mation would be found in the same section of the index, under the specific notation.

A second possibility is the generation of a functional group index for all chemical structures appearing in a catalog, textbook, journal, or secondary source publication.

Another possibility that is now being explored is the incorporation of biological data into the index. The inclusion of such data would permit investigation of structure–activity relationships. A glance at a given section would reveal how many compounds contained a given ring structure(s), functional group(s), or combinations thereof, as well as the type and level of activity exhibited. This could be a powerful aid to research. Any type of data could be included; the possibilities are unlimited. The data can be ordered by any desirable parameter with the linearized structures associated with it for comparison.

f. Updating the Index. Updating the index does not present a problem. Since the program is available, supplements, which will include compounds received after the major index was generated, can be prepared at suitable intervals. When the supplements become too numerous for easy searching, the tapes used for each index can be blended and used to create a new master index. The frequency of updating would depend upon the growth rate of the file.

SUMMARY

The preparation of a computer-produced index of permuted Wiswesser chemical line notations is described. The uses and limitations of this powerful and economical retrieval tool are discussed. The utility of such an index may be markedly increased by the inclusion of biological, source, cost, etc., data.

ACKNOWLEDGMENTS

The authors are indebted to Dr. D. H. Frear for permission to use the Second Edition of the "Pesticide Index" as a source for compound to be coded and used as examples (Figure 4); Mr. W. Nugent and Mr. W. Matthews, Diamond Alkali Company Data Center, for their efforts and excellent cooperation in preparing the IBM 1401 program; and the Edgewood Arsenal Data Processing Center for preparing and implementing the Univac program.

LITERATURE CITED

(1) Sorter, P. F., Granito, C. E., Gilmer, J. C., Gelberg, A., Metcalf, E. A., *J. Chem. Doc.*, **4**, 56 (1964).
(2) Granito, C. E., Gelberg, A., Schultz, J. E., Gibson, G. W., Metcalf, E. A., *ibid.*, **5**, 52 (1965).
(3) Gelberg, A., Nelson, W., Yee, G. S., Metcalf, E. A., *ibid.*, **2**, 7 (1962).
(4) Landee, F. A., Abstract of Papers, Division of Chemical Literature, 147th National Meeting of the American Chemical Society, April 6–9, 1964, Philadelphia, Pa., p. 3F.

Computer Searching of Chemical Patents*

P. T. O'LEARY,[a] J. M. CATTLEY,[b] J. E. MOORE,[a] and D. G. BANKS[a]
Gulf Oil Corporation and Gulf Research & Development Company, Pittsburgh, Pennsylvania

Received April 29, 1965

In the fall of 1963, Gulf Research & Development Company bought from Information for Industry, Inc., an index, on magnetic tape, of the more than 100,000 U. S. chemical patents which had issued during and since 1950. The index is of "uniterm" or "coordinate" type. The indexer selects from the text of the patent he is indexing those words he believes characterize the patent. Frequently used words accumulated in this way comprise the "vocabulary" of now more than 9000 terms, or "descriptors," used for the magnetic tape form of the index. Thus, keywords, rather than concepts, are indexed. The index is "inverted"; that is, under each descriptor are listed the accession numbers of patents partially characterized by that descriptor.

The index was available in dual dictionary form for clerical "coordination" and on magnetic tape for search on an IBM 1401 computer with 8k memory.

Better to fit our computer facilities, we adapted the index for search on our IBM 7094. The adaptation of the IFI Index involved re-inversion of the index, during which clerical errors were corrected. A retrieval program was developed for the 7094 to search this re-inverted, or serial, file. This program has greatly increased speed and flexibility in searching, which we feel is essential to our use of the index.

