

- ences; TOXLINE, National Library of Medicine, toxicologic abstracts and references; CHEMCON, Chemical Abstracts Service, chemical references; PATS, Predicasts Inc., chemical marketing abstracts and references.
- (3) Benson, F. R., Tillmanns, E.-J., Charles, R. D., Hoiness, C. M., and Soltzberg, S., "A Current Literature System Based on CA Condensates and Central Patents Index," Mid-Atlantic Regional Meeting, Philadelphia, Pa., Feb 1972.
- (4) Tillmanns, E.-J., Boylen, J. B., and Kendis, M. S., "Atlas Biomedical Literature System—A Computerized Current Awareness and Information Storage and Retrieval System," *J. Chem. Doc.*, **11**, 242–248 (1971).
- (5) Benson, F. R., and Walck, R. A., "The Atlas Chemical Research Information System," *J. Chem. Doc.*, **8**, 88–93 (1968).
- (6) Kirschner, S., Kravitz, S. H., and Mack, J., "The Application of Computers to the Retrieval of Selective Information Regarding the Anticancer Activity of Coordination Compounds," *J. Chem. Doc.*, **6**, 213–217 (1966).

The Hercules Technical Information Division: Services, Special Systems, and R&D†

HERMAN SKOLNIK

Hercules Incorporated, Research Center, Wilmington, Delaware 19899

Received April 1, 1974

The following information groups and services in the Hercules Technical Information Division are described: Library (open literature); Files (Hercules proprietary literature); Translating; Editorial; Report, Journal, and Patent Literature; Literature Research; Chemical and Market Data; Computer and Mathematics; and Computerized Information Systems Design and Programming. Several special information systems and the place of R&D in the various operations and services are discussed.

Five years ago, I presented a paper¹⁰ on "Management of Operations and Services in the Hercules Technical Information Division" at the 4th Middle Atlantic Regional Meeting, February, 1969. Two points I directed attention to in that paper were: the need to change technical information operations and services in harmony with changes in objectives and goals of the parent organization; and the need to design new operations and services to meet new needs.

Many changes have occurred over the past five years, not only in chemistry and chemical technology, but also in the organizational set-up of Hercules, the Research Center, and the Technical Information Division. The charts I used five years ago to illustrate the organizational set-up of the Research Center and of the Technical Information Division are now a part of our history. The current organizational set-up of the Research Center is shown in Figure 1 and that of the Technical Information Division in Figure 2.

Obtaining, storing, and circulating documents are basic to all information operations.^{2,8} There are, however, three primary categories of documents which are essential components in information services within an industrial R&D environment:

1. Documents generally housed in a library, such as books, journals, patents, trade publications, and available to everyone for purchase.

2. Technical reports, generally housed in a special Report Files, written by or for R&D personnel and relevant to the ongoing R&D programs.

3. Technical correspondence, generally housed in a Correspondence Files, which, in a sense, is supplementary to and complements the Report Files.

Both the Report Files and Correspondence Files are unique for Hercules; they exist to protect proprietary infor-

mation but within a system that makes the documents readily accessible to those with a need to know.

These three document services, Library, Report Files, and Correspondence Files, constitute our information problem when viewed in terms of size, growth, and complexity. They also constitute an important and essential raw material base on which we can build new science and new technology, and by which scientists and engineers can know what is known in all areas of interests pertinent to the company's business and future growth.

LIBRARY

The Hercules Research Center Library contains over 30,000 books and bound volumes of journals, thousands of U. S. and foreign patents, all trade publications of interest to the needs of R&D personnel for chemicals and equipment, and other documents, and subscribes to approximately 700 journals. One of the objectives of the Library is to have books on the shelves before someone from the laboratories asks for them. Because the head librarian maintains a constant awareness of our R&D programs and of who is doing what, he is unusually successful in anticipating our needs for new books and in keeping journal subscriptions in harmony with our scientific and technological interests.

An extremely important service of the library is the circulation of journals routinely to those who need them on a continuing basis. So that the library can have journals available in the library during the working day and still serve readers routinely, the library subscribes to as many copies as necessary to accomplish both and to handle requests for specific articles. Routine journal circulation is handled by a computer system which matches each copy of a requested journal with the requesters by building number and room number on a computer-produced circulation card. The computer system is also tied in with the library's subscription order.

