(4) J. L. Carmon, Final Report to the National Science Foundation on Grant GN-851 to Expand the University of Georgia Information Center, Oct 1, 1969–Sept 30, 1972, p 1.
(5) Reference 4, p 5, A-5.
(6) Reference 4, p A, 9–10.
(7) M. E. Williams et al., "Four Year Summary, Educational and Commercial Utilization of a Chemical Information Center, June 25, 1968 to June 25, 1972", NSF Contract No. NSF-C554, IITRI Report No. C6156-18, July 30, 1972, p 4.
(8) C. M. Bowman and M. T. Brown, "The Development, Cost and Impact of a Current Awareness Service in an Industrial Organization", *J. Chem. Doc.*, **11**, 72 (1971).
(9) B. Lawrence, B. H. Weil, and M. H. Graham, "Making On-Line Search Available in an Industrial Environment", *J. Am. Soc. Inf. Sci.*, **25** (6), 365–66 (1974).
(10) S. A. Elman, "Cost Comparison of Manual and On-Line Computerized Literature Searching", *Spec. Libr.*, **66**, 12–18 (1975).
(11) M. L. Calkins, "On-Line Services and Operational Costs", *Spec. Libr.*, **67**, 559–567 (1976).
(12) J. H. Bement, "The New Prices—Some Comparisons", On-Line, Vol. 1, No. 2, April 1977, pp 9–22.
(13) D. T. Hawkins, "Impact of On-Line Systems on a Literature Searching Service", *Spec. Libr.*, **67**, 559–567 (1976).
(14) "On-Line Searching Continues Strong, but True Costs Prove Elusive", Vol. 4, No. 6, Advanced Technology Libraries, 1975, pp 4–5.
(15) A. D. Collette and J. A. Price, "A Cost/Benefit Evaluation of On-Line Interactive Bibliographic Searching in a Research and Engineering Organization", presented at the ASIS Mid-Year Meeting, May 19–21, 1977.
(16) C. E. Wilmot, "On-Line Opportunity: A Comparison of Activities in America and the United Kingdom", *ASLIB Proc.*, 134–43 (1976).
(17) Chemical Abstracts Service "Specifications Manual for Computer-Readable Files in the Standard Distribution Format", Vol. 2, 1975, Data Element Section, ID No. 0059 01-0A.
(18) See ref 17, ID No. 0071 01-0A.
(19) Chemical Abstracts Service, "Subject Coverage and Arrangement of Abstracts Sections in Chemical Abstracts 1975 Edition".
(20) Chemical Abstracts Service, "History of CA Section Names and Numbers Beginning with the Eighth Collective Period", 1974.
(21) Lockheed Information Systems User Education Materials, June 25, 1974.
(22) On-line searching is the use of a cathode ray tube (CRT) or teletype-like terminal to directly access computerized bibliographic files at locations distant from the searcher.
(23) It is the intention of this paper to make the reader aware of the various forms of keywords which should be considered prior to searching; the truncation technique is omitted deliberately to keep the focus on keyword selection.
(24) SDC text-searching option allows a serial search of an alphabetic or alphanumeric character string once a subset of the database has been created by the searcher through a direct search.
(25) A strategy arranged in algebraic notation and computer processed from the innermost to the outermost level, i.e., (A + B * (C + D * (E + F))).

On-Line Searching of the American Petroleum Institute's Databases†

LEONORE ROGALSKI
Corporate Research Center, UOP Inc., Des Plaines, Illinois 60016

Received July 27, 1977

The databases produced by the Central Abstracting and Indexing Service of the American Petroleum Institute consist of about 212 000 literature citations and 98 000 U.S. and foreign patents from January 1964 through December 1976. On-line searching of these files utilizes many special features to improve selectivity and relevance of the retrieved citations. These features include: (1) a thesaurus for controlled vocabulary, (2) assignment of roles, links and section headings; and (3) weighting of terms by indicating whether they are of major or minor importance. Experiences during two years of on-line searching of these files are discussed.

