

Review Articles

Storing and Retrieving Information

Emerging Solutions for Mechanizing the Storage and Retrieval of Information. Compiled by Mortimer Taube. (Studies in Coordinate Indexing, vol. 5). Washington, Documentation, Inc., 1959. 178 p.

The first volume of *Studies in Coordinate Indexing* appeared in 1953, the second in 1954, and the third in 1956. Volume four of the series, under the title *The Mechanization of Data Retrieval*, appeared in 1957. A separate work, not a part of this series, appeared in 1958: *Information Storage and Retrieval; Theory, Systems, and Devices*, edited by Mortimer Taube and Harold Wooster, and published by the Columbia University Press.

The first half of the present volume contains five papers by Dr. Taube. The first paper, "Problems of mechanizing storage and retrieval of information," is the broadest of the lot, and in it Taube presents his basic views. A good deal of the material was also used in the earlier *Information Storage and Retrieval*. There is a considerable amount of this sort of repetition from volume to volume, as might be expected in a series which has didactic purposes as well as the purpose of communicating results of study and research.

The second paper is a fascinating study on "The distinction between the logic of computers and the logic of storage and retrieval devices." This topic also was covered, to a lesser degree, in *Information Storage and Retrieval* (pp. 90-96), and was prefigured in some of the earlier series volumes. This is an alluring piece of work, which is nevertheless not entirely convincing, because incomplete. If the circuitry of computers is excessive to the needs of information storage and retrieval *per se*, how much of this excess is nonetheless useful and even necessary for the convenient reordering and manipulation of items along subordinate axes such as journal title or language or permuted article title, as in the new *Chemical Titles* being issued by the American Chemical Society? Well, Taube

would say, it all depends on what you want. And of course it does. It also depends on what computer you are talking about; the excess capacity in UNIVAC will be less than in LARC. And perhaps there is something to be said along the lines that some excess capacity can be economically justified in a mass-produced standard machine. Taube doesn't discuss any of this; his paper covers the area precisely stated in its title. But it leaves hanging these implications for practice. If we can first agree on what degree of "specialism" we mean, we can all agree to his conclusion that "Machines are designed for special purposes; the design and logic of any individual machine should reflect such purposes. If a complete abstraction is made from purpose and efficiency, there remains no basis for design; that is, no basis for the logical and physical arrangement of parts and functions which constitute a machine. Hence the concept of a universal machine is in essence contradictory." He goes on to say that "Computers and storage and retrieval devices are different types of information handling machines. Having different purposes, they differ in design, operating characteristics, and logic. The distinction which has been drawn between a two-valued propositional calculus and an algebra of classes illustrates the fundamental character of these differences." Very well. The suspicion remains that there is a difference between the logic of the relationship of parts of an information retrieval problem, on the one hand, and the logic of the circuitry of the computer, on the other hand, but that these are by no means incompatible with each other, and that in fact their relationship is symbiotic.

In other words, Taube implies that the logic of internal design of hardware circuitry dictates the logic of the applications of the system. He refers to the propositional calculus as the logical method for describing computer networks; I see no valid reason for assuming that, as a consequence, the logic of computer applications also must be described in terms of the propositional calculus only. In fact, this is manifestly not the case.

In the third paper, with L. B. Heilprin, Taube discusses "The relation of the size of

the question to the work accomplished by a storage and retrieval system." Even if one cannot understand the mathematics involved, as I cannot, the piece has in large measure the virtues of stimulation and suggestibility.

The fourth paper, "An evaluation of use studies of scientific information," is a marvelous *tour de force*. Here Taube abstracts the papers on this subject presented at the International Conference on Scientific Information held in Washington in 1958. Since the Conference papers reviewed all previous papers, says Taube, his review of the Conference papers is in essence a review of reviews, and since the Conference papers indicate that use studies have generally provided ineffective results, therefore the general conclusion must also be to this effect. Taube goes on to try to show that the results are necessarily ineffective, ambiguous, and disappointing. He makes some good points, as for example the simple distinction between primary and secondary publications, and the supremacy of professional judgment in designing secondary publications and services. And yet, once more, there is much which could and should be added. To take a simple case, is there nothing to be said for a use study of an abstracting service, for instance, which would inquire into such matters as frequency of issue, rough subject breakdown to serve interests of maintaining current awareness as opposed to retrospective search, and matters such as provision of foreign-language materials in translation? But the effect of the essay is salutary; it is sad but probably true that we need to be reminded again and again of the vast differences among concepts such as "communications" and "dissemination of information" and "information storage and retrieval."

In the fifth paper Taube describes "The COMAC: an efficient punched card collating system for the storage and retrieval of information." It is an ingenious idea, and well presented.