With the index as now set up, we make many searches per week at an average time of about 1.5 minutes per question on an IBM 7094 computer. Most of these are on chemical ideas submitted by our research people. These span many petroleum industry operations: refining, devel-

* Presented before the Division of Chemical Literature, 148th National Meeting of the American Chemical Society, Chicago, Ill., Sept. 1964.
[a] Gulf Oil Corp.
[b] Gulf Research & Development Co.

opment of new products, petrochemicals, and even aspects of petroleum production, such as drilling fluid formulations and reservoir flooding. The search, where successful, shows what features of the idea are new and allows a sounder evaluation of it. In contrast, formerly only a small percentage of the ideas submitted could be searched, and time and money were spent on some of the rest, which we ultimately found to have little or no novel content. The index has proved helpful in state-of-the-art searches, too. When necessary, we supplement the computer searching in the IFI Index by conventional searching in other pertinent indexes. We feel that in addition to the value of the searches themselves, the use of the index has given us helpful experience in the use of a computer to retrieve documents from a file of moderate size.

Patents to be retrieved must be specified in terms of descriptors from the vocabulary. Requests for information, therefore, must be translated into specifications in terms of descriptors. The computer program provides four types of specification. The first type (an "and" group) is a list of descriptors all of which must appear on (that is, all of which must have been assigned by the indexer to) each patent retrieved. The second type (an "or" group) is a list of descriptors at least one of which must appear on each patent retrieved. The other two types of specification provide for selecting patents on the basis of the absence of descriptors, rather than on their presence. Thus, the third type (a "not any" group) is a list of descriptors none of which may appear on any patent retrieved. The fourth (a "not all" group) is a list which rejects a patent only if every descriptor in the list appears on it. The first three of these types are illustrated in Figure 1, which shows how information on documents retrieved is preceded in the computer printout by a description (in terms of descriptors) of the type of document requested.

Any number (within limits to be listed below) of any of these types of specification may be grouped together to form a "question" or "request," although there is really no point in using more than one "and" group or more than one "not any" group in a question. The input to the computer is in the form of punched cards one card to head each group of descriptors and one card for each descriptor. In addition, there is a card to identify the question or request and specify the type of output and another card (an "end request" card) to instruct the computer that no more questions are forthcoming. The person framing the questions needs only to describe the cards on data input forms. Descriptors are identified and specified on these forms by number, each descriptor having been assigned a number by the IFI people.

```
                    UNITERM INDEX

                U.S. CHEMICAL PATENTS

            INFORMATION FOR INDUSTRY, INC.

REQUEST NO. 640085     NO. RESULTS =   54 FOR 1950-1962
                       LIMIT WAS   =  200 TAPE 0


 EACH DOCUMENT MUST HAVE ALL OF THE FOLLOWING ---
00310   ACETONITRILE


 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM ---
01845   ALKENE
10650   BUTENE
10660   BUTENE-1
26530   ETHYLENE
37300   ISOBUTYLENE
37480   ISOOLEFIN
48220   OLEFIN
57070   PROPYLENE
56894   PROPENE
23955   DODECENE
19520   DECENE
19523   1-DECENE
42490   2-METHYL BUTENE-1
42510   3-METHYL-1-BUTENE
32940   HEXENE
32580   HEPTENE
47910   OCTENE
47920   OCTENE-1
47925   1-OCTENE

 DELETE DOCUMENT IF ANY OF THE FOLLOWING OCCUR ---
27120   EXTRACTION
65040   SOLVENT EXTRACTION
54640   POLYMERIZATION
54650   POLYMERIZATION X CATALYST
```

Figure 1. Example of question showing three types of specification of descriptor presence or absence.

Questions are assembled into batches, which for economy should contain at least about 20 questions and may contain as many as 36. Each batch must conform to the following upper limits:

Questions (requests)	36
AND, OR, etc. groups	500
Descriptors	1000

To the person framing the questions, this poses no problem. If a batch submitted exceeds one or more of the limits, the computer is programmed to process only as many questions as will not exceed any of the limits. The remaining questions may be included in a subsequent batch.