† Hercules Research Center Contribution No. 1634. Presented in the Symposium on Information Groups in Chemical Companies—Services, Special Systems, and R&D, 9th Middle Atlantic Regional Meeting of the American Chemical Society, April 24, 1974, Wilkes-Barre, Pa.

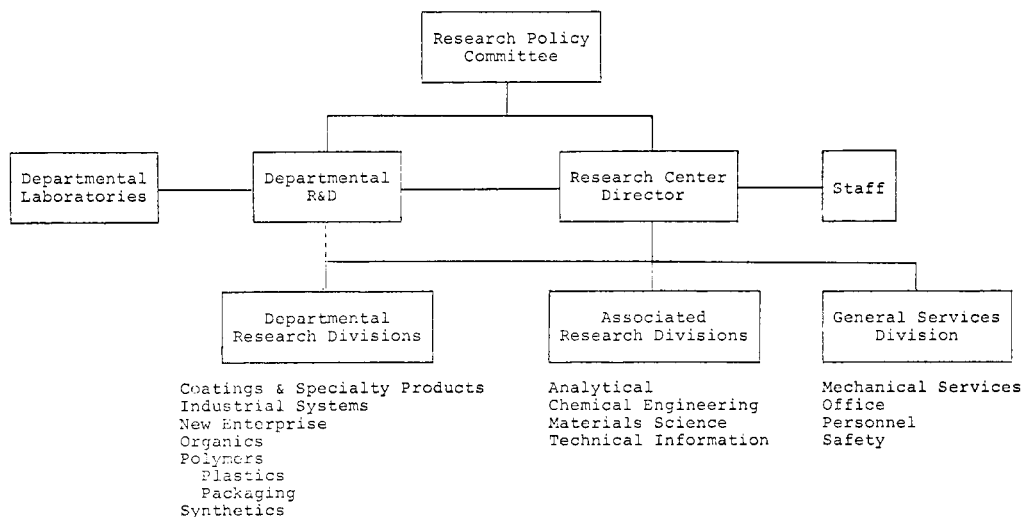


Figure 1. Hercules Research Center Organization.

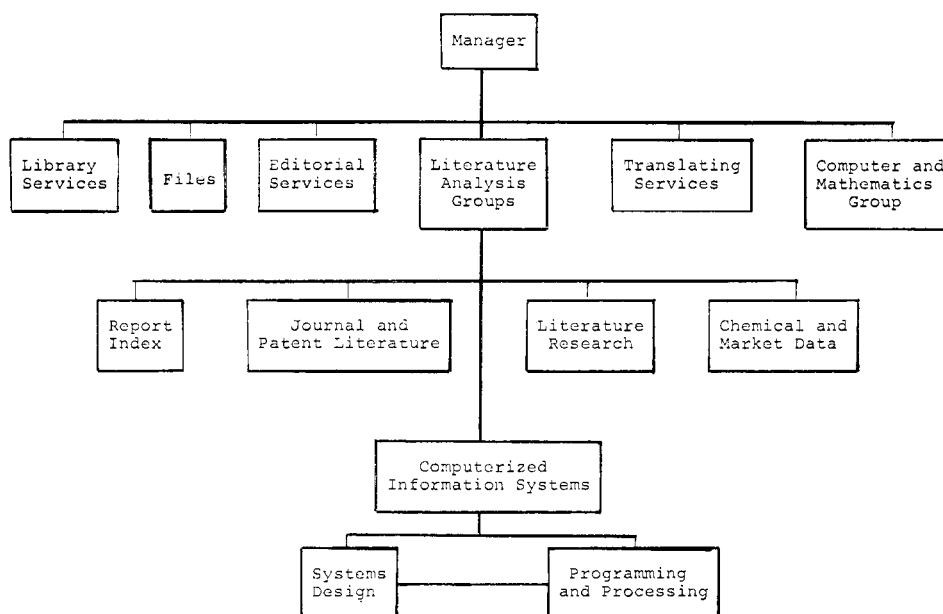


Figure 2. Organization of the Technical Information Division.

