Du Pont Central Report Index: System Design, Operation, and Performance*

BARBARA A. MONTAGUE and ROBERT F. SCHIRMER
Information Systems Division, Secretary's Department, E. I. du Pont de Nemours & Co., Inc., Wilmington, Del. 19898

Received January 11, 1967

The Central Report Index was established in the Du Pont Information Systems Division in 1964 to consolidate the scientific and technical report indexing, storage, and retrieval systems of nine separate departments. Objectives were improved services to technical personnel and better operating economy from more effective use of information personnel, facilities, and resources. The design of the consolidated system was based on an analysis of user requirements and the operating experiences of the predecessor systems. Development studies led to specifications for document handling, abstracting, indexing, searching, mechanization, file conversion, and organizational structure. Operation and performance of the current consolidated system are described.

In the highly competitive atmosphere of today's market place, improvements in current products, and research and development of new products are vital to the health of an industrial organization like the Du Pont Company. But research and development efforts are expensive and to be effective must build upon the foundation of previous technical achievements. To ignore past accomplishments can lead to costly duplication of effort.

Recognizing this requirement for effective access to past work, many of the departments in the Du Pont Company organized centers to make more readily available the technical information contained in their internal reports. These centers were based on modern methods of information storage and retrieval utilizing concept coordination associated with computer processing of indexes. The first such center was established by the Textile Fibers Department in 1950. By 1963, 11 separate departmental organizations were in existence, all using similar methods for information processing.

About this time, management became concerned with the proper flow of information among the various departments and encouraged information centers to exchange inquiries. This interdepartmental exchange greatly broadened the availability of Company technical information, but it was an obviously non-optimum method of operation because one inquiry could generate 11 separate searches.

Reappraisal of needs for internally generated technical information, and study of routes for more effective information transfer, led to adoption in April 1964 of recommendations that the separate information centers be consolidated as a single operating unit in the Information Systems Division. Objectives of consolidation were: improving information distribution across departmental lines, standardization of techniques and methods, better manpower handling, simplified and more efficient computer programming and services, better coordination of research and development work on information handling, and reduction of costs. The consolidated center, to be known as the Central Report Index, was assigned responsibilities to continue operation of the separate indexes for participating departments and to develop an optimum system for central operation.

Nine departments accepted the recommendations, and transfer of personnel and files began in July 1964 and was completed in September. Thereafter, the Central Report Index continued the operation of the nine separate indexes as agreed while the design and consolidation into a merged system was in progress.

DESIGN PHASE

The design phase was conducted by a development team of Information Systems Division consultants and a development coordinator from the Report Index. They reviewed and considered analyzed user needs, currently used information processing techniques, and advanced concepts of potential merit. Final system design was based on the cumulative experience of the development team, their interviews with information system operators and users within Du Pont, and their contacts outside the Company. The consolidated system recommended for the Report Index was a balance of requirements, capabilities, costs, and factors involved in smooth transition to the new system.

User Interviews. Interviews with users and operators of the nine systems disclosed the following guides for system design:

> Depth of indexing should depend on the type of report. Chemically oriented reports require deeper indexing than nonchemical reports.
>
> Review of indexing is desirable for consistency and current awareness of indexers.
>
> Links are valuable retrieval aids under specific circumstances.
>
> Roles are valuable for chemicals but have less merit for nonchemicals.
>
> Second-level search capability is needed for chemical compounds.

*Presented in Symposium on Automation of Information Operations, ACS Middle Atlantic Regional Meeting, Philadelphia, Pa., February 1, 1968.

A thesaurus of terms is valuable for indexing and searching.

Manual and computer search capabilities should be provided.

Abstracts, supplemented by the documents themselves, should be provided for screening answers.

Indicative abstracts as answers to inquiries are preferred by users.

Most searches should be answered within 24 hours.

Inquiry Analysis. A representative sample of 5% of the questions addressed to the nine indexes during the previous year was analyzed to determine frequency of use and value for retrieval of the following inquiry parameters in a typical operating environment:

Chemical *vs.* nonchemical (general) terms.
Specific *vs.* generic terms.
Various types of question logic.
Links and roles.
Interchange of inquiries among the several indexes.

The analysis showed that 37% of the search terms were chemical and 63% nonchemical. A further breakdown revealed that 20% of the total terms were generic chemicals, 17% were specific chemicals, 20% were generic nonchemical and 43% were specific nonchemical terms.

Analysis of question logic showed negation to be of value although it is not used as frequently as intersections and unions.

In analyzing the value of roles and links, many of the questions were reprocessed in order to measure the effects of search variables. From the findings, it was concluded that chemical roles aided relevance, but nonchemical roles contributed only a small increase to relevance while imposing a large reduction in recall. This information was of particular significance to system design considerations and is discussed further in the following section.

Study of interchange of inquiries showed that for each inquiry received by any of the indexes, almost three searches were made among the indexes.

Design Studies. Various studies were undertaken by the development team for design of the consolidated system.

A document control system was designed for recording receipt of reports and determining the status and location of all reports during the indexing, abstracting, editing, and keyboarding operations. Also a microfiche program was initiated for storage of the 43,000 reports in the inventory of the nine indexes.