In addition to calling for patents characterized by sets of descriptors, one may instruct the computer to list the descriptors assigned to a patent whose number is known. This is done by including in the data input to the computer a card containing the code PATNO followed by the numbers of the patents whose descriptor list is desired; the PATNO card which signals this type of request must be counted with the AND, OR, etc. cards in evaluating approach to the batch limits listed above.

Several types of output are available. The simplest is a list of patent and accession numbers. There is a one-to-one correspondence between accession numbers and patent numbers. Because, however, the accession numbers are consecutive whereas the patent numbers are not necessarily so (not all patents being on chemical inventions), it is more convenient to use the accession numbers with books, provided by Information for Industry, Inc., containing the principal claim of every patent they have indexed. A letter preceding the accession number in the printout corresponds to the year in which the patent issued, A designating 1950, B 1951, etc. Selection of this type of output—a list of patent and accession numbers—is effected by entering "2" in the 23rd column of the request card.

By placing a "1" in the 23rd column of the request card, one obtains a printout not only of patent and accession numbers of patents meeting the specifications of the request, but also lists of the descriptors assigned to each of these patents, as illustrated in Figure 2. Only occasionally do we choose to have these descriptor lists printed. They may serve as a basis for selecting patents for closer study. They also show the indexer's habits—what features of the patent he thought significant, what features he ignored—and if the patents retrieved have not satisfactorily disclosed the information sought, the lists of descriptors may provide a clue as to how, in a subsequent try, to get closer to what is wanted.

Finally, the patent numbers, accession numbers, and descriptor numbers of patents retrieved may be stored on a magnetic tape, which we call a "save tape." To do this one merely places "2" in the 26th column of the control card. The output for all questions so designated in the same batch is stored on the same save tape. This save tape, containing a small fraction of the parent index, may be searched in a fraction of the time and at a fraction of the cost of searching the parent index. Hence, by asking a question broadly and storing the patents retrieved on tape, subspecific searches can then be made very cheaply on the save tape. For example, one may store all hydrocracking patents on a save tape and then search only that tape for particular hydrocracking catalysts.

```
UNITERM INDEX     RESULTS FOR REQUEST WLK03A     PAGE 5


     A01140   2500146
02760   ALUMINA ALUMINUM OXIDE
02790   ALUMINA GEL
03690   AMMONIUM MOLYBDATE
05950   AROMATIZATION DEHYDROCYCLIZATION
07620   BAUXITE GIBBSITE
12690   CARRIER
12890   CATALYST-PRODUCTION
14590   CHROMIC OXIDE CHROMIUM SESQUIOXIDE
14690   CHROMIUM NITRATE
19900   DEHYDRATION DEHUMIDIFICATION DEHYDRATOR
19980   DEHYDROGENATION
20660   DESULFURIZATION
22220   DIETHYL KETONE
23170   DIOXIDE
27640   FERRIC OXIDE
30180   GAS
30320   GEL
32330   HEAT TREATMENT
32350   HEAVY METALS AND SALTS
34150   HYDROGEN
35840   IMPREGNATION IMPREGNATING IMPREGNATED
40710   MAGNESIUM OXIDE MAGNESIA
43970   MOLYBDATE MOLYBDIC ACID
44030   MOLYBDENUM OXIDE MOLYBDENA
44830   NAPHTHENE NAPHTHENIC
45610   NITRATE
48160   OIL-TREATMENT
49380   OXIDE
51450   PETROLEUM
58940   RECYCLE GAS
59230   REFORMING REFORMATE
63010   SILICA SILICON DIOXIDE
63380   SIMULTANEOUS
64610   SODIUM OXIDE
70920   TITANIUM DIOXIDE TITANIA
71140   TOLUENE
77190   ZIRCONIUM OXIDE ZIRCONIUM DIOXIDE
```

Figure 2. Example of output in answer to question. This type of output includes accession number, patent number, and list of descriptors.