The Central Abstracting and Indexing Service (CAIS) of the American Petroleum Institute produces three databases which can be searched on-line through the System Development Corporation's Orbit system. The three bases cover literature (APILIT), patents (APIPAT), and business data (P/E NEWS) for the petroleum and petrochemical industries.

As of January, 1977, APILIT comprises about 212 000 citations from selected technical and trade journals of the petroleum and chemical fields. This on-line file is updated monthly with the addition of about 1500 items.

The APIPAT covers approximately 98 000 U.S. and foreign patents with updating adding about 700 patents per month. Both APILIT and APIPAT are currently available only to subscribers.

The third file, P/E NEWS, was loaded in 1975 and is open to both subscribers and nonsubscribers but at different prices. P/E NEWS contains about 46 000 items taken from the following publications: *Platts Oilgram News Service*, *Middle East Economics Survey*, *Petroleum Intelligence Weekly*, *Petroleum Economist*, and *Oil Daily*. Coverage is business oriented with indexing of company activities and developments in the petroleum and energy fields. Updating adds about 500 items per week.

APILIT and APIPAT have been available since January 1964, for batch searching. Beginning with an experimental operation for the year 1974, the entire file then went on-line. The history of the transfer of the CAIS indexing to computer-searchable files was presented by Irving Zarember of the CAIS staff at a conference in 1975.[1]

Because of the 13 years of data available and the speed with which such material is entered, UOP has found the API bases extremely helpful for both retrospective search and the location of current information on new processes or announcements. The staff of the UOP Research Library has conducted over 400 searches of these files with questions varying from simple ones, such as finding papers by a certain author, to more comprehensive requests, such as locating selected data on hydrocracking from a certain period to date. Searches are

Table I. Abbreviations, Searchable Fields, and Print Options for the API Databases

		APILIT		APIPAT		P/E NEWS	
		SEARCH	PRINT	SEARCH	PRINT	SEARCH	PRINT
AN	ID Number	X	X	X	X	X	X
TI	Title	(1)	X	(1)	X	X	
AU	Authors	X	X	X	X		
AA	Author Affiliation	X (2)	X				
ED	Entry Date	Date ranging		Date ranging		Date ranging	
SO	Source		X		X		X
PN	Patent Number	–	–	X	(3)		
CC	Section Headings	X	X	X	X (3)		
IT	Index Terms	X	X	X	X		
ST	Free Terms	(1)	X	(1)	X		
LT	Linked Terms	(1)	X	(1)	X		
UP	Update Code	X		X		X	
DE	Descriptors	–	–	–	–	X	X
JC	Journal Code	–	–	–	–	X	

(1) TI and ST are searchable by forming a subset then searching the subset with the STRSEARCH command. The LT field can be searched with the SENSEARCH command after producing a subset.

(2) Searchable 1972 through October 1974 through the source field. From October 1974 to date AA appears as a separate field in the unit record and is, therefore, searchable as such.

(3) The patent number is printed as the source.

† Presented at the 11th Middle Atlantic Regional Meeting, American Chemical Society, Newark, Del., April 20–23, 1977.

0095-2338/78/1618-0009$01.00/0

Table II. Orbit Abbreviations

CONT?	Continue?
D	Down
GEN	General
MM	Multimeaning
N	No
NBR	Neighbor
NP	No Postings
OVFLW	Overflow
PROG	Progress
P or PRT	Print
QUAL	Qualify
SCHD	Searched
SS 1/C	Search Statement 1 or Command?
SS 1/PSTG	Search Statement 1 - Number of Postings
SENS	Sentence Search
STRS	String Search
Y	Yes

conducted via Telenet or Tymshare on a Texas Instruments terminal of the Silent 700 series, Model 733KSR300BAUD.

INTRODUCTION

Before discussing the various special features of the API files and how they are used, a definition of the terminology is in order. Table I shows the abbreviations, searchable fields, and print options of the API databases.

The Boolean operators for Orbit are AND, OR, and NOT. Since AND takes precedence over OR and parentheses are not permitted, complex logic statements cannot be used in searching.

Table II shows additional Orbit abbreviations that will be used throughout this paper. Standard Orbit instructions apply, such as a colon for unlimited truncation and a hash mark (#) for one-letter truncation.