The effectiveness of abstracts as screening devices was studied. This included an analysis of indicative *vs.* informative abstracts, author *vs.* indexer-generated abstracts, and length of abstracts. It was concluded that abstracts should be indicative, with a maximum of 225 words in length. Author or modified-author abstracts should be used whenever possible, although there are occasions when the entire abstract requires re-creation by the information analyst.

Study of the indexing function included depth of indexing, review indexing, links, roles, and term coding. It was concluded that chemical reports would normally be indexed to an average depth of 50 terms per document and nonchemical reports to an average depth of 25 terms. Both levels of depth, however, were dependent on the length of the report.

Examination of review indexing showed that it was worthwhile in providing consistency and completeness of indexing and provides current awareness for the reviewer. Review indexing should require only 5 to 10 minutes per report.

The study of links resulted in the following conclusions:

Links increase relevance but have more effect with chemical than nonchemical information.

Links do not reduce recall by blocking relevant documents; therefore, there is no penalty in their use from this standpoint.

Complex linking techniques such as master links (*1*) or sublinks (*2*) are generally not justified because they provide small return on investment and result in burdening the indexing and searching functions.

From the inquiry analysis, and study of indexing practices among the nine indexes, the following significant findings resulted: *For Chemical Information*: Roles improve relevance by reducing false retrieval, reduce recall unless applied under rigorously controlled conditions, and were extensively used by searchers in processing inquiries. *For Nonchemical Information*: Roles do not significantly improve relevance, reduce recall even when applied under careful control, and because of lack of reliability were used redundantly by both indexers and searchers, thus adding to operating costs without producing benefits.

It was concluded that roles for nonchemical terms were not needed and should be discontinued, but that certain roles had high value for chemical information and reactions and would continue to be employed. Consequently only the following roles of the ones previously used (*3*) were retained:

Role 3—Reactant
Role 4—Special Agent
Role 5—Solvent, Medium, or Catalyst support
Role 7—Product of a Reaction

The results of these studies on indexing depth, review of indexing, and the use of links and roles are consistent with those reported by other workers in this field (*1*, *4–9*).

Experience indicated that numeric term coding was the largest error-producing element in the existing mechanized information retrieval system. It was desirable, therefore, to find a quick, easy, and economical method of storing terms in a computer without manually assigning numeric term codes. An analysis of computer sorting techniques indicated that the size of the field to be sorted and the configuration of the field (numeric, alphabetic, special characters) had relatively little effect on the over-all speed and cost of the computer. It was decided, therefore, to represent all general terms (other than names of chemical compounds) in their alphanumeric form up to 35 characters.

Examination of general concept terms employed in the nine indexes and analysis of search questions processed during a one-year period indicated that the total number of terms needed for effective inquiry handling could be substantially reduced with resultant economy (*10*). Analysis of the 28,000 general terms in the combined indexes showed that only 2750 different concepts were contained in the 8000 most frequently occurring terms, the duplicate terms in the several indexes, and that 89% of the concepts employed in the one-year search question study involved these concepts. From these data it was

estimated that a consolidated thesaurus of 4000 to 6000 general terms was practical and could be attained by appropriate combining of some terms, unbinding of others, and consolidation of synonyms.

Analysis of search question experience showed also that for chemical terms the Report Index needed the ability to store and search organic chemicals, inorganic chemicals, and polymers at both the specific and generic levels. The analysis suggested, furthermore, that there was a substantial need for a generic search capability for substructure components by completely arbitrary and random combinations of atoms and bonds, not attainable through classification and schemes based on fixed combinations of atoms and bonds (fragments). An additional over-all system goal, in recognition of current file size and estimated future growth, was the elimination of variability in input and retrieval of chemical structure information that could result from extensive dependence on manual (human) processing of these steps. Accordingly, specifications were adopted that input and retrieval operations be mechanized and/or computer based to the maximum degree practical, and that such portions of the operations remaining on a manual basis have reasonable probability of being subsequently mechanized.

The system appearing to offer the best balance of generic capability, mechanization, and practicality for the environment at hand was the Chemical Structure Storage and Search System (CS_4) then being cooperatively developed by Du Pont (*11*) and Chemical Abstracts Service. This system is characterized as a topological scheme providing for the identification of each atom and its connecting bond(s) to any other atom(s) in the structure for indexing, storing, and searching of chemical structures. This system provided the additional feature of being aligned closely with the broad based system being developed for national use by Chemical Abstracts Service and of offering the potential of readily accepting chemical structure information generated by Chemical Abstracts Service for internal Du Pont processing should such handling eventually become desirable. It was judged also that the flexibility of this system would facilitate the production of special collections of chemical structure information into alternate system formats (fragmentation, cipher notation, and classification) if desired. This system likewise appeared to have special advantages for the merging of the several departmental chemical vocabularies since it provided a mechanism for combining identical compounds into one representation, regardless of the fragmentation or nomenclature system previously used by any of the nine indexes.