The sixth paper, by R. W. Murphy of the IBM Corporation, describes "The IBM 9900 Special Index Analyzer," which is IBM's commercial version of the COMAC. A reading of this paper is almost sufficient, in itself, to instruct the librarian in all he needs to know about symbolic logic. What he needs to know is only the first page of the first chapter of the large encyclopedia on the

subject, but he does need to know this much, which happens to be of great simplicity. It is a pity that most librarians are stuck with the general view expressed by John Metcalfe in his book *Information Indexing and Subject Cataloging* (Scarecrow Press, 1957) that "classical logic [is] . . . a more useful discipline for indexing theory and practice of any kind than symbolic logic and its algebra or calculus of classes." (p. 27). It would be difficult to be more dead wrong.

In the seventh essay H. P. Luhn of the IBM Corporation describes "The IBM Universal Card Scanner for punched card information searching systems." Luhn is the acknowledged master in the field of punched card devices for information storage and retrieval; here he demonstrates once more his inventiveness and consummate skill in designing ingenious coding patterns.

The eighth paper is by J. C. Costello, Jr., and Eugene Wall of the DuPont Company. They discuss "Recent improvements in techniques for storing and retrieving information." (Some of this material has reappeared in the recently issued report of the Committee on Government Operations, *Documentation, Indexing, and Retrieval of Scientific Information*, Senate Document No. 113, 1960.) It is a fine article. Among other things, it discusses the important concept of "role indicators," which, from one way of looking at it, may be thought of as the coordinate indexing machine system answer to subheadings, which of themselves are irrelevant within the context of such systems. Here, too, is a discussion of the "thesaurus" of technical terms. This word thesaurus has become fashionable recently. The big Webster says that "thesaurus" means "treasury or storehouse; a repository, especially of words, as a dictionary." The introduction of the word into the documentation field originally served a useful purpose, but now, like so many jargon words, it has lost very much of its sharpness. Just look at the use of the word in the Luhn essay (and Luhn was one of the first to use it in a special sense, along with Bernier and some of the British mechanical translation people), and then look at the way Costello and Wall use it here and especially in the Senate Report; the word is now being used in two quite different senses, and one more pitfall has been added for the unwary.

The ninth and last essay is on "The Mag-

nacard system," and is by R. M. Hayes, formerly of the Magnavox Company. The nine pages of text simply are not enough to encompass an adequate explanation of the total system. It is good, however, to have between hard covers such data as are here available.

It is curious to note the relative lack of attention which Taube's work has received in the Library press. Without making an exhaustive survey, I am aware of only five reviews of substance—my own review of Volume 1 in *Bulletin of the Medical Library Association* 42: 380-4, July 1954; Saul Herner's review of Volume 3 in *American Documentation* 8: 56-8, January 1957; and three reviews of *Information Storage and Retrieval*: by Henry Dubester in *College and Research Libraries* 20: 254-5, May 1959; by I. A. Warheit in *Library Quarterly* 29: 223-5, July 1959; and by B. C. Vickery in *American Documentation* 10: 319-20, October 1959. It is curious for many reasons; Taube's writing is of brilliant clarity, marred only occasionally (and not at all in his two most recent and best volumes) by excessive polemical zeal; always full of apt metaphor; rarely padded with a single extraneous phrase; loaded with seminal ideas of great power. It is possible that the early claims for the Uniterm system, pushed too far, wearied and alienated a large part of the profession. If this be true, it has its ironic aspects; to my mind, what was and is essentially wrong with the Uniterm system is not its unusual posting system, and certainly not its central coordinate indexing concept, but its employment of article-derived catchwords in preference to a carefully chosen controlled vocabulary of terms. (Taube, Wall, and many others would presumably still dissent vigorously from this view.) Taube himself was careful and precise in distinguishing his Uniterms as a single variety of coordinate indexing, but still and notwithstanding, the profession, ox-like, insists on equating the two.

The immediacy and relevance of some of the essays in the early volumes has faded. Some of the projected plans have not materialized. But many of the essays, and the ideas in them, were and still remain fundamental advances. Of Volume 1 of the *Studies*, the chapter on The functional approach to bibliographic organization (reprinted from Shera and Egan, *Bibliographic Organization*,