There is no way of knowing beforehand how many patents will be retrieved in response to a request. To obviate the possibility of being overwhelmed by paper, a limit is placed, in columns 17–20, of the request identification card, on the number of documents to be printed out. The limit applies to both simple lists of accession and patent numbers and to lists accompanied by descriptors; it does not, however, limit the storage of retrieved patents on a save tape. The maximum limit of 9999 allows complete printout of the answer to all but the broadest questions, for those who wish a complete answer. It has proved useful to save patents retrieved on tape and to print out a small sample, say twenty, of patent numbers with their accompanying descriptors. As mentioned above, these descriptors then may serve as guides in framing subspecific questions to be searched in the save tape. Unfortunately, the sample will consist of the first twenty (or whatever other number is selected) patents which fulfill the requirements of the question. They are, therefore, not necessarily a representative sample but rather are more likely to indicate the habits of the indexers employed over the period during which the patents issued. The fact that the index now occupies three reels of tape allows some measure of selection of the period represented by the sample by specifying in what order the reels are to be interrogated by the computer.

The selection of descriptors in the framing of questions needs contrivance because of the many different names which have been used for chemical compounds, processes, concepts, etc. Because a familiarity with the vocabulary or list of descriptors is helpful here, we funnel all searches through one person, who translates requests for information into descriptor specifications. Forms for search requests have been developed (Figure 3) for those who wish to use them—forms which consist of questions designed to elicit from the requester a fairly clear definition of the information he wants. The form in Figure 3 was devised by Mr. D. J. Zeloyle of the Gulf Oil Corporation.

GR&DCo. COMPUTER PATENT SEARCH REQUEST No. _____

FROM: _____ DIVISION REQUESTED BY: _____
DATE: _____

TO: Dr. P. T. O'Leary APPROVED BY: _____
Patent Information Section, 2-204 DATE: _____

TITLE: (Also include I.R. number if an Idea Record exists.)

1. Disclose the proposal briefly. Report the utility of any new composition of matter.

2. List the reactant(s), product(s), by-product(s), and catalyst(s) or initiator(s). Determine the scope of these by stating the broad classification and the specific names (e.g., is the reactant any hydrocarbon, any aromatic hydrocarbon, any mononuclear aromatic hydrocarbon or benzene?).

   (a) Reactant(s)

   (b) Product(s)

   (c) By-Product(s)

   (d) Catalyst(s) or Initiator(s)

3. Record the reaction conditions.
   (a) Time (b) Temperature (c) Pressure (d) Other

4. Explain, if necessary, the chemistry involved by means of structural formulas equations, etc. on the back of this sheet.

5. Designate all pertinent patents and other pertinent publications known to you.

DATE RECEIVED _____ APPROVED _____

Figure 3. Search request form for chemicals or chemical reactions.

With keyword indexing, one concept does not necessarily correspond to a unique descriptor. Indeed, synonymous, or nearly synonymous, descriptors occur in the list. In framing requests, it is important not to overlook any synonyms. Several aids are available to prevent this. First of all, we keep, in packets ready for use, descriptor cards corresponding to particular, frequently-used concepts. "Catalysts" is a good example. Its packet contains over 200 cards. A subset of these is used corresponding to the particular catalyst of interest. For example, a Friedel–Crafts catalyst subpacket would include, among others, cards for Friedel–Crafts catalysts, ferric chloride, aluminum chloride, aluminum bromide, and boron trifluoride.