Several special features contribute to the ease of searchability and to the usefulness of the API files. The statements made here apply to APILIT and APIPAT, unless P/E NEWS is specifically mentioned. The reason is that the indexing for P/E NEWS does not follow all of the rules set forth for the other two files. One of the most important of these features is the controlled vocabulary. The API Thesaurus provides a display of all acceptable indexing terms. The vocabulary is hierarchical in nature with autoposting to "broader" or generic terms from the specific ones. Although controlled, the Thesaurus is not considered stagnant. It has been growing as new terms have been needed and has been modified and enriched with Scope Notes that are intended as aids to both the indexer and searcher. An Indexer's Manual prepared by the CAIS provides the ground rules for the system.

Searchers should be familiar with both the Indexer's Manual and the Thesaurus. In preparing search strategy, the searcher must follow through on the "See Also" index terms and other directions such as "Scope Notes" especially if a comprehensive search is desired. For this reason searches of API files at UOP are conducted by Research Library personnel who have been trained to use the system, rather than by the person requesting the information. The requester is encouraged to be present when the search is run particularly when overlapping concepts or large numbers of postings are anticipated, or if the questions are somewhat vague.

Figure 1 shows a section from one of the pages in the API Thesaurus. Acceptable index terms appear in bold print, e.g., ANAEROBIC MICROORGANISM, ANALOG COMPUTER, and -ANALYSES AND TESTS. The dash in front of the latter indicates that it is a section heading. The section headings are directly searchable. The numbers following the terms are aids to the API staff for their computerized text editing and indexing system. As usual any term such as -ANALYSES AND TESTS must have the AND portion disguised as A#D for searching to avoid confusion with the AND operator.

Cross reference terms direct the user to "Use: PENTENES" as shown under AMYLENES or to "See: ANALOG COMPUTER, COMPUTER SIMULATION, DATA CORRELATION, DIMENSIONAL ANALYSIS, and MODEL" as shown under ANALOG. In both cases the searcher should check the entries under the cross references. There might be additional aids or related concepts that might not have occurred to the searcher when first planning strategy. "Broader terms" (BT) which are autoposted are shown as are the "See also" (SA) terms. SA's are related concepts that again may guide the user to a better or related term to search for other than the one he thought of to express the desired concept. Scope Notes appear under ANAEROBIC MICROORGANISM and -ANALYSES AND TESTS.

AMYLENES
use: PENTENES

ANAEROBIC MICROORGANISM *(1652)*
Index only when specified in the abstract. After 1972 they are not recorded as identifiers.
BT: MICROORGANISM
SA: BACTERIA

ANALOG
see: ANALOG COMPUTER
COMPUTER SIMULATION
DATA CORRELATION
DIMENSIONAL ANALYSIS
MODEL

ANALOG COMPUTER *(1653)*
BT: COMPUTER
SA: COMPUTER SIMULATION
HYBRID COMPUTER

ANALOG INSTRUMENT
VALID IN 64-65 ONLY.

-ANALYSES AND TESTS *(1654)*
This is an index term for a section heading. Each abstract in or cross referenced to this section in the API abstract bulletins is indexed with this heading. Searchable on disk only, 1964-1970. From 1971 on searchable in all API indexes. For the scope of this term, see the list of section headings near the end of this book.

Figure 1.

MOTOR GASOLINE *(3828)*
Does not include aviation gasoline.
BT: MOTOR FUEL
NT: LEADED GASOLINE
LOW LEAD GASOLINE
UNLEADED GASOLINE
SA: AVIATION GASOLINE
GASOLINE STOCK
NAPHTHA
OCTANE NUMBER
UF: OUTBOARD MOTOR GASOLINE
plus OUTBOARD MOTOR
UF: PETROL
PREMIUM GASOLINE
plus PREMIUM
UF: REGULAR GASOLINE
plus STANDARD QUALITY
UF: SUBREGULAR GASOLINE
plus SUBSTANDARD QUALITY
UF: SUPERPREMIUM GASOLINE
plus SUPERPREMIUM

Figure 2.

