

# Selective Dissemination of Information and the Academic Science Library

*The meaningful discharge of academic duties requires the effective use of up-to-date factual and conceptual knowledge. The wealth of the material available and its perishable character necessitate an expeditious procedure for information sampling, storing, and retrieving at the individual level. The feasibility of such a procedure, administered by the science library, is explored. The discussion consists of three parts: The inherent inadequacies of biomedical bibliographic services to fill the need of an individual for current awareness; the magnitude and cost of a modest personal information system; and the potential supportive participation of the science library in the teaching and research activities of members of the university community.*

REGULAR AND SIGNIFICANT CONTRIBUTIONS to the pool of universal knowledge are expected from the members of the academic community, but this privilege seems something of a liability because the accelerated pace of scientific research has generated such a wealth of publications that it is more and more difficult to keep abreast of developments even in a narrow field. It can be difficult sometimes to judge whether a scientific effort represents original work, and reliance is often placed upon the mere probability that research has been anticipated. In 1963 the number of journals with scientific value was estimated to be approximately thirty-five thousand.<sup>1</sup> A substantial portion of the in-

formation published in these journals has a limited useful lifetime.<sup>2</sup> Therefore, if a publication is to serve its purpose, it must find its audience quickly. The multitude of journals and the perishability of their contents make the need for an effective current awareness service most urgent. The purpose of this paper is to indicate mechanisms by which the science library could render a valuable service to teaching and research by establishing a current awareness service geared to the interests of individual members of the academic community. A brief discussion of a personal information system illustrates the direction of further developments.

No bibliographical service currently functions satisfactorily as a means of keeping individuals up to date, although several comprehensive services are available as retrospective searching devices. In biomedicine, for example, three major indexing and abstracting journals are available in English alone. Each of these journals, *Biological Abstracts*, *Excerpta*

<sup>1</sup> Charles M. Gottschalk and Winifred F. Desmond, "World-Wide Census of Scientific and Technical Serials," *American Documentation*, XIV (July 1963), 188-94.

*Dr. Kolder is Associate Professor of Physiology and Mr. Simpkins is Science Reference Librarian in Emory University, Atlanta. This work was supported in part by the U. S. Public Health Service, National Institute for Neurological Diseases and Blindness Grant NB 21,768.*

<sup>2</sup> R. E. Burton and R. W. Kebler, "The 'Half-Life' of Some Scientific and Technical Literatures," *American Documentation*, XI (January 1960), 18-22.

*Medica*, and *Index Medicus*, now lists more than one hundred thousand publications annually.<sup>3</sup> None of these journals, however, is prompt enough in reporting the literature to constitute an effective current awareness service. The listing in *Index Medicus* runs at least three months behind publication of the original report, the citations in *Biological Abstracts* are six months to one year late, and the abstracts in *Excerpta Medica* sometimes trail the original publication by two years. *Current Contents* might be considered an acceptable current awareness service but it lacks a subject approach to the content pages. Another disquieting factor in the evaluation of bibliographic services is the unpredictability of editorial acceptance of publications for inclusion. In 1965, *Biological Abstracts* listed 6,735 serial titles, and contained 110,119 abstracts or titles.<sup>4</sup> Assuming an even distribution, this would mean that fewer than seventeen titles per journal were editorially selected for inclusion. *Excerpta Medica* listed about forty-four publications per year for each journal covered in 1964, while the comparable figure for *Index Medicus* was sixty-three publications per journal. The inclusion ratios quoted constitute only a fraction of all papers published in the source journals used. This selectivity may be a desirable device for protection of the user from an inordinate amount of noise; but Oehlerts<sup>5</sup> reported that 1,635 of the nearly five thousand journals on the "List of Serials Abstracted" in *Biological Abstracts* in 1960 were not represented by a single citation. Among the omitted titles were the *Comptes*

*Rendus* of both the Academie des Sciences and the Société de Biologie. Nor did Oehlerts find any citations to *Genetica*, *Bibliographia Genetica*, *Archiv für Protistenkunde*, and *Planta*. Other research journals of like stature were represented by fewer than five citations. The difficulty of retrieval from bibliographic journals further diminishes the comfort derived from their services. Neither the permuted title index nor the conventional analytical index permit indexing in enough depth or specificity to develop confidence in finding the material that the editors have seen fit for inclusion. Martyn and Slater<sup>6</sup> found in tests on retrievability that, at best, 85 per cent of publications found through the author index of *Biological Abstracts* were retrievable through the permuted title index. The *Index Medicus* analytical subject approach yielded, in a less comprehensive series of tests, a low of 5 per cent and a high of 70 per cent of the references known to have been included in that bibliographic journal.<sup>7</sup> All this is not to say that bibliographic journals are without merit, but one should be reminded that they have grave limitations.