Almost every industrial library I have visited directs considerable attention to the cataloging and shelving of books by one of three systems: Dewey, Library of Congress, or Universal Decimal. We observed that each of these systems failed to reflect the realities of science and technology within an industrial R&D environment. Consequently, several years ago, we designed a new book classification system that was considerably more relevant to the areas of science and technology of interest and importance to our R&D personnel. In brief, the new book classification system recognized, for example, Polymer Chemistry, Agricultural Chemistry, Explosives and Chemical Propulsion, and Chemical Technology (Uses and Applications of Chemicals) as major classes, equal in importance with Organic Chemistry, Physical-Organic Chemistry, Physical Chemistry, Inorganic Chemistry, and Analytical Chemistry. In each major class, the books are cataloged and shelved in subclasses progressing from the most generic to the most specific.

The Library issues a monthly publication, *New Additions to the Research Center Library*, which lists with annotations new books, journals, translations, trade publications, government publications, and other documents it acquired during the month. It also issues monthly a list of

meetings, local and national, and courses of interest to Hercules technical personnel.

REPORT FILES

The Research Center Report Files is the central repository for all technical reports issued at all Hercules locations since the founding of the company in 1912. In addition, other reports, such as analytical data reports, testing reports, and pilot plant runs, and reports received through intercompany relationships or from outside contracting laboratories are also housed in the Report Files. Currently, the Report Files is accumulating technical reports at the rate of about 4,000 per year and data reports in the many thousands per year, each of which must be retrievable quickly.

CORRESPONDENCE FILES

The Correspondence Files processes about 40,000 letters per year to and by the Research Center technical staff, of which about 20,000 per year are retained for possible future value to our R&D programs.⁷

REPORT INDEX GROUP

Technical reports serve two primary functions in an industrial R&D environment:

1. As the major medium of communications to research management and others of the progress in, status of, and contributors to the R&D programs, technical reports provide the facts and conclusions for reaching decisions for action.

2. As the permanent record of the information, knowledge, and experience gained in the course of carrying out the programs.

Hercules research management was unique in 1928, 14 years after the founding of the company, when it assigned a chemist to the responsibility of setting up a report index. At that time, the chemist could not find a report indexing operation within the United States, but had to travel to Scotland (Ardeer) to seek aid and advice. The report index was set up, using the Kaiser system, although over the years the system did go through a series of relatively minor modifications. A completely new system was designed in 1958 primarily to bring the indexing language and concepts in harmony with the many changes that were occurring in chemistry and chemical technology, in the conduct and complexity of our research and development, and in the number and kinds of scientists in our environment.⁴ Again, in 1969, the system was changed radically to a computerized information system, using the Multiterm indexing concept which we conceived and developed in the late 1960's.¹²

Relating indexing with the information needs of the scientists in our environment has been an important factor in the design of all of our information systems. This has required that we ascertain these needs from the scientists themselves and by keeping abreast of the progress in their R&D programs and of the evolving literature relevant to these programs. Our extensive indexing activities have motivated us to be knowledgeable in every indexing and classification system described in the literature. But having learned these systems and using them experimentally, we have ended up generally by modifying them radically or by conceiving completely new systems.^{3,4,12}

JOURNAL AND PATENT LITERATURE

Being aware of and keeping abreast of the advances and developments in relevant areas of science and technology is a responsibility that each scientist needs to handle on a personal basis. This is not to say, however, that he does not seek help and aids in finding the majority of documents of interest to his research and to his continuing scientific education. Indeed, this responsibility is so great that research management motivates the scientists to meet the responsibility and provides whatever help and aids the scientists consider necessary to be agents of change rather than victims of change.

Obviously the most important step in aiding the scientists is to ensure that all important documents, such as journals and patents, are available. The documents, however, must be available for reading in the library, in offices and laboratories, and at home, as scientists do not fit one reading pattern, nor do they fit one pattern in requiring help. Consequently, it is necessary to have several mechanisms to serve these varying needs.

Currently, we employ the following mechanisms:

1. Multiple subscriptions to journals for availability in the library during the working day (one copy), for overnight reading, for routine circulation, and for specific requests.

2. Overnight reading, involving a messenger service, by which documents are delivered to offices and laboratories

at the end of the working day and picked up the following morning.

3. Routine journal circulation, involving our internal mail delivery, by which scientists receive selective journals they deem necessary to read on a regular basis.

4. Mail delivery of journals which contain articles brought to their attention in the Journal Literature Bulletin or of patents brought to their attention in the Patent Awareness Bulletin.