System Recommendations. As a result of the user interviews, inquiry analysis, and design studies, the development team made the following system recommendations:

> Abstracts should be provided for all reports, to be used as a primary screening device and as a product for the inquirer.
>
> Microfiche should be used as the permanent report record and as an additional screening device to supplement abstracts.
>
> Daily computer searches should be scheduled.
>
> The indexing function should include:
>
> a. Use of links.
>
> b. Use of chemical roles only.
>
> c. Review of indexing.
>
> Computer files should be updated monthly.
>
> Vocabularies and existing computer files of the nine indexes should be consolidated to provide an efficient, low-cost, one inquiry–one search method of operation.
>
> Information analysts should be organized by technology, not by function.
>
> A continuing program of long-range development and short-range improvements should be established to assure progressive and economical services.

CONSOLIDATION OF FILES AND IMPLEMENTATION OF NEW SYSTEM

Seven of the nine indexes used the IBM 705 for their file manipulation, one index used the IBM 650, while the other index utilized Termatrex optical coincidence cards, as shown in Figure 1. Since the Termatrex cards did not contain links and roles, it was decided to keypunch the 1000 indexing sheets to create temporary term/document computer files compatible with the other mechanized systems.

From these systems, approximately 73,000 punched cards representing the chemical and general term vocabularies of the nine separate indexes were created for consolidation. These cards contained a departmental code, a term code, and an alphabetic description for each term. The cards were manually separated into 45,000 chemical and 28,000 general terms.

Chemical Terms. Although consolidation of the chemical term vocabularies was a substantial task, this step was greatly facilitated and high accuracy of merging of identical structures assured by use of the Chemical Structure Storage and Search System (CS^4). The alternative manual route was evaluated as requiring greater expenditure of manpower (structure identification and combining of equivalent names) and having inherently lower accuracy.

For consolidation, technical personnel created structure drawings from the chemical names and computed a molecular formula for each drawing. Drawings were based on conventions adopted after consultation with Chemical Abstracts Service. Additional conventions, based on significant repeating units, were created for drawing polymer structures (*11*). Each structure drawing was coded manually by clerical personnel to produce a list of atom-bond connections for keyboarding input. Keyboarded input included additionally, the molecular formula, departmental code identification, and associated document (report) number(s). All structure input records were proc-

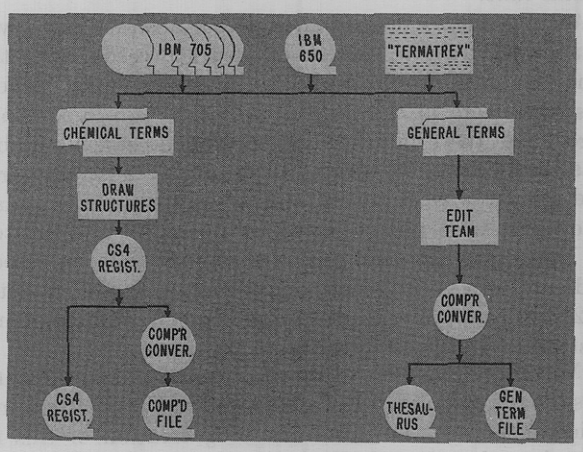

Figure 1. Consolidation of files

essed through a computer routine that created a unique atom-bond connectivity table in canonical form, identified those structures that were identical, assigned each such structure a unique number (Registry Number), and posted the corresponding report numbers to the Registry Number. Thus, as a result of these operations two files were created: a magnetic tape Registry File of chemical structures as atom-bond connectivity tables, to be used for substructure (generic) searches, and a Chemical Compound/Document File.

Since the consolidation operation was carried out over a period of 2 years, an additional 15,000 chemical compounds from current indexing work were included in the conversion collection. For these 60,000 structure representations, 43,000 unique chemical structures were produced in the Chemical Compound/Document File. At this stage of consolidation, compound/document postings totaled 320,000.

General Terms. A two-man editing team consisting of a PhD chemist and an engineer who were experienced in general term editing and thesaurus preparation was assigned the task of creating a consolidated thesaurus from the 28,000 different general terms from the nine indexes. The final product consisted of 11,000 terms of which 5000 were names of equipment, devices, techniques, processes, and properties. The remaining 6000 terms were trade names.

This editing assignment required 18 man-months of effort to review each of the 28,000 terms for choice of word form, elimination of synonyms, and building of relationships and cross references (*9*). The theasurus relationships used were USE and USED FOR for see and seen from references, BROADER TERM and NARROWER TERM for generic-specific relationships, and RELATED TERM for near synonyms. Traditional parenthetical scope notes were used to define meanings and differentiate terms with identical word forms but different meanings, such as TANKS (Military) and TANKS (Containers).

Conversion of the nine general term files followed completion of the editing effort (Figure 1). The previous departmental codes and the new alphanumeric description of the general terms were punched into IBM cards. These cards were merged against the existing term/document files of the nine systems to produce a consolidated General Term/Document File containing 11,000 terms with 750,000 postings. A thesaurus of general terms was also produced at this time.

Abstract File. A consolidated file of abstract cards was established from a variety of abstract cards from the nine indexes which had abstracts available. These abstracts were in the form of 3 × 5, 4 × 6, 5 × 8-inch cards and IBM punched cards. The accession number was handwritten on each of the available abstract cards and the cards were sorted into accession number order. The cards were photoreduced on microfilm and printed by Xerox copyflow onto continuous 5-inch wide card stock and cut into 3 × 5-inch unit abstract cards. A total of 20,000 cards were made in this manner. The remaining abstracts were typed to complete the file.