Second, our Computer Section has tabulated the frequency of posting to descriptors, year by year. A page

PAGE 104

DESCRIPTOR AND TEXT		FREQUENCY BY YEAR												
		50	51	52	53	54	55	56	57	58	59	60	61	62
21500	DICALITE	1	1	0	0	0	0	0	0	0	0	0	2	2
21510	DICARBAMATE	0	2	1	0	0	0	0	0	0	0	6	10	5
21514	DICARBOXYLATE	0	0	0	0	0	0	9	0	0	0	0	0	0
21517	CICARBOXYLIC ACID	0	0	0	0	0	100	108	0	0	0	0	0	0
21518	DICARBOXYLIC ACID ANHYDRIDE	0	0	0	0	0	0	0	0	12	0	0	0	0
21518	DICARBOXYLIC ACID DICARBOXYLATE	80	101	89	125	120	0	0	127	204	271	186	180	259
	SPECIFIC ACIDS													
21530	DICARBOXYLIC ANHYDRIDE	6	7	4	0	0	0	0	0	0	0	14	16	22
21540	DICHLORIDE	50	63	46	18	50	32	30	37	17	34	31	60	8
21550	DICHLOROACETIC ACID	5	6	7	0	11	0	0	0	0	0	10	2	7
21560	DICHLOROACETYLATION	0	0	0	0	0	1	0	0	0	0	0	2	0
21565	AR-DICHLOROANILINE	0	0	0	0	0	0	0	0	14	0	0	0	0
21570	DICHLOROBENZENE	6	8	2	0	0	12	12	27	0	0	10	7	26
21580	1,2-DICHLOROBENZENE	16	24	16	0	0	0	25	0	0	0	30	26	26
21590	1,3-DICHLOROBENZENE	1	1	1	0	0	0	0	0	0	0	8	2	1
21600	1,4-DICHLOROBENZENE	3	8	3	0	0	0	0	0	0	0	11	6	8
21604	O-DICHLOROBENZENE	0	0	0	14	33	9	0	15	46	40	0	0	0
21605	P-DICHLOROBENZENE	0	0	0	4	0	0	0	0	14	17	0	0	0
21610	O,O-DICHLOROBENZOYL PEROXIDE	0	5	15	0	0	0	0	0	0	0	3	4	11
21620	1,4-DICHLOROBUTANE	1	2	0	0	0	0	0	0	0	11	3	4	5
21630	3,4-DICHLOROBUTENE-1	4	6	3	0	0	0	0	0	0	0	6	6	6
21640	DICHLORODIETHYL SILANE	6	13	0	7	0	0	0	0	0	0	0	2	2
21650	DICHLORODIFLUOROETHYLENE	0	5	4	0	0	0	0	11	0	15	4	6	3
21660	DICHLORODIFLUCROMETHANE	3	10	14	12	13	0	20	13	20	27	26	22	29
21670	DICHLORODIMETHYL SILANE	11	26	24	20	13	0	0	0	0	30	17	24	18
21680	DICHLOROETHANE	72	69	63	63	0	0	97	103	99	99	102	93	116
21683	1,1-DICHLOROETHANE, 1,2-DICHLOROETHANE	0	0	0	0	50	0	0	0	0	0	0	0	0
21685	DICHLOROETHYL ETHER	0	0	0	12	0	0	0	0	0	0	0	0	0

Figure 4. Page from tabulation of frequency of posting to descriptors, year by year.

from this tabulation is shown in Figure 4. A quick check of posting frequency to descriptors in a proposed question may give a clue as to the need for seeking a synonym or for approaching the desired concept by a different combination of descriptors. For example, it is clear that indexers during 1955 and 1956 employed descriptor number 21517—in other years, 21520—to designate dicarboxylic acids. Hence, we would include both descriptors in an "or" group. In the 1964 edition of the IFI Index tape, and correspondingly in the second, revised edition of the vocabulary, such synonymous pairs of descriptors do not occur. Each pair has been replaced by a single descriptor, and patents indexed by members of the pairs have been re-indexed accordingly. Thus, for example, in the current edition of the index, patents formerly posted under descriptor number 21517 or number 21520 now are posted under descriptor number 21520, and, in the second edition of the vocabulary, descriptor number 21517 does not appear.

