Thermochemistry," Vol. 1, F. D. Rossini, Ed., Interscience, New York, N. Y., 1956.
(26) Rossini, F. D., and Deming, W. E., "The Assignment of Uncertainties to the Data of Chemistry and Physics, with Specific Recommendations for Thermochemistry," *J. Wash. Acad. Sci.* **29,** 416–41 (1939).
(27) Shaw, R., "Heat Capacity of Liquids. Estimation of Heat Capacity at Constant Pressure and 25°C using Additivity Rules," *J. Chem. Eng. Data* **14,** 461–5 (1969).
(28) Sokolov, A. V., Kolesov, V. P., and Vorob'ev, A. F., "From the Scientific Council on Chemical Thermodynamics at the Institute of General and Inorganic Chemistry of the U.S.S.R. Academy of Sciences. Recommendations concerning the Publication of the Results of Calorimetric Measurements," *Russ. J. Phys. Chem.* **39,** 693–4 (1965).
(29) Stohmann, F., Kleber, C., and Langbein, H., "Ueber den Wärmewerth von Carbonsäuren der aromatischen Reihe," *J. Prakt. Chem.* [2] **40,** 128–57 (1889).
(30) Wagman, D. D., Evans, W. H., Parker, V. B., Halow, I., Bailey, S. M. and Schumm, R. H., "Selected Values of Chemical Thermodynamic Properties," *NBS Tech. Note* **270-3,** January 1968.
(31) Washburn, E. W., "Standard States for Bomb Calorimetry," *J. Research NBS* **10,** 525–58 (1933).

Multi-Level Retrieval Systems
IV. Large Systems*

LEE N. STARKER,** K. C. OWEN, and J. W. MARTIN, JR.
Science Information Services Department,
Warner-Lambert Research Institute, Morris Plains, N. J.

Received June 28, 1971

The computer phase of a Multi-Level Retrieval System designed for handling large data collections is discussed. This is an outgrowth of the techniques that were described for earlier systems where the data load was not heavy and/or where the system was queried on an infrequent basis. In all instances, the same data base has been carried along from one phase to the next. The system has been implemented with a minimum of cost and effort, as it utilized Inquire, a general information retrieval package of programs that had been purchased for a different application, but has also been readily applicable in this case.

In earlier papers of this series,[1, 2, 3] we described the first two steps in the development of a comprehensive information retrieval system. The design of this system was such that the mechanical aspects of the retrieval hardware could be applied to the problem of the moment, without fear that in a very short time the application would be outdated. When this occurred in the past, it often became necessary to redesign and restructure the system, and consequently, to modify and recreate the data base.

Designers are often accused of using an "atom bomb to kill a fly," when a flyswatter would have been more than adequate, because they tend to overdesign so as to avoid the need to rework the system in the immediate future. On the other hand, such overdesign may be justified if it is felt that there is a good chance that file size and/or usage of the system will grow with sufficient rapidity to warrant a more sophisticated approach. This is particularly so when it becomes difficult and expensive to reformat and reprocess the file for each new level of activity.

The Multi-Level approach that we have developed obviates many of the reasons for starting a simple system in a complex environment. This is possible because we have adopted, as a basis for this technique, keyword indices as the method for retrieval from our data bases. Keyword indices are of course readily adaptable to many manual, semi-manual, and automatic systems, and thus present a most desirable index format, when maximum versatility is required. The same data bank can then be used as we progress from the simplest to the most complex hardware. It will also be relatively independent of the hardware aspects of the system.

Thus the variable-field visual collation (peek-a-boo)[1] card (Figure 1), based on a modified IBM card, served to initiate new systems that were either small in size or to which infrequent reference was made. Use of this card was predicated on a do-it-yourself basis, but information-center assistance was often desirable. As these peek-a-boo systems grow they can be extended by the use of additional decks of cards (one new deck for each 500 documents). Each search of the file must now be carried through all of the decks, and when frequent referral is necessary, this procedure can become quite tedious.

At this stage of growth then, consideration can be given to conversion to a Termatrex system.[2] Since the J-400 Termatrex drill can use IBM cards as input data, it was only necessary to convert our original peek-a-boo cards to term/document IBM card formats. This was done by computer (Figure 2), and the resultant deck was used to generate a Termatrex system that corresponded to the earlier peek-a-boo cards, but eliminated many of the objections that had arisen.