Microfiche File. One of the features of the consolidated system is a complete set of all indexed reports in 4 × 6-inch microfiche form. Details concerning creation of this file have been described by Crow (*12*).

Description of Computer System. The consolidated system employs an IBM 360/40 computer emulating the IBM 1410/7010 for updating and searching. The system is batch oriented and is characterized by two inverted term/document master files, one for compound Registry Numbers and the other for general terms as shown in Figure 2. Integrated with these master files are the Chemical Structure Storage and Search System (CS[4]) Registry File, a Thesaurus File, and a Chemical Name/Compound Number Validity File. This latter file was implemented to eliminate errors in manually recording compound numbers. Indexers have the opportunity of writing chemical names consisting of a maximum of 48 characters on the indexing sheet which, after keyboarding, are machine matched against this Validity File to obtain valid compound numbers. Files are updated monthly and are available for daily searching. The capability to manipulate all or any combination of these files has resulted in a flexible computer processing system.

CONSOLIDATED SYSTEM OPERATION AND PERFORMANCE

Organization of Staff. The technical staff was organized into three groups, each specializing in a different area of technology and responsible for indexing reports and answering inquiries for the client departments assigned to that group. This grouping arrangement was adopted to cope with the spectrum of scientific and technical information in Du Pont reports, to enhance communications with client departments, and to provide appropriate supervision. The technical force consists of chemists, biologists, and engineers. A senior advisor was appointed for each of the three groups and is responsible for quality control and consistency in indexing and searching, training new employees in the departmental technologies with which they are concerned, and daily direction of technical efforts.

Indexing. Figure 3 illustrates the work flow associated with indexing of a report. Two copies of each report are received from the participating departments by a clerical function designated Document Control. For each report received, a document flow card is maintained on which is recorded the date of completion of each of the steps in the input procedure. An accession number is assigned to both copies of the report, consisting of a prefix of two alphabetic characters designating department