The "21530-dicarboxylic anhydride" situation is less clear. It seems unlikely that such anhydrides were not of concern in any patent issuing during 1953–1959. More likely, that chemical type was indexed under 21517, 21518, 21520, or 04220-anhydride, or under combinations of these.

Thirdly, if we know the number of a patent or patents pertinent to the subject of a proposed search, we can, by the use of a PATNO request (as described above) discover how it was indexed. We may then assume that other pertinent patents have been indexed similarly, an assumption not invariably valid but, none-the-less, a useful expedient. We can, of course, and should, complement the descriptors revealed by the PATNO output by whatever synonyms come to mind.

Another problem to cope with in framing questions is control of noise, that is, limiting retrieval of nonrelevant documents. Two approaches are available: the first, to frame several questions of increasing narrowness in attempting to retrieve information sought; the second, to frame a broad question, store the retrieved patent numbers on a save tape, and subsequently search the tape with more specific questions. In some instances, several questions, none subspecific to another, may be framed to retrieve the desired information.

We might illustrate the first approach by our search for information on the fungicidal action of hydroxyaromatic acids extracted from lichens. The first question called for patents bearing the descriptor, "fungicide–fungistatic." So indexed had been 2461 patents issuing between 1950 and 1962, inclusive. We stored these on a save tape. We did not, of course, review this noisy output. We had framed narrower questions, as illustrated on Figure 5. This was an attempt to find plant extracts, regardless of chemical composition, with fungicidal activity. The 26 retrieved was not an unreasonable number to review. An approach *via* chemical composition is shown in Figure 6. Here 41 patents were retrieved. Another question, shown in Figure 7, was designed to search by a generical descriptor for hydroxyaromatic acids (with fungicidal action). This question could have been combined with the previous one, but we were, at the time these questions were framed, acquainting ourselves with the index and wished to compare indexing under generic and specific descriptors. The patent, 2,802,029, retrieved in the last question covered a salicylamide and had also been retrieved by the previous question. Notice that in the last question the aromatic character of the acid was ignored. If we had included a descriptor or descriptors for the concept of aromaticity, we would not only have reduced noise but also lost our signal. This is, of course, in keeping with keyword indexing and emphasizes the value, in using an index of this type, of submitting several questions of increasingly narrower scope, and selecting for review the

```
                UNITERM  INDEX

              U.S. CHEMICAL PATENTS

            INFORMATION FOR INDUSTRY, INC.

        REQUEST NO.PT029A   NO.RESULTS =  306 FOR 1950-1962
                            LIMIT WAS  =  500 TAPE 1

 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
54440   POLYETHYLENE
54890   POLYPROPYLENE POLYPROPENE
54290   POLYBUTENE POLYBUTYLENE
54690   POLYMETHYLENE
54374   POLYDIOLEFIN
54280   POLYBUTADIENE
54590   POLYISOBUTYLENE POLYISOBUTENE
54610   POLYISOPRENE


 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
71970   TRIALKYL ALUMINUM
72430   TRIETHYL ALUMINUM
22040   DIETHYL ALUMINUM CHLORIDE
20980   DIALKYL ALUMINUM HALIDE
01950   ALKYL ALUMINUM
01960   ALKYL ALUMINUM HALIDE
02950   ALUMINUM HYDRIDE
76870   ZIEGLER CATALYST
```

Figure 5. Question used to search "fungicide–fungistatic" save tape for patents on plant extracts.

```
                UNITERM  INDEX

              U.S. CHEMICAL PATENTS

            INFORMATION FOR INDUSTRY, INC.

         REQUEST NO.WEH 03   NO.RESULTS =   41 FOR 1950-1962

 EACH DOCUMENT MUST HAVE ALL OF THE FOLLOWING ---
29760   FUNGICIDE FUNGISTATIC


 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
30030   GALLIC ACID GALLATE
30390   GENTISATE
34560   HYDROXYBENZOIC ACID HYDROXYBENZOATE
59800   RESORCYLIC ACID
60790   SALICYLIC ACID SALICYLATE
68670   TANNIC ACID TANNIN
73530   TYROSINE
74440   VANILLIN AND DERIVATIVES
22500   CI-2-HYDROXYETHYL TEREPHTHALATE
```