We have used both facets of this Multi-Level System in our Library as a means of maintaining control over collections of documents dealing with our Warner-Chilcott Division's products. A separate file is maintained for each product and several of them have moved through the steps that we have presented.

*Presented before the Division of Chemical Literature, 161st Meeting, ACS, Los Angeles, Calif., March 30, 1971.
**To whom correspondence should be addressed.

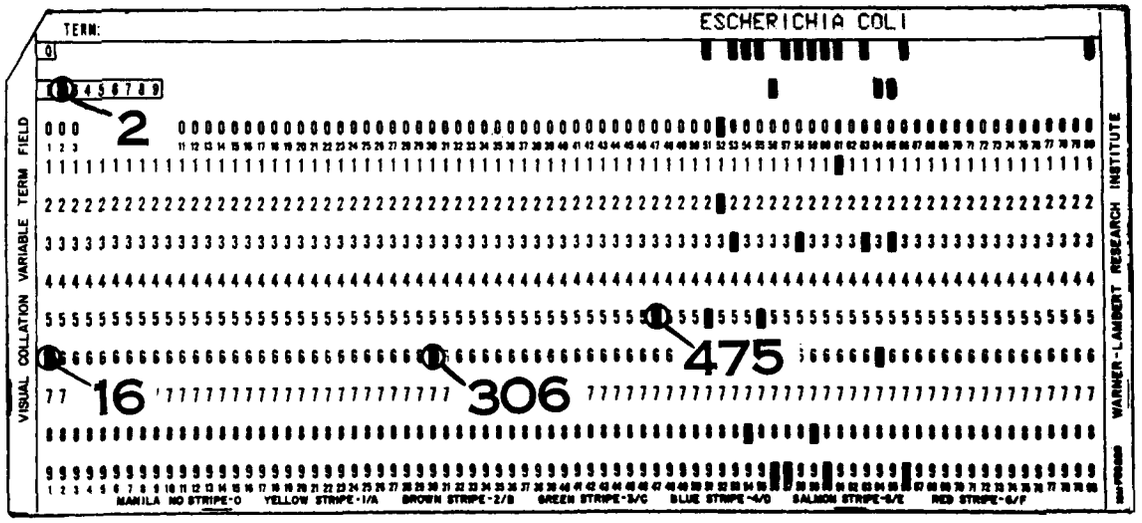

Figure 1. Visual collation variable term-field card

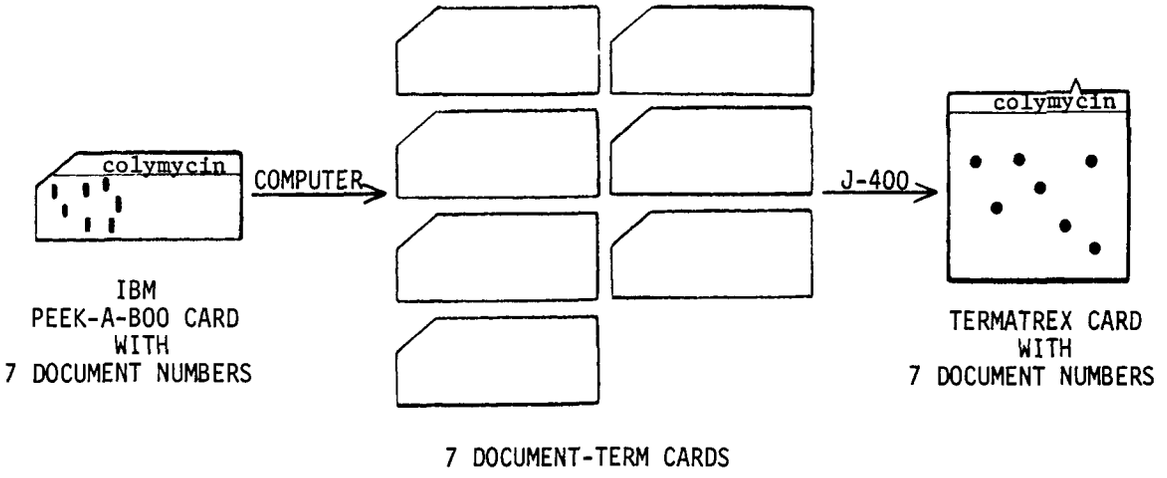

Figure 2. Conversion of IBM peek - a - boo cards to Termatrex cards

The major challenge to each aspect of the system has always been our Coly-Mycin file. The file currently contains nearly 5000 papers indexed with about 1500 terms and at a depth of 30 terms/paper. This number of documents is still well within the range of the Termatrex application insofar as the number of items is concerned, and most collections of this size could be handled very satisfactorily in this manner.

Termatrex systems yield document numbers only, as their response to queries, and many of our clients require full titles of the papers that answer their questions. We were, therefore, spending inordinate amounts of time in preparation of the response to the client. The actual typing time was, of course, further added to by the need to proofread, correct, retype, and reproof the lists.

We did consider an automatic typewriter approach to this problem, and have also used cut-and-paste techniques where feasible, but finally decided that our best answer was to computerize the system, so as to provide both search and report facilities in one operation. Similarly, a substantial saving in clerical time could also be anticipated.

Our considerations in this direction were also affected by the fact that we had acquired a proprietary search and retrieval package called Inquire.[4, 5] The Inquire System consists of a set of programs which create, maintain, and access data files. The System was written mostly in PL/1 and runs on an IBM System/360 computer.

INQUIRE

In particular, Inquire requires an OS/360 environment on a model 40 or larger. The programs may operate in a partition under MFT-2 or MVT. The data bases can be stored on any direct access device supported by OS/360, such as 2311, 2314, and Data Cells. PL/1 was chosen as the programming language because of its extensive, powerful data handling features. PL/1 MACRO facilities were used to provide well-defined interfaces between programs and files.

An Inquire data base is composed of three main files, the Index File, Search File, and Data File. The index file is an indexed sequential (ISAM) file containing one record for each unique keyword in the data base. Each record contains the name of the keyword, keyword code, keyword frequency count, and an address pointer to the search file. The pointer points to the disk address of a chain in the search file which relates all items referenced by that keyword.