In order to compensate for the deficiencies of bibliographic journals, most academic teachers and scientists have to maintain a more or less elaborate personal information storage and retrieval system. The input may consist of books, reprints, personal communications, proceedings, minutes, graphs, etc. The information is retained in subsystems of varying form, like shoeboxes, folders, notched cards, or 3 x 5 in. cards and may be found in alphabetic order or under broad subject headings without sufficient subdivision. The multiplicity of subsystems reflects the variety and magnitude of the input. One of the authors has maintained a personal information

<sup>3</sup> "List of Serials Abstracted," *Biological Abstracts*, XLVI (November 1965), 7373, and XLVI (December 15, 1965), 8927; "A Review of Current Activities and Future Plans," *Excerpta Medica*, Sec. I, XIX (January 1965), iii-xii; U.S. National Library of Medicine, *Annual Report Fiscal Year 1964* (Washington: The Library, 1965), p. 23. See also, *Index Medicus*, VI (December 1965), i.

<sup>4</sup> *Biological Abstracts*, loc. cit.

<sup>5</sup> Donald E. Oehlerts, *The Most-Cited Serials in Biological Abstracts in 1960* ("Occasional Paper no. 65," Urbana: University of Illinois Library School, 1962).

<sup>6</sup> John Martyn and Margaret Slater, "Tests of Abstract Journals," *Journal of Documentation*, XX (December 1964), 212-35.

<sup>7</sup> *Ibid.*

system for the past fourteen years. Since 1959 this has been done systematically. The main input to the system originates from the regular examination of some eighty biomedical journals related to human physiology. The total file presently consists of 7,400 precoded 5 x 7 in. cards arranged in serial order. Each card represents a document, which may be a book, reprint, bulletin, specification sheet, manual, or a citation. Administrative material, correspondence, and audiovisual aids are not included in the system. The documents are kept in serial order, but separate from the filing cards. Retrieval of information is accomplished by an author index consisting of about seven thousand entries and a coordinate index with about four thousand keywords. The system now demands approximately three hours weekly of the author's time for scanning and indexing of the source journals. The posting of new serial numbers on author or keyword cards and similar tasks are performed by a secretary. Out of approximately ten thousand publications examined in 1965, eleven hundred were added to the file. On the average, 124 articles were found to be published per source journal scanned. This represents a signal-to-noise ratio of 1:9; but it must be noted that those items retained are highly relevant. The average estimated cost per document retained in the whole manual system is \$1.25, based upon a two-year sample of time and expense accounting. This figure includes the acquisition of some 350 books, about four thousand reprints, filing facilities, and prorated secretarial help. The figure does not include compensation for time spent by the author.

Presently a computer program is being written by J. C. Ziegler (Emory biomedical data processing and analysis center) for transfer of bibliographic information from this personalized system to tape for machine retrieval on an IBM 1410 computer. The computer-based sys-

tem is expected to reduce the retrieval time considerably, especially when the coordination of several keywords is required. Five simultaneous searches for coordination of up to seven keywords each through the whole tape memory require about fifteen minutes. Updating with new information and elimination of obsolete information are standard procedures, as is the preparation of an alphabetic list of authors and keywords. On the other hand, the computer involvement diminishes the accessibility of the file since the computer is not always instantly available. A published figure for a general, non-individualized, computer-operated bibliographic information storage and retrieval system is \$5 for each entry into the system, subject to variation with the number of entries and the number of searches.<sup>8</sup> The information system outlined above has been a valuable aid to both teaching and research. The time and money invested are a small price for the current awareness aspect alone. In addition, there is no problem in compiling references for research projects or reading lists for single students or whole classes.