Figure 6. Question used to search "fungicide–fungistatic" save tape for patents on specific hydroxyaromatic acids.

```
                UNITERM  INDEX

              U.S. CHEMICAL PATENTS

            INFORMATION FOR INDUSTRY, INC.

         REQUEST NO.WEH 05   NO.RESULTS =    1 FOR 1950-1962

 EACH DOCUMENT MUST HAVE ALL OF THE FOLLOWING ---
29760   FUNGICIDE FUNGISTATIC


 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
34587   HYDROXYCARBOXYLIC ACID
34588   HYDROXYCARBOXYLIC ACID, HYDROXYCARBOXYLATE
34593   HYDROXYESTER
```

Figure 7. Question used to search "fungicide–fungistatic" save tape for bearing generic descriptors for hydroxy acids.

```
                UNITERM  INDEX

              U.S. CHEMICAL PATENTS

            INFORMATION FOR INDUSTRY, INC.

         REQUEST NO.WEH 02   NO.RESULTS =   26 FOR 1950-1962

 EACH DOCUMENT MUST HAVE ALL OF THE FOLLOWING ---
29760   FUNGICIDE FUNGISTATIC

 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
53510   PLANTS BIOLOGY
74610   VEGETABLE
32620   HERB
28620   FLOWERS
38990   LEAF
60410   ROOT
29750   FUNGI FUNGUS
01610   ALGICIDE ALGAE


 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
27110   EXTRACT
27120   EXTRACTION
65040   SOLVENT EXTRACTION
```

Figure 8. Question used to make save tape on polymers made by Zeigler catalysis.

```
                UNITERM INDEX
              U.S. CHEMICAL PATENTS
            INFORMATION FOR INDUSTRY, INC.
         REQUEST NO.PT029A NO.RESULTS =  27 FOR SAVE TAPE
                           LIMIT WAS = 500 TAPE 0
 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
14700   CHROMIUM OXIDES
15420   COBALT OXIDE
27640   FERRIC OXIDE
27800   FERROUS OXIDE
37160   IRON OXIDE
41130   MANGANESE OXIDE
44030   MOLYBDENUM OXIDE MOLYBDENA
45500   NICKEL OXIDE
44060   MOLYBDENUM TRIOXIDE
15940   COLUMBIUM OXIDE
68710   TANTALUM OXIDE
70460   THORIUM OXIDE
70970   TITANIUM OXIDE
73270   TUNGSTEN OXIDE
74390   VANADIUM OXIDE
77190   ZIRCONIUM OXIDE ZIRCONIUM DIOXIDE

 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
54440   POLYETHYLENE
54890   POLYPROPYLENE POLYPROPENE
54290   POLYBUTENE POLYBUTYLENE
54690   POLYMETHYLENE
54374   POLYDIOLEFIN
54280   POLYBUTADIENE
54590   POLYISOBUTYLENE POLYISOBUTENE
54610   POLYISOPRENE

 EACH DOCUMENT MUST HAVE AT LEAST ONE TERM FROM -
71970   TRIALKYL ALUMINUM
72430   TRIETHYL ALUMINUM
22040   DIETHYL ALUMINUM CHLORIDE
20980   DIALKYL ALUMINUM HALIDE
01950   ALKYL ALUMINUM
01960   ALKYL ALUMINUM HALIDE
02950   ALUMINUM HYDRIDE
76870   ZIEGLER CATALYST
```

Figure 9. Question used to search "Ziegler" save tape for catalysts containing transition-metal oxides.

answer which appears to provide the best balance between reduction of noise and loss of significant information.