The search file is a direct access (BDAM-by relative record number) file containing one record for each item in the data base. Each record contains a keyword code and a pointer to the next record referencing that keyword for every keyword used as an index for a particular item. In addition, each record stores the disc address, in the data file, of the actual item.

The data file is a direct access (BDAM-by relative track

and key file containing one record for each item in the data base. Each record contains the complete text of the item, segmented into fields. Only the contents of the fields are stored, not the names of the fields.

The names of the fields and the characteristics of the fields are chosen by the user at the time the data base is generated.

The Inquire System of programs are divided into three main subsystems, the Command System, the Loader System, and the Utility System.

The Command System performs search and retrieval, maintenance, output formatting, and editing. Queries to the system utilize DESCRIPTIONS to retrieve specific information. A description is a Boolean combination of keywords, field value conditions, and/or item numbers. An item number is a unique number assigned by the system to each item. A keyword is a combination of characters used to index an item. A field value condition is a relationship imposed on the values of a field, such as Field Name "contains" value, or Field Name "GT" value.

Maintenance commands allow the user to add (delete) items, change field values, change keywords, and add (delete) keywords. Print commands allow the user to custom design his output report including sorting the information, and computing simple expressions and totals.

The Loader System is used to create a new data base, and to make bulk updates to an existing data base. Validation and editing features are provided. These include checking for duplicated or undefined fields, too long fields, and omitted required fields. Items containing such fields are rejected and can be corrected for a subsequent update run.

The Utility System loads, unloads, reorganizes, and prints the index file. It also provides for moving, rearranging, and expanding the data file and search file.

The program was originally purchased to permit us to convert[6] a complex biological data system from the UNIVAC 1050, for which it was originally designed,[7] to the IBM 360

Our experience with this software package in our Bio-Data application led us to believe that it would be ideally suited for conversion of the Library's literature files since the Inquire programs provide us with a generalized information retrieval package that was particularly adapted to keyword searching. The program also provided for the searching of fielded information and for scanning or text searching. We were, or course, already dealing with a keyword system and we were certain that the ability to scan text titles in this instance would be of value.

