: (1) title, authors, and original summary or abstract of the article, (2) its list of references, (3) copies of abstracts and reviews of the article from various sources, and (4) a citation index *to* the article. Of this, all but (4) could be photographically reproduced. Author and permuted-title indices for each volume would need to be provided.

Given a set of volumes of an INFORMATION LEDGER, it would be possible to browse quite extensively without turning to anything else, or, alternatively, to decide which one, three, or six articles needed to be looked at to find the specific information sought. With these volumes widely available, so that only a few articles needed examination in each instance, reference to journals by non-returnable photocopy could perhaps become quite effective, so the INFORMATION LEDGERs could be a key step in the transition to a whole new pattern of literature use.

Real problems in the way of INFORMATION LEDGERs include questions of copyright arrangements for reproduction, and of effective means of updating citation indices. These are not trivial, but they do not seem insoluble.

CLOSE. In closing, a word should be said about quite another topic: mechanized, or perhaps, in the eyes of some, machine-made-possible information retrieval. Nowhere above have we claimed that machines can now do things no human could before. Why? Because our concern here has been with the tools that can be *developed* and put to use in the near future. *Research*, as always, holds brave promise for the more distant future, but only if we distinguish it clearly from development. And even then we are likely to find the machine doing faster and cheaper what we have, in the meantime, learned to do as persons. More importantly still, if mechanized information retrieval is going to meet the aims we have concentrated upon here, aims of interaction between seeker and literature, it will have to bring specific persons, namely, those who are seeking, far more deeply and more intimately into the mechanized process than many have so far supposed.

APPENDIX: A SAMPLE FORMAT

Table I illustrates what a portion of a restricted citation index might look like. The code used here has been developed for citation-index use with four principles in mind: (1) reasonable compactness is very desirable, (2) coding should be by rule, rather than of necessity by code book, to the greatest extent possible, (3) coded material should have as large chance of identification by a reasonable casual user as is compatible with (1) and (2), (4) it is unimportant a small fraction of all references are somewhat equivocal. It uses 16 characters, which is bearably sortable with classical punched-card (EAM) equipment. These characters are divided as follows (we take 32WLS 24BMTA 471 as an example):

2 figures for the year (special arrangements involving 1 letter, 1 figure before 1900) = 1932.
3 letters for (first) author's last name (see below): *Wilks*, S.S.
3 figures for volume number: _24
4 letters for journal identification: *Biometrika*
4 figures for initial page number: _471

The author code is basically FCL, namely, First letter, next Consonant, Last letter, with (i) double letters treated as single, (ii) the second of three or more vowels regarded as a consonant, (iii) y as vowel, (iv) the last capital letter in split or hyphenated names used as C. (A small trial indicated that this code transmits only about one information bit less than an author's full name, at least among statisticians.)

The journal code is made according to a set of rules which, as a compromise between simplicity and resolution, are too long to give here. There are many instances where 2 or more journals will receive the same abbreviation. (Thus *J. Clinical Psychology*, *J. Counselling Psychology*, and *J. Consulting Psychology* all come out JCPY.) In most of such cases the relationship between year and volume number will suffice to identify the correct journal. (And if there are a few cases per thousand where one has to look in two or three journals, this raises the look-up effort by less than 1 per cent.

REFERENCES

(1) W.C. Adair, "Citation Indices for Science," *Amer. Document.*, 6, 31 (1955).
(2) Dale B. Baker, "Growth of Chemical Literature—Past, Present and Future," *Chemical and Engineering News*, July 17, 1961.
(3) E. Garfield, "Citation Index for Science," *Science*, 122, 108-111 (1955).
(4) E. Garfield, "Citation Indexes—New Paths to Scientific Knowledge," *Chem. Bull.*, *43*, 4, 11-12 (1956).
(5) E. Garfield, "Breaking the Subject Index Barrier—A Citation Index for Chemical Patents," *J. Pat. Off. Soc. 39*, 583-595 (1957).
(6) E. Garfield, °°A Unified Index to Science," *Proc. Int. Conf. Sci. Information* (Washington, 16-21 Nov. 1958) 461-469. (contains the suggestion of including abstracts as citations), 1959.
(7) Derek J. De Solla Price, Chapter 5 of "*Science Since Babylon*," Yale University Press, 1961. Condensations of Chapter 5 have appeared as: The Acceleration of Science—Crises in our Technological Civilization, *Product Engineering*, March 6, 1961; The Beginning and End of the Scientific Revolutions, 1670-1970, *Lehigh Alumni Bulletin*, March, 1961.