To a large degree, these document services are a part of an overall computerized system which also includes computer printouts of journal subscription orders; a list of journals received by the library for circulation to scientists to select journals for routine circulation; a list of journals each scientist is receiving to remind him periodically to add to or delete from the list as his assignments and responsibilities change; a list of journals received by the library during the week so scientists can know they are available for reading in the library or for overnight reading; and circulation lists of requests for specific articles in a journal or specific patents brought to their attention in the bulletins.

In general, a scientist tends to concentrate his regular reading on the few journals, from five to 15, in which he will find the majority of articles of interest to his research and to his continuing scientific education. For the remainder of the articles he may need on a relatively current basis, and for a retrieval system covering his total needs, he has recourse to our Journal Literature Bulletin and to its accompanying indexes.

Each issue of the 700 journals we subscribe to is examined on receipt for articles that are pertinent to Hercules products, processes, and R&D programs. The selected articles are classified, indexed, and abstracted for terminal input into a computer for production of the biweekly Journal Literature Bulletin which is sent to all Hercules technical people, over 500, who wish to use it for awareness of the evolving journal literature and for requesting specific articles.

A parallel operation based on patent gazettes and Derwent Abstract Services is also provided for keeping Hercules scientists aware of the evolving current patent literature and for their ordering of patents.

LITERATURE RESEARCH

Literature studies prior to and during a laboratory investigation are a vital and necessary activity of every mature scientist. No one should displace this obligation. But many occasions arise within an R&D environment, particularly for R&D teams, when this assignment is delegated most effectively to a literature scientist. Quite often in the preliminary stages of deciding on whether an R&D program should be undertaken, the literature scientist's report on background information can supply the facts for making the decision, and, should the decision be favorable, the report then becomes the basis for the laboratory scientists assigned to the program to extend their literature studies. Most importantly, when a literature scientist is given the assignment, a report is written and is available to others; a laboratory scientist may consult the same literature, but in general he will keep the information to himself and rarely will he write it up for others to use. Literature reports are particularly effective when they relate to the needs of members of an R&D team and when they evaluate the ongoing R&D with the facts in the open literature.

To be of optimum value in an R&D environment, the literature research chemist must be technically competent, have a good knowledge of the literature and of indexes to the literature, be a skillful searcher, have a good reading knowledge of several languages, and have better than average skill in writing.

MARKET RESEARCH

Market research data and information are essential within an R&D environment to inform R&D personnel of market needs and opportunities and of the current and potential business areas of the parent company. The primary problems involve the gathering, analyzing, and communication of data and information which occur in widely dispersed media. Finding and evaluating reliable information requires a person with unique competence and skills, which must be combined with skills in setting up and maintaining a viable information and communication system.

TRANSLATING SERVICES

Chemists with linguistic reading and speaking skills are an asset in an R&D environment, as language can be a major barrier to the chemical literature or in international scientific and technological interrelationships. Although many of our scientists are multilingual or have linguistic skills, the majority is not skillful beyond one or two languages.^{1,17} We have had full-time translators on our staff since 1927, and currently we have linguistic skills in 14 languages spread among about half of the chemists in the division. One chemist with outstanding linguistic skill in six languages is available full time for oral or formal translations on request.

COMPUTER AND MATHEMATICS SERVICES

This group provides the computer facilities, software, programs, and mathematical skills for handling the variety of complex mathematical problems that may arise in the conduct of R&D work. The group has had unique success in automated data analyses for analytical and testing instruments and in using computer simulation techniques for a variety of management and process problems.

EDITORIAL SERVICES

The division has had the delegated responsibility for maintaining a relevant technical reporting system that meets the needs of R&D management and laboratory personnel throughout the company.¹⁴ The division is also the responsible liaison between Hercules technical personnel preparing papers for presentation at technical meetings or for publication in technical journals and those in the company who must review the papers for release. Several members of the division are nomenclature experts and are consulted by Hercules scientists and engineers for solutions to nomenclature problems.

SPECIAL SYSTEMS

That the Technical Information Division has extensive holdings of documents, such as books, journals, proprietary reports, and that the members of the division are experts in all areas of information science are well known throughout the company. Thus, in addition to the approximately 300 scientists at the Research Center, hundreds of scientists at over 40 other locations are heavy users of our document resources and of our services and skills. Hardly a day goes by without a request from one of these locations for answers from and questions of the literature. We are currently involved in over 50 special information systems for as many groups throughout Hercules.