The system also provided the necessary sort facility. We were, in addition, impressed with the versatility of the report generator capabilities of the package. A number of different output formats could be readily devised and implemented by the information scientist, at the time that each query was formulated. Frequently used formats could be written, debugged, and stored, to be called on as needed.

CONVERSION

Although our major needs were related to the Coly-Mycin file, we decided to pilot our conversion operation with the smaller Mandelamine decks. These covered approximately 750 documents which had generated about 3000 cards in the citation file and about 20,000 cards in the term file. Also available for the system was the Mandelamine master keyword file of approximately 1000 cards. The keyword cards carried the keyword number, the document number, and in some cases, the keyword name. The master cards carried the keyword name, the keyword number, and the designation of the related Termatrex card. They were originally used either to generate a list of keyword name *vs.* Termatrex card numbers or as master cards for gangpunching the keyword names into the cards generated by routine keypunching. We had built up over the years decks of cards punched with the citations of the thousands of papers in our various product-related files. These cards were used to generate periodic bibliographies —both alphabetically by author (Figure 3), and numerically by document number (Figure 4). Because of the volume of cards, and the fact that we maintained two complete decks to avoid constant sorting, card storage was becoming a problem. Ideally, we should be able to use these citation cards as well as our keyword deck for direct input to the new system, again without a need for rekeypunching, and make it possible to both search and print out listings of the papers.

Initially, input to Inquire required a matching up (Figure 5), of the citation cards and the term cards for each document. Subsequent update procedures permit random additions to the file.

VIEGAS, AULO PINTO AND MARTINS, J.D.: (ANOREXIGENS.) REV. BRASIL. MED. 20:309-13, JUNE 1963:	PS 0103
VOUDOUKIS, IGNATIOS J. AND PRASAD, A.G.S. (HENRY FORD HOSP., DETROIT): CHLORPHENTERMINE HYDROCHLORIDE (PRE-SATE): A PROMISING APPETITE DEPRESSANT FOR THE OBESE HYPERTENSIVE. CLIN. RES. 14:449, OCT. 1966. (SOCIETY REPT.)	PS 0117
WAGNER, H. (GIESSEN JUSTUS LIEBIG UNIV.): (MEDICAMENTOUS TREATMENT OF OBESITY.) MED. ERNAEHR. 2:25-7, 1962.	PS 0033
WAMBSGANSS, HEINZ (KARLSRUHE MUN. HOSP.): (CLINICAL EXPERIENCES WITH AVICOL.) MED. WELT PP. 2739-41, NO. 51, DEC. 22, 1962.	PS 0100
WAREMBOURG, H. AND JAILLARD, J.: (CLINICAL TRIALS WITH S. 62-2: CONCERNING 40 CASES.) LILLE MED. 9:962-7, NO. 10, 1964.	PS 0065
WEISCHER, MARIE-LUISE AND OPITZ, KLAUS (UNIV. MUENSTER, WESTF.): (DETERMINATION OF CHLORPHENTERMINE, FENFLURAMINE AND RELATED SUBSTANCES IN URINE.) ARZNEIMITTELFORSCH. 17:625-7, MAY 1967.	PS 0130
ZIEGLER, G. (ALSTERBERG NURSING HOUSE, HAMBURG-FUHLSBUETTEL): (ATTEMPTS AT CONTROLLING THE APPETITE IN OLD PATIENTS AND DIABETIC PATIENTS.) MED. ERNAEHR. 3:284-7, 1962.	PS 0054

Figure 3. Citation listing (alphabetic)