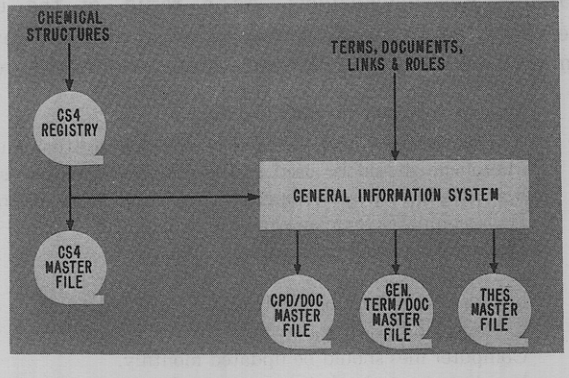

Figure 2. IBM 1410/7010 general information system

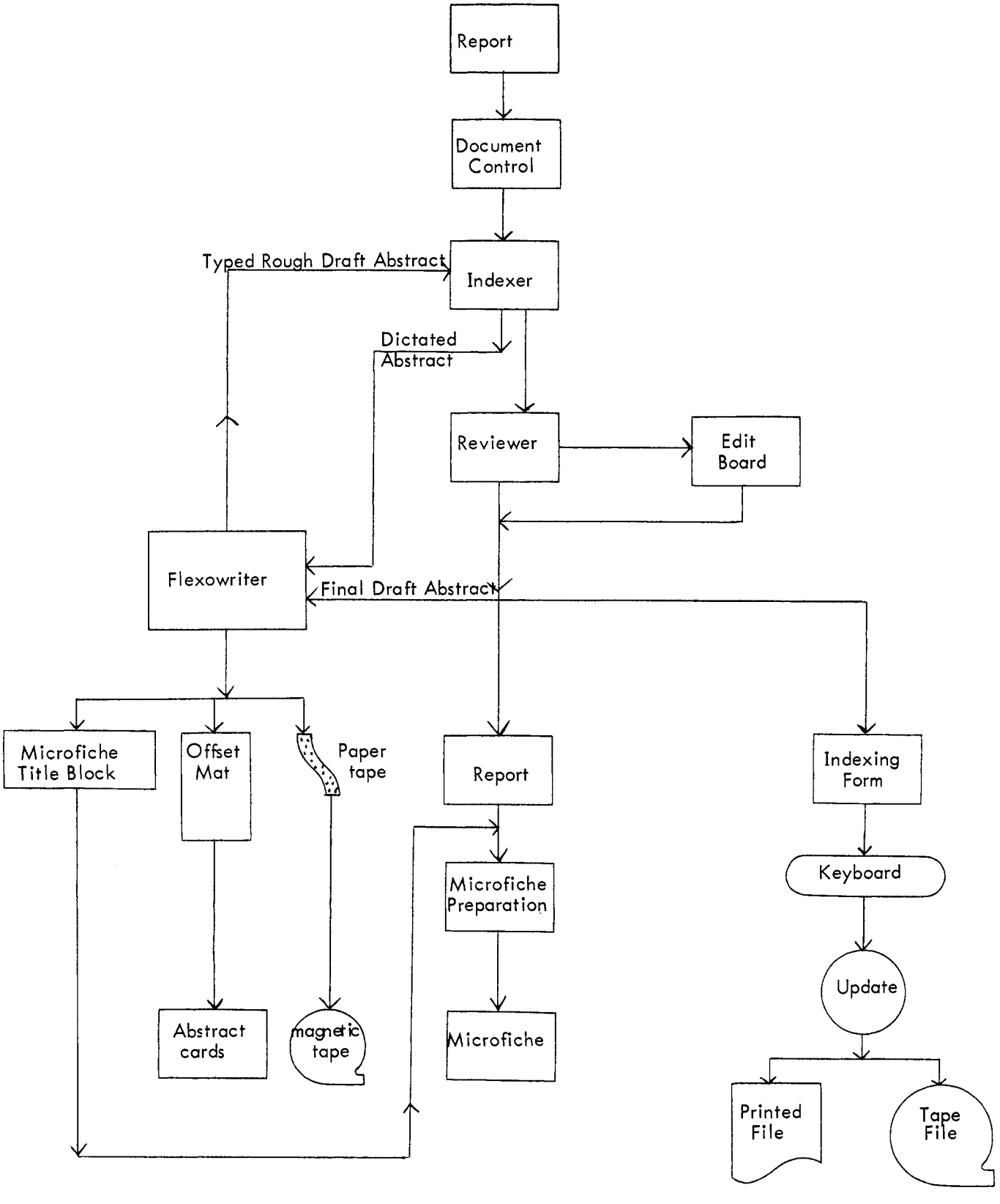

Figure 3. Flow of report input

of origin and type of document (*e.g.*, research, sales, or plant process) plus six numeric characters. Prior to indexing, a report is placed in a file folder with a document routing slip affixed to the cover which displays successive steps in the indexing procedure.

A report is assigned for indexing and abstracting to a member of the technical group responsible for handling the technology involved. Reports are indexed using a standard, edited vocabulary of general terms displayed in the Thesaurus and a chemical vocabulary based on *Chemical Abstracts* systems of nomenclature which is displayed in the Master Chemical List.

The indexer differentiates general terms from chemical terms by appropriate marking on an indexing form. Frequently used chemicals such as BENZENE and ACETONE are marked as common chemicals for subsequent machine matching against the Chemical Name/Compound Number Validity File at which time the compound number is substituted for the term name, BENZENE. The accession number is posted to the compound number during the updating of the Compound File. The Chemical Name/Compound Number Validity File is not limited to a standard system of nomenclature and contains various alternate names and synonyms of commonly used chemicals. It is an auxiliary tool to the Master Chemical List.

Abstracts are indicative rather than informative and include within space limitations, all the concepts that were indexed for system storage. This method of coverage is considered vital because the abstract is used by the searcher and the client to screen search answers for relevance. A close relationship between the indexing concepts which retrieve the report at search and the concepts included in the abstract which serves as a screening tool is considered imperative.

Abstracts are dictated using dictating machines to conserve technical time unless the author's abstract is satisfactory or easily edited. Many reports—those which do not contain complex chemical reaction information—also lend themselves to dication of the indexing terms together with links and roles.

Following release by the indexer, the indexing-abstract information is processed by clerical personnel to produce a typed abstract draft via an automatic paper tape activated typewriter (Flexowriter) which simultaneously creates a punched paper tape record of the draft. Clerical personnel also enter chemical compound Registry Numbers on the indexing form by lookup in the Master Chemical List (for those chemical compounds not automatically accepted on computer updating from the Chemical Name/Compound Number Validity File).

The indexing folder is then routed back to the indexer who edits the rough draft of the abstract and the indexing forms, and completes any new general term forms by suggesting relationships and any new chemical term forms by drawing the structures and computing the molecular formulae. Structure drawing is done according to rules and conventions in a manual of standard representations. These conventions represent vocabulary control in terms of chemical structure.

The indexing folder is routed next to the senior advisor of the group who reviews the indexing, abstract, and the nomenclature and structure of new chemical terms for consistency, accuracy, and coverage from both a systems point of view as well as technology.

The reviewer returns the indexing folder to the indexer to note any comments or corrections and to take necessary action. This procedure also serves to reinforce the learning process for newer employees.

If new general terms are generated by the indexer, the report folder is routed to the Edit Board, a group of senior advisors who control terms for addition to the Thesaurus. Comments or changes by the Edit Board are routed to the indexer for information. At this point the indexer is responsible for a final check of the contents of the indexing folder for completeness and accuracy. The folder is then returned to Document Control where the indexing form is removed and sent to keyboarding on Mohawk data recorders for subsequent computer updating.

The other contents of the folder are sent to their various destinations. The corrected abstract goes to the Flexowriter operator for final typing. The original punched paper tape is used to drive the Flexowriter and the operator intervenes only to make corrections, thus saving manual retyping of the entire abstract. The abstract is prepared in final form on a multilith mat from which the desired number of 3 × 5-inch abstract cards are duplicated. Two other products are obtained from this automatic typing operation: a title block for the microfiche of the report; and a punched paper tape containing the title, author, report number and abstract. The paper tape is converted periodically to magnetic tape for potential future uses under consideration.