Chicago, University of Chicago Press, 1951), the chapter by C. D. Gull on "Substitutes for the card catalog," and the chapter on "Evaluation of information systems for report utilization," are still of great interest and usefulness. Of Volume 2, the first chapter on "Machines and classification in the organization of information" is outstanding; one could only wish that the cryptic remark on the impossibility of "categorization" in large systems had been more fully explained, and followed up in subsequent work. Volume 3 contains a chapter by Wildhack, Stern, and Smith of the National Bureau of Standards on the peek-a-boo system used by them. Volume 4 contains an interesting paper on meaning and linguistic structures, a critique of the Minicard system which suffers from being somewhat precious in manner, and a table of "dropping fractions" which specify the number of false drops which may be expected in superimposed coding under various conditions. *Information Storage and Retrieval* contains a good beginning toward a glossary of SR terms, an excellent series of working papers by Taube and his associates, and first-rate papers and discussion from Ralph Shaw, Calvin Mooers, Eugene Wall, Charles Bernier, H. P. Luhn, and others. If the discussions are inconclusive, they are yet pregnant; it should be evident, for all who care to see, that a new era is struggling to be born. In these essays and in those of other contemporaries, greater and more lasting programs has been made in the last decade in theory of subject bibliography than in many decades preceding. Contrary to the view prevailing in some quarters, Cutter, Kaiser, and Hulme have found worthy successors, whose contributions will surpass them.

In a paper which appeared in *American Documentation* in 1953 (Vol. 4, p. 163-73) Jesse Shera said that any system of bibliographic organization "must be designed to make readily available the extensive and increasingly intricate accumulations of technological and operational literature as well as the literature of scholarship," and that "failure on the part of the librarian to recognize the importance of [the] radical change in his role in the modern industrialized and highly specialized society has resulted in a cultural lag on the part of librarianship that could, if it is permitted to continue, result

in institutional obsolescence." He deplored the "devastating schism" which has arisen between librarians and documentalists; two years later in an editorial in the same journal he continued the same theme—"the latter still equate librarianship with inactive storage, and the librarians still persist in their insistence that documentalists are mere bibliographic amateurs who clothe traditional library processes in an esoteric and incomprehensible jargon. Unfortunately, there is just enough truth in these two points of view to give each some validity. . . . Yet both documentalist and librarian are seeking a common goal Some way must be found to unite the peculiar strengths of each into a single cohesive force."

To this view I unreservedly subscribe. Regrettably, the schism seems to be widening, rather than narrowing. The persistence of the documentalist in defining his craft as something separate from librarianship may be interpreted as a reaction of outrage in the face of the reluctance of traditional librarianship to reassess, *in depth*, the principles and techniques of our calling. We typically go around muttering pitiful platitudes to the effect that "it's all right in theory but it won't work in practice" with an air of sanctimonious solemnity. I have the strong feeling, reinforced by the intellectual thrust of books such as those here under review, that we librarians would be well advised to have the vision and the good grace to find ways of admitting some of the lesser documentalist heresies into the body of library canonical doctrine.—*Frank B. Rogers, National Library of Medicine.*

IBM Circulation Control

IBM Circulation Control at Brooklyn College Library. By Henry Birnbaum. White Plains, N. Y.: International Business Machines Corporation, 1960. 32 pp., Free.

Compromise between the desirable and the economically feasible has dominated circulation control records of libraries for three quarters of a century. The application of modern technology (simple as it was) to the

problem a little more than three decades ago merely increased the variety of experiments in compromise.

These first applications (Dickman and Gaylord) merely mechanized the recording of borrower identification on the book card systems of the day. The second type of application utilized IBM equipment to create punched call cards, thereby eliminating the established book cards, but maintaining a reference file essentially equivalent to the former book card files.

The third type of application of technology (Photocharging) abandoned the classed reference file and maintained the records of loans in transaction sequence. The key to this file, and to those of the numerous adaptations of it, was the prenumbered transaction card. Later adaptations utilized IBM punched transaction cards to further mechanize the clearance of the record of books returned. With this type of system the compromise moved far to the side of economic feasibility.

The transaction card systems appealed primarily to public libraries, but Brooklyn College adopted and used one modification for some time. As indicated by the author, who is chief circulation librarian it was found that too much had been sacrificed particularly in collegiate libraries. It was necessary to provide answers to the question, "Where is the book I need?" The Brooklyn answer was an ingenious combination of the second and third approaches to mechanization. By combining the IBM call card and the IBM transaction card, automation is carried further than with call cards alone, and more information is provided than by transaction cards alone.

Yet there are still compromises between the desirable and the expedient. By maintaining the file in sequence by the numerical portion of the Cutter number, the amount of key-punching is reduced and a numerical colator, rather than the more expensive alphabetic model, will suffice for filing, but the limitation to one thousand combinations means that the file loses convenience of consultation. If there are 10,000 volumes on loan at one time, an average of ten call cards will be grouped without further arrangement under each punched number. It is reasonable to assume that hand sorting through fifty cards would not be unusual.