PS 0137	SCHALLY, ANDREW V., REDDING, T.W., AND LUCIEN, H.W. (TULANE UNIV.): ENTEROGASTRONE INHIBITS EATING BY FASTED MICE. SCIENCE 157:210-11, JULY 14, 1967.
PS 0138	SEDAM, R.L. AND TICE, L.F. (PHILADELPHIA COLL. PHARMACY SCIENCE): NEW DRUGS OF 1965 and 1966. AM. J. PHARM 139:70-1, MAR.-APR. 1967.
PS 0139	GOZARIU, L., FLORESCU, O., AND MADAR, J. (UNIV. CLUJ, ROMANIA): (THE EFFECT OF CHLORPHENTERMINE (LUCOFEN) ON THE "FREE" AND "BOUND" PLASMA INSULIN IN RATS.) ENDOKRINOLOGIE 50:36-40, NO. 1-2, 1966.
PS 0140	VAGUE, J., ET AL. (UNIV. MARSEILLE): (CLINICAL STUDIES OF THYROGLOBULINE 100.) GAZ. MED. FRANCE 73:1-3, APR. 25, 1966.
PS 0141	DUNLOP, DERRICK AND ALSTEAD, STANLEY (EDS.): TEXTBOOK OF MEDICAL TREATMENT. LONDON, E. & S. LIVINGSTON LTD., 10TH ED., 1966, PP. 369-70.
PS 0142	BARNES, H.W., ET AL. (SANDOZ): EFFECTS OF VARIOUS CNS-ACTIVE SUBSTANCES AND CNS-MODIFYING INFLUENCES ON MOUSE-KILLING BEHAVIOR OF RATS. PHARMACOLOGIST 9:200, FALL 1967. (SOCIETY REPT.)
PS 0143	CLARKE, E.G.C. (ROYAL VETERINARY COLL., LONDON): THE IDENTIFICATION OF AMPHETAMINE TYPE DRUGS. J. FORENSIC SCI. SOC. 7:31-6, JAN. 1967.
PS 0144	MC IVER, A.K. (NORTH LONSDALE HOSP., BARROW-IN-FURNESS): DRUG INTERACTIONS. PHARM. J. 199:205-10, SEPT. 2, 1967.
PS 0145	GYLYS, J.A. (WLRI): ANORECTIC AGENTS AND PROGRESSIVE RATIO IN THE RAT. ARCH. INTERN. PHARMACODYN. 169:354-61, OCT. 1967.

Figure 4. Citation listing (numeric)

One unexpected problem that occurred resulted from a limitation of the space available for the keywords associated with each document. The depth of our indexing was such that two of the first 97 papers entered were rejected because of insufficient space for the assigned keywords. The problem, however, was a minor one since the space problem applied only to the number of keywords that could be printed out if needed. But the number of keywords that could be put into the search portion of the system was not altered. We were able to include all of the keywords for these two documents, simply by rearranging them to make maximum use of the space available. The price we paid was to place several keywords out of alphabetical order.

We also discovered that during the development of the Termatrex system, we had not always been wholly consistent in the spellings of given terms (anti-infective, a/infective; ammonia, ammonium; benezene, benzene; etc.). The Inquire system yields on a routine basis, a keyword count on each input and these variations in spelling appeared on the initial loading as different keywords. Once the keyword list has been established, however, new terms —or incorrect spellings—are rejected, unless they are introduced by accepted "add" and "delete" routines.

Current operations now require that only a term number be punched. This speeds up punching and eliminates the spelling problems.

Other advantages include:

The possibility of combining all of our own documents into one master file. All searches, which were intended to be limited to papers on a single drug would have as one of the terms in the search query the symbol of the drug required (MA for Mandelamine, CO for Coly-Mycin, etc.). On the other hand, the desire to run a search through all of the materials in our files—how many clinical studies have been run on populations of over 100 patients—could be handled very simply, by eliminating any term that referred to a specific drug.

Up-to-date term lists, with frequency counts are generated on a routine basis with each update of the file. This permits ready analysis of the terms in use and their frequency, making possible a continuing evaluation of the indexing dictionary. This list can also be used to estimate the number of citations that will result from a given query.

Other analytic data also become more readily available. Not only can we determine the number of papers per term, but also terms per paper, number of papers per year, number of papers containing any given combination of terms, number of papers per author, or per institute, etc. In most cases, this information is obtained by phrasing a search question and then adding the command COUNT. This yields only the number of citations found and does not print the bibliographic information.

The system is also suitable for use as an SDI medium. Profiles are coded into appropriate queries, which are retained and searched on a periodic basis. Since each update of the file carries an INDATE, these SDI profiles can be limited to only that material added to the file since the last update. The system provides for the storage of a certain number of questions, so that they need not be rephrased each time.