The use of a save tape is illustrated in Figures 8 and 9. The patents on polymers made by Ziegler catalysis amounted to 306 and were stored on tape (Figure 8). Subsequently a question was asked to determine in how many of these transition-metal oxides figured in the catalysis. Of the 27 retrieved, a review showed that 9 were of interest.

When, as in this instance, related questions are saved on the same tape, it is advisable to repeat the listing of the original descriptors in searching the save tape. When other questions are not related or when only one question is saved, such relisting is unnecessary.

We have been curious to compare the performance of the computer-searched IFI Index with that of other indexes we have been using and continue to use. In one instance, patent searchers in Washington found, by using the U. S. Patent Office classification, four patents, three of which issued since 1950 having also been retrieved by use of the IFI Index. By an IFI search 105 patents were retrieved. From this number 18 of possible interest were selected by checking principal claims, which serve as abstracts. These 18 patents included the three post-1950 patents retrieved by search in the U. S. Patent Office. A search in other indexes available at Gulf Research & Development Company's Research Center had furnished a list of patents whose 1950+ portion had all been retrieved by the IFI search.

We have mentioned briefly above supplementing computer searching of the IFI Index with hand searching of other indexes. For example, in an exhaustive search, we list the Patent Office classes to which the patents retrieved in the IFI search have been assigned. This list may then serve as a guide to the searcher using the collection of classified patents in Washington. The inverse process, recalling the IFI descriptor list of a known pertinent patent, has already been mentioned.

A problem still to be solved is how best to reduce the number of patents printed out by the computer to the text which is to be read by the person or persons requesting the information. We assume here that the list has been pared to a reasonably low number by reframing questions. Some people who request searches are willing to work with the raw computer output themselves. In other instances, a member of the Patent Information staff does this task. In general, a person can screen 30–60 patents per hour; thus, this remains the expensive phase of searching.

Whoever screens the output may advantageously begin by referring to the IFI books of representative claims, mentioned above and illustrated in Figure 4. We keep these books in an air-conditioned room with convenient work space. We feel that, except in state-of-the-art searches, it is reasonable to limit the number of patents to be screened to 100. When the output exceeds that number, the question should be rephrased to obtain a less noisy output.

Copies of many of those patents whose full text the searcher wishes to read are on file in the same room. *All* patents from 1959 on in the IFI Index are available on microfilm, and we have a reader-printer in the search room so that these microfilmed copies can be conveniently read and, if desired, printed. Thus, we have tried to facilitate the use of the computer-search output by phrasing questions to obtain a reasonable signal-to-noise ratio and by putting all the equipment for screening the output into one place.

Installation and Operation of a Registry for Chemical Compounds*

DON P. LEITER, Jr., HARRY L. MORGAN, and ROBERT E. STOBAUGH
Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210

Received July 7, 1965

Since 1958 the Chemical Abstracts Service has been working toward establishing a computer-based system for handling chemical information.** Briefly, the concept of the CAS system consists of sets of special subject files in the following categories: (1) physical properties, (2) chemical reactivities, (3) biochemical activities, and (4) applications. With the importance of compounds in correlation studies, and the need to interrelate compounds and the huge collections of chemical and other data, a highly developed subsystem, called the Registry System, for handling compounds must be the first step in the actual operation of an over-all computer-based service. The Registry System will include files of compounds interconnected with files of associated data that permit identifying the compounds and retrieving them from the files.

The process of registration or entry into the system includes the assignment of a unique machine address, called a Registry Number, to each compound which is new to the system or the retrieval of the previously assigned Registry Number from the files. It is intended that ultimately the system will include a record of every chemical substance reported in the literature and iden-

* Presented before the Division of Chemical Literature, Symposium on Current Applications of Mechanized Procedures for Handling Chemical Information, 149th National Meeting of the American Chemical Society, Detroit, Mich., April 1965.

** CAS is pleased to acknowledge the support of the National Science Foundation during all of this work and of the National Institutes of Health during part of the work.