Reproduced abstract cards are employed for maintenance of an Accession Number File, and Author File, and for distribution to the report originating points.

The second copy of each report, accompanied by a copy of the report abstract card and the Flexowriter paper tape produced title block information, is sent by Document Control to a separate internal Du Pont operating group for the preparation of a microfiche copy of the report (*12*).

Once a month, the Compound and General Term Files are updated as shown in Figure 2. Report accession numbers plus links are posted to compound number–role records in the Compound File and to general term records in the General Term File. Newly registered chemical terms are added to the CS[4] Registry File and new general terms are added to the Thesaurus along with their related terms.

A variety of listings of the different files are produced at updating. Complete master reference copies of both the Compound File and the General Term File are listed once a quarter with supplements being listed on intervening months. These listings are used for manual searching by visual inspection, and for system maintenance. The general term Thesaurus, the Chemical Name/Compound Number Validity File, and the Master Chemical List are printed quarterly and distributed to the indexing and searching staff.

Searching. The work flow associated with searching is shown in Figure 4. The answering of inquiries received from clients is assigned to the technical members of the departmental group who also index those departments' reports. A client calls a telephone number assigned for his department's inquiry use. The searcher negotiates the question with the inquirer by clarifying, suggesting, and

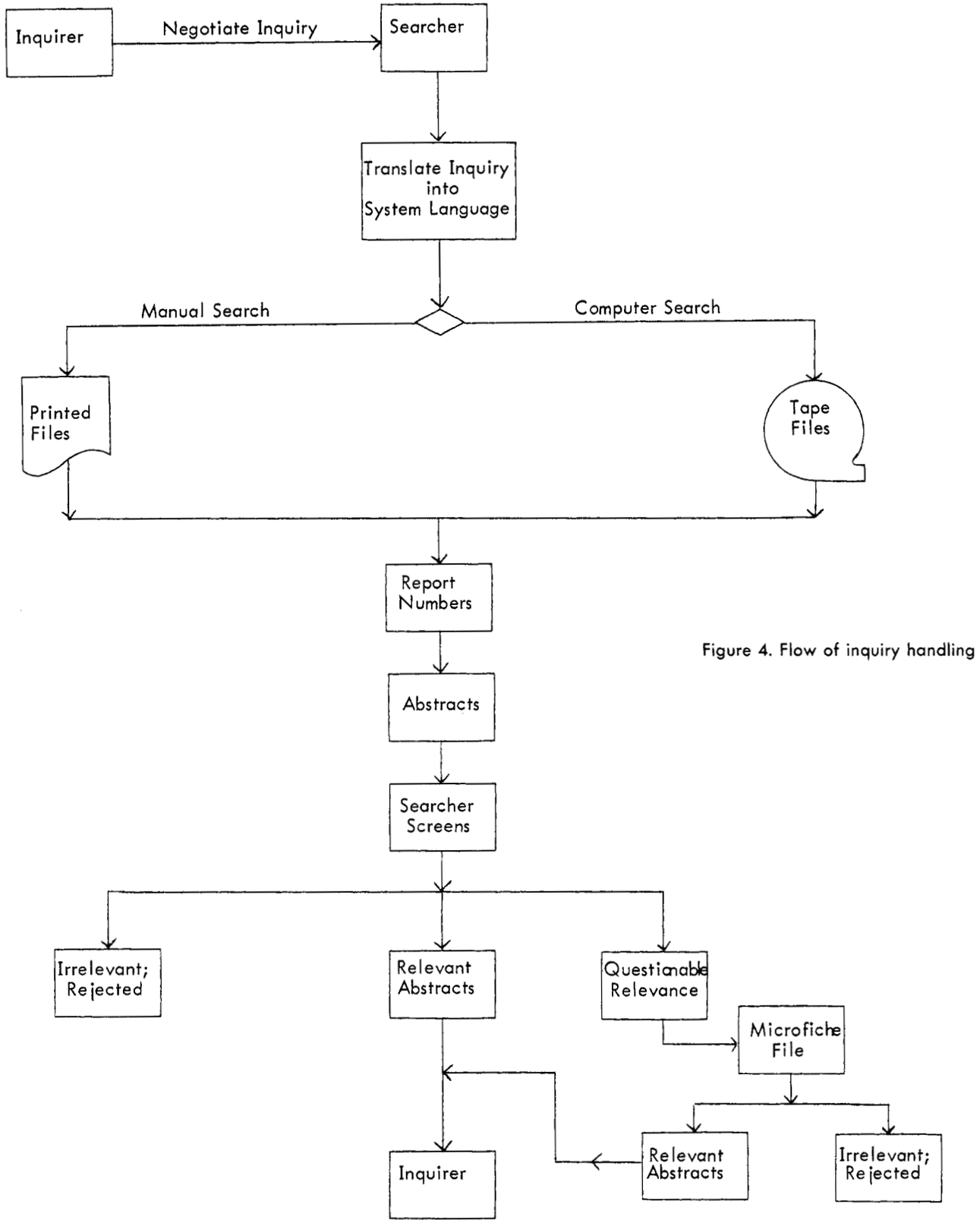

Figure 4. Flow of inquiry handling

narrowing down the inquiry. The searcher then translates the inquiry into system language by selection of chemical terms and roles, general terms, and inquiry logic. A decision is made between doing a manual search of the printed indexes, assisted by the clerical force, or a computer search. This decision is based upon the complexity of the search, the clerical time required to perform a manual search, and the need for a prompt answer. Four hours of clerical time is considered the break-even point for manual searching.