Certain objections can be raised against maintaining an extensive personal file. It may appear redundant to index material that will be included in one or more of the major bibliographic services, and it seems doubtful whether a single indexer can hope to cover more than a fragment of the great mass of relevant literature. The charge of redundancy has already been disposed of in the discussion of the deficiencies of three major indexing and abstracting journals. These services are too slow, too general in their indexing, and too unpredictable in their selection to render superfluous some supplementary personalized device. The justification for indexing a small fraction of the biomedical journals derives from ev-

<sup>8</sup> Charles P. Bourne and Donald F. Ford, "Cost Analysis and Simulation Procedures for the Evaluation of Large Information Systems," *American Documentation*, XV (April 1964), 142-49.

idence suggesting that the great bulk of significant publications is concentrated in comparatively few primary journals. Brown<sup>9</sup> counted citations in seven leading journals of physiology and found that fifty-eight journals yielded 90 per cent of the citations; publications in chemistry clustered around even fewer journals. The citation count in six representative journals in chemistry revealed that thirty-eight journals produced 90 per cent of the citations. Circulation statistics reinforce the implications of the citation count. Fleming and Kilgour<sup>10</sup> found that sixty-seven journals in the Yale medical library represented 50 per cent of serials borrowed and 262 titles accounted for more than 80 per cent of loans. An earlier study by Urquhart<sup>11</sup> at the Science Library in London established that forty journals provided half the circulation. It is conceivable that the spread of primary information is not so narrow as the citation count and the circulation statistics would indicate. It does seem likely, however, that inspection of a relatively small number of well-chosen journals will retrieve a large fraction of the most important scientific publications pertinent to personal and specific interests.

Since current awareness is vital to the academic teacher and scientist as an individual and since both theoretical and empirical reasons indicate that a small scale, *i.e.*, university-wide, service can be effective, the science library could enhance its value by providing a current awareness service. This service might take, as a first step, the form of a daily list of accessions, mailed to each par-

ticipating member of the academic community. It might provide, as a second step, copies of the content pages of journals requested. If electronic data processing equipment is available, the library as a third step, might distribute weekly a permuted title index to the content of scientific journals received. A logical final step might be the establishment of a system known as Selective Dissemination of Information. This system requires that participants profile their research and teaching interests in terms of weighted keywords. The keywords can be authors' names, subject headings, names of organisms, reactions, or even citations. Each participant may submit as many profiles as are necessary to describe all aspects of his interests and needs. The publications in selected journals are indexed, using terms in the participants' profiles as a subject authority. The profiles of the participants are matched, by computer, against the profiles of the indexed publications. If a preselected level of correlation is obtained between the profile of the participant and the profile of a publication, an abstract is sent to the participant, who, in turn, is expected to furnish feedback on the actual relevance of the publication. The feedback results in a continuous revision of the user's profile and increases the ratio of relevant information to noise delivered by the system. This information system was devised by the late Hans Peter Luhn of IBM, who also wrote a program for it, which is available from IBM upon application. The system has proved itself in many non-academic environments, among them IBM itself, NASA, and the Ames Laboratories. One major difficulty arises with the maintenance of a Selective Dissemination of Information system, *viz.*, finding and keeping capable indexers. Graduate students can be used effectively, but student help usually implies a fast turnover. The most practical way,

<sup>9</sup> Charles Harvey Brown, *Scientific Serials* ("ACRL Monograph No. 16," (Chicago: Association of College and Reference Libraries, 1956), p. 116-19; 98-99.

<sup>10</sup> Thomas P. Fleming and Frederick G. Kilgour, "Moderately and Heavily Used Biomedical Journals," *Medical Library Association Bulletin*, LII (January 1964), 234-41.

<sup>11</sup> D. J. Urquhart, "Use of Scientific Periodicals," in *International Conference on Scientific Information, 1958. Preprints of Papers* (Washington: National Academy of Science, 1958), p. 277-90.

and possibly the final step in improvement toward an individualized system at the university level, seems to be the direct participation of potential users. Each participant could make himself available for the indexing of five or six journals in his field of specialization. Judging from the Cranfield reports,<sup>12</sup> this effort should require not more than five to six hours per month, and would seem a small investment for a hand-tailored service prepared competently and covering an area limited only by the number of participants and their diverse interests. With little more trouble and expense the library could maintain this

system as a retrospective device for literature searches.

The progressive build-up through the steps discussed could be established and maintained by the academic science library. The service should improve markedly the signal-to-noise ratio at the input to a personalized information storage and retrieval system. Admittedly, this would be a departure from the traditional role of an academic science library, but special libraries have been attempting services like this for several years. The availability of computers, the growth of knowledge, and the accumulating emphasis upon research have combined to make it all but mandatory that the science library assume a more active role in the research process. ■ ■

<sup>12</sup> Cyril W. Cleverdon, *Report on the First Stage of an Investigation into the Comparative Efficiency of Indexing Systems* (Cranfield, England: College of Aeronautics, 1960).