Whereas information systems such as for journals and patents and for Hercules proprietary reports, as already described in this paper, are essentially for all Hercules technical people, special systems may be set up to meet the needs of an R&D team, a group of scientists within a specific area of a discipline of science, a sales or development group

within an operating department, or an entire department.

In designing and setting up special systems for various groups throughout Hercules, we take advantage of our operations for the journal, patent, and Hercules report information systems. We also designed these three major information systems for processing and production by computers. Having examined a document for indexing and abstracting in one of the three systems, it is relatively easy to process it for a special system, with the same or different indexing and abstract, and to have the computer produce as many products as may be required.

Some special systems, however, are unique in terms of documents, data, indexing, and abstracting.

One such system is our computerized information system for the Systems Group of the Industrial Systems Department.¹³ Scientists at four locations required an awareness and retrieval system for the report literature from Hercules R&D and from U. S. government agencies and contractors work in chemical propulsion and all related areas. The system we designed and set up consists of a weekly bulletin which informs the readers of the over 100 reports, and their contents, received at each location. Because the weekly bulletin is computer produced, the index terms, abstracts, and other information are accessible for the production of a variety of indexes, such as subject, author, company, government agency, contract number, report numbers, as a retrieval tool at each location.

Another unique system is our pesticide information system.⁶ This was designed to provide a bookkeeping mechanism for the thousands of chemicals synthesized in our laboratories and tested for their effect on plants and insects; to provide a mechanism for retrieval by chemical, plant, or insect; and to provide a mechanism for correlating chemical structures, functionalities, and moieties with their effect on plants and insects.

R&D

R&D is an essential activity in the design, setting up, and maintenance of information operations, services, and special systems in an industrial R&D environment. There are many variables that need to be considered in information operations and systems, ranging from methods for recognizing and establishing information needs to systems design. The art and science of chemical documentation, or chemical information science, is not such that we have solutions to our problems. Indeed, most of our problems are without universal solutions. Among the unsolved problems, that each of us must seek the best answers to, are classification, indexing, and notation systems; microform systems; filing systems; photoduplication and printing methods; communication systems; correlative methods for relating chemical structures with properties and applications; and the design of computerized information and data systems.^{3,6,9,11,12,15-17} But, even if there were universal solutions to our problems, R&D would still be essential for keeping our information operations and services within reasonable manpower and economic limits.¹⁰ The doubling of the literature over the past seven years is not a justification for doubling the manpower—it is, however, a justification for doubling the R&D effort toward finding solutions to our information problems.

LITERATURE CITED

- (1) Clouser, B. E., and H. Skolnik, "A Procedure for the Identification of Languages," *J. Chem. Educ.*, **32**, 118-122 (1955).
- (2) Skolnik, H., "Planning a Research and Development Information Center," *Research & Eng.*, **2**, 26-29 (1956).
- (3) Skolnik, H., and J. K. Hopkins, "Simplified Stoichiometric Formula Index," *J. Chem. Educ.*, **35**, 150-152 (1958).
- (4) Skolnik, H., "Designing an Index," *Percolator*, **35**, 26-31 (1959).

- (5) Skolnik, H., and W. R. Payson, "A New Posting Method for the Preparation of a Cumulative List," *J. Chem. Doc.*, **3**, 21-24 (1963).
- (6) Skolnik, H., and A. Clow, "A Notation System for Indexing Pesticides," *J. Chem. Doc.*, **4**, 221-227 (1964).
- (7) Skolnik, H., and M. H. Payson, "Designing an Author-Based Correspondence Information System," *J. Chem. Doc.*, **6**, 240-244 (1966).
- (8) Skolnik, H., "The Art and Science of Chemical Documentation," *Ind. Chim. Belge*, **32**, Special Suppl. No. 1, 100-102 (1967).
- (9) Skolnik, H., and R. E. Curtiss, "An Evaluation of TEXT360 for Producing Reports," *J. Chem. Doc.*, **9**, 150-154 (1969).
- (10) Skolnik, H., "Management of Operations and Services in the Hercules Technical Information Division," *J. Chem. Doc.*, **9**, 213-217 (1969).
- (11) Skolnik, H., "A New Linear Notation System Based on Combinations of Carbon and Hydrogen," *J. Heterocycl. Chem.*, **6**, 689-695 (1969).
- (12) Skolnik, H., "The Multiterm Index: A New Concept in Information Storage and Retrieval," *J. Chem. Doc.*, **10**, 81-84 (1970).
- (13) Skolnik, H., and B. E. Clouser, "Designing an Information Awareness and Retrieval System for Chemical Propulsion Literature," *J. Chem. Doc.*, **11**, 39-43 (1971).
- (14) Skolnik, H., and L. F. McBurney, "Technical Reports and Decision Making," *Chem. Tech.*, **1**, 82-85 (1971).
- (15) Skolnik, H., and W. L. Jenkins, "Evaluation of the IBM Administrative Terminal System and Magnetic Tape Selectric Typewriter for Text Processing," *J. Chem. Doc.*, **11**, 170-173 (1971).
- (16) Skolnik, H., "A Computerized System for Storing, Retrieving, and Correlating NMR Data," *Appl. Spectrosc.*, **26**, 173-182 (1972).
- (17) Skolnik, H., "A Multilingual Index via the Multiterm System," *J. Chem. Doc.*, **12**, 128-132 (1972).