Searches are programmed in Polish notation for the computer by selection of logical operations to be performed on the terms in the question. The inquiry program provides five logical operations: unions, intersections, nega-

tions, an intermediate answer for an intersection, and an intermediate answer for a negation. Intermediate answers list the document references discarded by the logical operations of the search.

Search options include ability to select reports by: designated departments or the type of report desired by use of the alphabetic prefixes of the accession numbers, and date of issue, by use of limits on the numeric portion of the accession numbers.

As shown in Figure 5, searches for both compounds and general terms can be made at either a generic or specific level. Substructure searches of the CS[4] File produce listings of compound numbers which can be either printed or forwarded by computer to the 1410/7010 search system. Generic searches of general terms can be processed against the Thesaurus File to produce unions of narrower terms. Searches for specific compounds and/or general terms are then processed against the Compound/Document and/or General Term/Document Master Files to generate listings of document numbers.

Initial search answers are obtained as report accession numbers regardless of whether an inquiry is processed manually or by computer. The technical searcher screens answers for the inquirer by inspecting copies of the abstracts of reports nominated as answers to the inquiry. Since candidate abstracts are intended only for quick reference use, they are prepared at minimum expense from the Master Abstract Card File (in accession number order) by clerical staff using Thermo-Fax office copying machines. Where an abstract does not provide sufficient information for determining relevance, the searcher consults the microfiche of the complete report. Thermo-Fax copies of abstracts considered relevant to the inquiry are mailed to the client. After consulting the abstracts, the inquirer obtains copies of reports in which he is interested through his file room or report library of the departments which originated the reports. The Report Index does not provide reports in answer to inquiries but locates references to reports.

System Performance. The indexing philosophy used by the Report Index represents the composite experience of many different information scientists responsible for operating a variety of systems over a span of some 18 years in Du Pont. One of the skills most difficult to develop in new technical people being trained to index scientific and technical information is judgement of the proper depth of indexing. Depth of indexing comprises two elements, specificity and exhaustivity. Design studies concluded that adequate recovery of report information requires indexing at the same specific level reflected in the report. Exhaustivity, or the number of concepts indexed, at an average level of 25 terms per report for nonchemical information and 50 terms per report for chemical information as recommended by systems design studies, has proven entirely acceptable during the 18-month period this level has been in effect.

Likewise, system performance for the reduced set of roles and continued use of links adopted upon consolidation has been gratifying; indexing and searching have been greatly simplified and made less ambiguous.

The Report Index processes approximately 6000 new internal technical report accessions per year using a basic technical force of 15 chemists and engineers. Standard

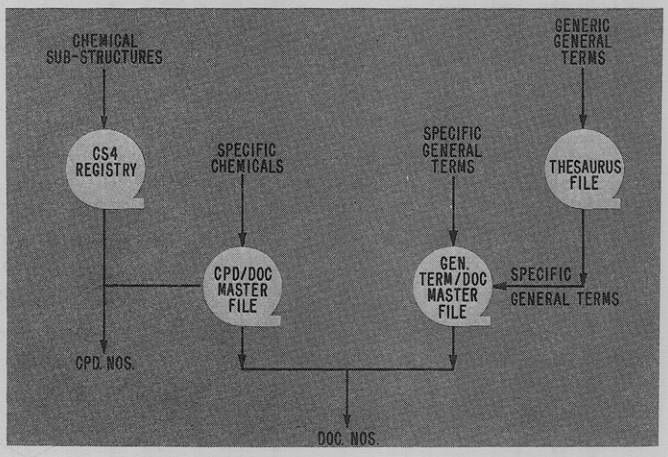

Figure 5. Computer Searching

levels of indexing production have been adopted to aid in attaining the best operating economics. Thus, average levels of technical time were set for the various elements of the indexing process. An over-all goal for indexing and abstracting the average report is 2.0 hours of technical time. This includes 1.6 hours for preliminary indexing and abstract preparation, 0.2 hour for review by the senior advisor, plus 0.2 hour for completion of indexing, drawing of chemical structures, and vocabulary control by the Edit Board. It is emphasized that these averages involve a range from a fraction of an hour per report to several hours.

The Report Index processes also approximately 3000 inquiries per year. Roughly 25% of these are readdressed by the Report Index to other information centers within Du Pont and to outside information sources. About 87% of the searches made on the consolidated system are run on the computer and 13% are made manually using the printed inverted indexes. Manual searches, however, do not provide generic search coverage for either chemical or general term information since the over-all system design specifies that generics will be created by computer at time of search.