We fully expect that once the bulk of our product files have been converted, computer output will be used to generate the master copy for our *Abstract Bulletins*. Currently, this publication carries the citation to the paper and a list of the keywords that have been assigned.

Computer outputs will be used also to prepare listings for routine distribution to all of our branches and affiliates—both locally and overseas.

Rather than attempt to prepare a program that would prepare alphabetic listings according to a complex series of rules, we have adopted the technique of manually assigning incoming documents to their correct alphabetic position in the file, and then assigning a sequence number based on that position. Alphabetic sorting is then accomplished by sorting on the sequence number. The document number, of course, is used to arrange material in numeric sequence.

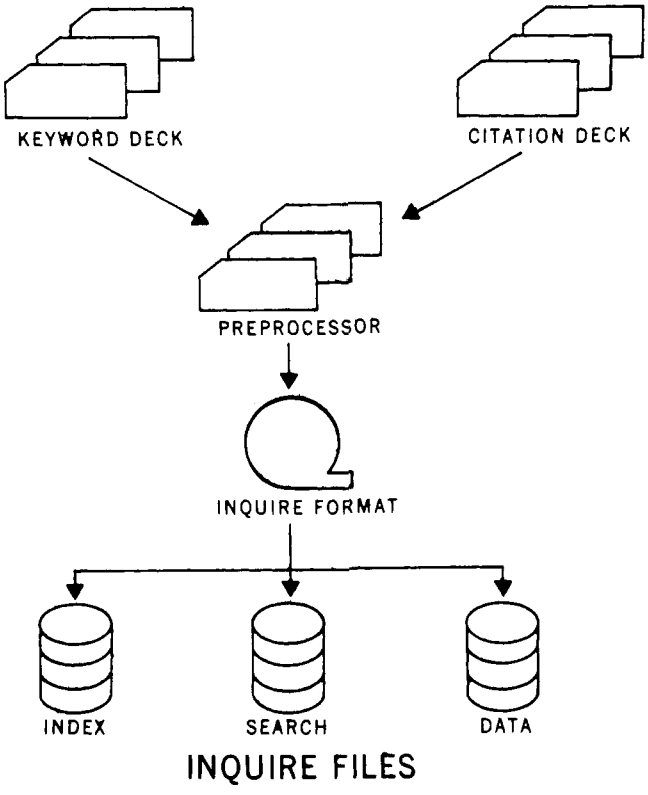

Figure 5. Literature file construction

SEARCH PROGRAMS

The query language used by Inquire is not overly complex and is easy to learn. Examples of several search requests and how they are framed should illustrate the techniques required:

A simple question might be: What U.S. publications on Mandelamine do we have that relate its activity with pH? Give me no more than 10 citations.

The Inquire command would be—

FIND PRODUCT=MA AND PH AND "UNITED STATES," TAB

PREVIOUS, LIMIT 10

The system commands are underlined:

FIND is the search command
AND denotes the Boolean logic to be followed
TAB starts the print format definition. In this case, PREVIOUS denotes the fact that the same format as was used in the preceding search should be followed; these instructions need not be repeated.
LIMIT 10 states that no more than 10 answers are to be printed.

A more complex question was asked for papers which discussed testing with Mandelamine sensitivity discs.

```
DOCMNTNO CITATION

MA 444   SENECA, HARRY (COLUMBIA UNIV.): CURRENT THERAPY OF IN-
         FECTIONS OF THE RENAL EXCRETORY SYSTEM. J. AM. GERIAT.
         SOC. 12:1100-27, DEC. 1964.

MA 483   SENECA, HARRY (COLUMBIA UNIV.): PYELONEPHRITIS; CLINICAL
         AND EXPERIMENTAL. J. AM. GERIAT. SOC. 13:947-66,
         NOV. 1965.

MA 820   ANON.: METHENAMINE HIPPURATE (HIPREX). MED. LETTER
         10:58-60, JULY 26, 1968.

MA 864   GERSTEIN, ALAN R., ET AL. (UCLA): THE PROLONGED USE OF
         METHENAMINE HIPPURATE IN THE TREATMENT OF CHRONIC
         URINARY TRACT INFECTION. J. OROL. 100:767-71, DEC.
         1968.

MA 989   RODMAN, MORTON J. (RUTGERS UNIV.): COMBATING URINARY
         TRACT INFECTIONS. RN 31:59-68, NOV. 1968.

ITEMS RETRIEVED        5

PROCESSING TIME -      7.4 SECONDS
```