On-Line Substructure Searching Using Fragment Code—A Proposal

MICHAEL E. D. KOENIG

Institute for Scientific Information, Philadelphia, Pennsylvania 19106

Received February 11, 1974

A file structure is proposed which makes use of a fragment code and patterns and "families" of patterns to accomplish on-line substructure searching.

Now that, thanks to reduced data storage and communication costs, the day of large scale on-line literature search systems is actually upon us—the adaptation of computerized chemical substructure searching to on-line techniques is of great importance.

This paper proposes a method of organizing fragment code representations of chemical structure so that on-line substructure searches can be accomplished with reasonable storage requirements and relatively responsive data manipulation times.

DISCUSSION

In order to make the exposition of this proposal as concrete as possible, it will be described in terms of Ringcode¹ (probably the most commonly used and widely known fragment code) as used in the literature services Ringdoc, Pestdoc, and Vetdoc, produced by Derwent Ltd.[†] Let us examine various approaches to organizing fragment code information on-line.

The most straightforward approach would be to treat each fragment or code punch as a separate entry point in an inverted file, but it is equally obvious that the number of postings per entry point would be rather formidable. In the Derwent literature services to date, there are nearly half a million abstracts with an average of about six Ringcode cards per posting, and about 25 punches per card. This amounts to about 75 million total Ringcode punch postings. File storage is obviously a nontrivial problem, but even more of a problem is the manipulation of the huge data sets that would be necessary. A typical search would require Boolean logic operations on numerous lists, frequently hundreds of thousands of items in size. Response time would be intolerably long.

A second approach would be to treat patterns as index terms, and to do substructure searching by searching the

inverted file for an appropriately broad pattern, then bringing in the corresponding document records from some quickly accessible sequential file, and doing a bit search on-line of those records. Substantial storage must still be relatively quickly accessible for the sequential file. Such a system is also critically dependent upon the previous definition of an appropriate pattern. Patterns that are rather specific run the risk of being too narrow, and patterns that are more general run the risk of matching and pulling in very large portions of the total file, in effect, becoming sequential searches, with serious degradation of response time resulting.

A third approach, which I am proposing, is an extension of the second, or perhaps more accurately, a combination of the first and the second. It combines carrying both patterns and some individual punch postings in an on-line inverted file.

Let us first look at the most direct method of constructing such a system. Each pattern, with its associated document numbers, would be an entry in the inverted file. In addition, when each document was processed to see which patterns (if any) fit, that pattern which matched the greatest number of punches would be designated as the Most Complete Pattern (MCP). The punches not included in that pattern would be separately entered in the inverted file. This, plus a directory of what patterns each punch was to be found in, would in fact be a complete representation of all the Ringcode information in inverted file format.

Storage requirements are substantially reduced. Assuming two pattern hits per Ringcode card (allowing for overcoding, and for both general and specific patterns), and an average of six Ringcode cards per document, the pattern portions of the file requires approximately six million postings. Assuming an average of three punches per pattern card that are not included in the most complete pattern (MPC) for that card (many MCP's will of course be perfect fits, but one must allow for cases where no pattern is appropriate), the punch directory portion of the inverted file requires another nine million postings. The directory of

[†] Further information and specifications available from Derwent Publications, Ltd., Rochdale House, 128 Theobalds Rd., London, England.