Computer searches are run daily during the night shift so that any inquiry received prior to noon will be run the same night and computer answers be received the next morning. Three office copiers are used to expedite copying of abstracts for screening answers to searches; most abstracts are copied by noon. Abstract copies are screened, a transmittal letter typed, and abstracts mailed to the inquirer during the afternoon. A minimum time for computer searched inquiries is a little over 24 hours from receipt of inquiry to mailing of answers. Some inquiries are answered immediately by consulting the printed indexes. Other inquiries may be in process several days if they require CS[4] substructure searches and do not justify an independent CS[4] search. Substructure searches via CS[4] are normally handled as batch search runs to attain lower unit inquiry search costs. Annually, about 1800 computer searches are run on the Compound and General Term Files and 300 substructure computer searches are made. An average inquiry requires 2.3 hours technical time for setup and screening. An average CS[4] substructure search requires for equivalent operations 0.3 hour technical time.

SUMMARY

After one year of operation of the consolidated system, the value of the Report Index has been demonstrated in several ways. Of primary importance is the ability to retrieve reports from many departments in answer to a single query of the consolidated file. The number of searches per inquiry as a result of consolidation has been reduced from 2.8 to 1.3. This ratio is still greater than 1.0 due to the forwarding of inquiries to other information centers within and outside Du Pont. Consolidation has resulted also in standardization of techniques and methods, improved manpower utilization, simplified and more efficient computer programming and services, better coordination of research and development work for improvement of information handling, and better over-all economics.

ACKNOWLEDGMENT

A project of this magnitude obviously involved the work of many people. We gratefully acknowledge the direction provided by Carleton C. Conrad, Manager, and James G. Van Oot, Assistant Manager, of the Central Report Index. We also acknowledge the contributions of Robert L. Taylor to the microfiche program, of James E. Crow as a System Consultant, and of Bart E. Holm, Manager of the Development Section, who directed the development studies.

LITERATURE CITED

(1) Van Oot, J. G. *et al.*, "Links and Roles in Coordinate Indexing and Searching: An Economic Study of Their Use, and an Evaluation of their Effect on Relevance and Recall," J. CHEM. DOC. **6**, 95 (1966).

(2) Walker, J. F., "The Singularity Sub-Link—A New Tool for Use in the Storage and Retrieval of Information," *ibid.*, **4**, 45 (1964).

(3) Costello, J. C., Jr., "Storage and Retrieval of Chemical Research and Patent Information by Links and Roles in Du Pont," *Am. Doc.* **12**, 111 (1961).

(4) Herner, S., F. W. Lancaster, and W. F. Johanningsmeier, "Investigation of the Effect of Roles and Links on the Performance of a Mechanized Retrieval System," presented before the Division of Chemical Literature, 148th Meeting, ACS, Chicago, Ill., September 1964.

(5) Hyslop, M. R., "Role Indicators and Their Use in Information Searching—Relationship of ASM and EJC Systems," American Documentation Institute, Proceedings of the annual meeting, October 5–8, 1964, Philadelphia, Pa., **1**, 99 (1964).

(6) King, D. W., "Evaluation of Coordinate Index Systems During File Development," J. CHEM. DOC., **5**, 96 (1965).

(7) Lancaster, F. W., and J. Mills, "Testing Indexes and Index Language Devices: The ASLIB Cranfield Project," *Am Doc.* **15**, 4 (1964).

(8) Montague, B. A., "Testing, Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles," *ibid.*, **16**, 201 (1965).

(9) Sinnett, J. D., "An Evaluation of Links and Roles Used In Information Retrieval," ML TDR 64-152, Air Force System Command, Wright-Patterson Air Force Base, Ohio, July 1964.

(10) Schirmer, R. F., "Thesaurus Analysis For Updating," J. CHEM. DOC. **7**, 94 (1967).

(11) Hoffman, W. S., "An Integrated Chemical Structure Storage and Search System Operating at Du Pont," *ibid.*, 8, 3 (1968).

(12) Crow, J. E., "Microforms and Technical Information," presented before the Division of Chemical Literature 154th Meeting, ACS, Chicago, Ill., September 1967.

A Mechanized Information System for Many Outputs from One Input*

HERMAN SKOLNIK and RUTH E. CURTISS

Hercules Incorporated, Hercules Research Center, Wilmington, Del. 19899

Received December 1, 1967

A highly mechanized information system, oriented to the needs and requirements of an industrial community of scientists and engineers, is described. From a single input of IBM cards, current awareness bulletins are produced from the cards by means of an IBM 870 and cumulative printouts for retrieval are produced by means of an IBM System/360 from magnetic tapes prepared from the cards. The information system is designed to give total cumulative printouts and also selective printouts of information related to disciplines and missions of science, research and development projects, and individual and group interests.

In a broad sense, the primary needs of today's scientist, as it was for his forbears, are mechanisms for being aware of the current literature within his area of activity and for retrieving from the total literature that which he requires to be knowledgeable in his area of activity (*4*). Because of the size, complexity, and growth of the chemical literature, today's scientist cannot afford to be tolerant of information systems and services that do not give him the maximum number of documents within his area of interest and the absolute minimum outside of his area of interest. On the other hand, the economics of information systems and services dictate a design for a community of scientists, not for an individual scientist.

* Presented in Symposium on "Automation of Information Operations," ACS Middle Atlantic Regional Meeting, Philadelphia, Pa., February 1, 1968.