Figure 6. Inquire printout long-term use of Mandelamine *vs.* Hiprex

The query was phrased for the computer as follows:

FIND PRODUCT=MA AND MANDELAMINE OR METHENAMINE/MANDELATE AND 'IN VITRO' AND SENSITIVITY AND DISK, SORT SEQUENCE, TAB DOCMENTNO 1 CITATION 9 FILLER 71, HEADER 'SENSITIVITY TESTING WITH MANDELAMINE DISCS'.

These commands order the system to locate all documents in the MA file which contain Mandelamine or Methenamine/Mandelate, in vitro, sensitivity, and discs as keywords. The results are to be sorted by sequence number (to give an alphabetic arrangement) and printed out so that the document number starts in print position 1, the citation starts in print position 9, but does not run beyond print position 71. The title of the listing is to be "Sensitivity Testing with Mandelamine Discs."

The system retrieved 39 items in 21 seconds, and a portion of the printout is shown in Figure 6. In this, as well as other instances, the printout can include the document number, the citations, and the keywords, singly or in any combination.

LITERATURE CITED

(1) Starker, L. N., and Cordero, J. A., "A Multi-Level Retrieval System. I. A Simple Optical Coincidence Card System," *J. Chem. Doc.* 8, 81-5 (1968).

(2) Starker, L. N., Owen, K. C., and Batson, B. C., "A Multi-Level Retrieval System. II. Medium-Sized Collections," *Ibid.*, **9**, 161 (1969).

(3) Starker, L. N., Kish, J. A., and Arendell, F. H., "Multi-Level Retrieval Systems. III. A Generic Chemical Search System Using Optical Coincidence Cards," *Ibid.*, **10**, 206 (1970).

(4) Infodata Systems, Inc., Arlington, Va. 22209.

(5) Bennett, R. E., and Frycki, S. J., "Internal Processing of External Reference Services," *J. Chem. Doc.* **11**, 76 (1971).

(6) Stein, J. D. Jr., Delaney, F. M., Peluso, S. D., and Starker, L. N., "A Computer-Based Comprehensive Bio-Data Information Retrieval System." Presented at the 6th Middle Atlantic Regional Meeting, ACS, Baltimore, Md., February 1971.

(7) Arendell, F. H., Lengler, C. R., Goebel, C. L., and Starker, L. N., "A Pharmaceutical Data Handling System for Scientific and Management Reporting Functions." Presented at 31st Meeting, International Federation of Documentation, Washington, D. C., October 1965.

Atlas Biomedical Literature System—A Computerized Current Awareness and Information Storage and Retrieval System*

EMMA-JUNE H. TILLMANNS,** JOYCE B. BOYLEN, and MERWIN S. KENDIS
Atlas Chemical Industries, Inc., Wilmington, Del. 19899

Received August 11, 1971

The computerized Atlas Biomedical Literature System produces a semi-monthly current awareness bulletin for published biological, biochemical, medicinal-chemical, and medical literature pertinent to Atlas research interests and products, and also stores keywords and bibliographic material for retrieval. Sources of input to the system are evaluated. Access to the systems and examples of the format of the data retrieved are discussed.

Various methods have been described for computerized storage and retrieval of technical literature. Some are used for storage and retrieval of company reports but can be extended to other types of literature, such as the GAF document storage and retrieval system[1] and the Olin SWIFT method.[2] Others are intended for files of published literature references.[3-5] Of the latter, the system described by Teal and Greenburg provides for the computer-printing of bibliographies in answer to queries and also the printing of bibliographies in the format and sequence required by the Food and Drug Administration (FDA).[5] None of these methods makes possible computer storage and retrieval and also computer-printing of a current awareness bulletin. Flexibility in changing the format of the output is also lacking in these previously published methods. The Atlas Bio-

*Presented before the Division of Chemical Literature, 162nd Meeting, ACS, Washington, D. C., September 13, 1971.

**To whom inquiries should be addressed.