Figure 2 shows the Thesaurus entry under MOTOR GASOLINE. The display shows its BT: MOTOR FUEL and all narrower terms (NT) that autopost MOTOR GASOLINE and MOTOR FUEL, i.e., LEADED GASOLINE, LOW LEAD GASOLINE, and UNLEADED GASOLINE. Also shown are the SA terms and "Used for" (UF) relationships. Thus, UF: OUTBOARD MOTOR GASOLINE plus OUTBOARD MOTOR means that both terms MOTOR GASOLINE and OUTBOARD MOTOR are necessary to index the concept of OUTBOARD MOTOR GASOLINE which is not an acceptable term. This cross reference assures consistency in indexing and helps the searcher retrieve a specific concept. Even though one or more of the individual terms called for in the cross reference may be highly posted, the coordination of terms narrows retrieval to the most relevant answers.

In addition to the acceptable index terms, the indexer assigns Free Terms (FT) which are taken from the natural language of the abstracts from which the index is prepared. The API staff keeps track of the number of times a free term is encountered. After five or more times it becomes a candidate for the Thesaurus. The form, whether an index term or a cross reference, and its place in the hierarchy, if it is inserted as an index term, is carefully considered to assure maximum acceptance and consistency in indexing. Free terms are searchable from 1970 for the APILIT file and from 1972 for the APIPAT file.

UOP SEARCHES

One of our people required a paper on the Selexol process published by someone at Mobil. Figure 3 shows the search.

```
SS 1 /C?
USER:
SELEXOL GAS TREATING AND ALL MOBIL:

PROG:
SS 1 PSTG (1)

SS 2 /C?
USER:
"P 1"

PROG:

AN  - 2382471
TI  - SELEXOL PROCESS. SELECTIVE TREATMENT OF
      SOUR NATURAL GAS
AU  - MOBIL OIL AG; UNLAND H; STEIN W H; MICHAELI
      W
AA  - EXPLOR PROD MOBIL OIL A G CELLE GER
SO  - ERDOEL ERDGAS Z 1975 V91 N.10 341-7 CHEM
      ABSTR V84-108121 IN GERMAN
CC  - NATURAL GAS, NAT. GASOL., LPG
CC  - TREATING
```

Figure 3.

```
SS 1 /C?
USER:
DIMERIZATION

PROG:
SS 1 PSTG (1247)

SS 2 /C?
USER:
SENS : CYCLODIMER: OR :CYCLODIMER: (ST)

PROG:
(57) SCHD (0) QUAL; CONT? (Y/N)

USER:
Y

     .....

PROG:
(1232) SCHD (28) QUAL; CONT? (Y/N)

USER:
Y

PROG:
SS 2 PSTG (28)
```

Figure 4.

SELEXOL GAS TREATING and ALL MOBIL: were the terms searched. Only one reference turned up but it was the right one. This type of search could not have been run on most other systems since author affiliations are not indexed consistently. API includes the company name even in instances where the name is in the text of the report and not in the bibliographic data.

Another experience along these lines was a request for all publications by personnel from the All Union Scientific Research Institute for Petrochemical Processes (USSR). What made this search different was the fact that the Institute's name began with "All" thereby confusing the computer which recognizes ALL as part of the search logic used to avoid a multimeaning message. By using the RENAME command, ALL was changed to ANY. The affiliation field could then be searched in the normal way with the computer now accepting the ALL and searching the proper alphabetical sequence. Incidentally, this same search required that the institute's name be abbreviated to ALL UNION SCI RES INST PETROCHEM: so as not to exceed the allowable number of characters for any one field. The correct abbreviation and point of truncation was easily ascertained from the appropriate printed affiliation and company indexes. This can also be determined on-line by utilizing the neighbor command.

One of our searchers required specific articles on cyclodimerization (Figure 4). This term is not an index term, but the acceptable term, DIMERIZATION, produced 1247 postings. The second step requested a SENSEARCH for :CYCLODIMER: (with left and right truncation) from both the title field and the free term field with the OR operator. This strategy narrowed the hits to 28.

Figure 5 shows the use of a chemical aspect in a search. The request was for data on the preparation of organic silicones. A *Chemical Abstracts* search was impractical because one would have to think of the names of all of the compounds of this type as well as the generic terms such as organosiloxane, organosilane, and organosilicone. The API search required only the term SILICON ORGANIC/P. This chemical aspect coordinated with the Role P for "product" produced a satisfactory search.

```
SS 1 /C?
USER:
SILICON ORGANIC/P

PROG:
S 1 PSTG (181)
```

Figure 5.

The use of roles is another feature of the system that comes in handy for limiting search response to relevant hits rather than producing noise. Roles are assigned to indicate whether a chemical is an agent or reagent, A, or a product, P, in a reaction. An intermediate in a reaction receives both roles. "Prior treatment" and "Substance analyzed for" were formerly designated by roles but now are linked to the index terms PRIOR TREATMENT and SUBSTANCE DETERMINED so that they can still be searched for as required.

This brings one to the use of linking in the system. All chemical compounds that are acceptable terms are assigned a link so that all of its autoposted terms will be linked. All Chemical Aspects used to index a single compound are assigned the same link. All index terms in the Common Attributes facet of the hierarchy are linked to the terms which they describe. Exceptions are those listed under Industry, Science/Technology, and Market/Affiliation. These terms should always be linked to at least one other term in some other facet of the hierarchy as indicated in the Scope Notes that appear with them in the Thesaurus. For example, concurrent flow of liquids in a system would be indexed as CONCURRENT linked to FLUID FLOW. Structure and the terms listed under this facet of property in the hierarchy are linked to the terms for materials having the structure. Materials are linked to each other, to chemical aspects, and to chemical index terms when used to index a single substance.

To illustrate one of the above, if a single organic silicone compound, such as TETRAETHYLSILANE, were the desired product in the search shown in Figure 5, a SENSEARCH would have been performed in the link term field for :C8:SATURATED CHAIN:SILICON ORGANIC: SINGLE STRUCTURE TYPE: in the subset formed by first searching some or all of these chemical aspects with the AND operator.

A search for Selected Studies on Hydrocracking for 1972 through 1976 utilized several features of the system (see Figure 6). Date-ranging limited the results to the proper years. The asterisk before HYDROCRACKING is a weighting factor and assures that only articles that discuss hydrocracking as the main topic are retrieved.

Sometimes when conducting a search for a new process the exact term requested is in the Thesaurus, but none or only a few postings appear when it is searched. This is caused by the delay between the introduction of new terminology in the published literature and its acceptance in a suitable form in the Thesaurus. The solution is to use the broader concepts or technology on which the process is based followed by a search for the companies known to be involved and/or searching for the requested term in the free term field. Figure 7 illustrates a case in point. HYDRANE COAL GASIFICATION was requested but the "No postings" message was returned. The searcher then asked for COAL GASIFICATION before 1977, since that was the date the index term HYDRANE COAL GASIFICATION became acceptable. A STRING SEARCH for HYDRANE with both left and right truncation was conducted on the title and free term fields. Seventeen answers were found.

```
SS 1 /C?
USER:
HYDROCRACKING AND FROM 72 THRU 76

PROG:
SS 1 PSTG (1041)

SS 2 /C?
USER:
1 AND THERMODYNAMIC PROPERTY OR 1 AND
THERMODYNAMIC ACTIVITY OR 1 AND

PROG:
CNT 2

USER:
KINETICS OR 1 AND REVIEW OR 1 AND MODEL
OR 1 AND COMPUTER SIMULATION OR

PROG:
CNT 2

USER:
1 AND REACTION MECHANISM

PROG:
SS 2 PSTG (469)

SS 3 /C?
USER:
2 AND *HYDROCRACKING

PROG:
SS 3 PSTG (229)
```

Figure 6.

```
SS 1 /C?
USER:
HYDRANE COAL GASIFICATION

PROG:
NP (HYDRANE COAL GASIFICATION)

SS 1 /C?
USER:
COAL GASIFICATION AND FROM 64 THRU 76

PROG:
SS 1 PSTG (2063)

SS 2 /C?
USER:
STRS :HYDRANE: OR :HYDRANE:(ST)

PROG:
(61) SCHD (0) QUAL; CONT? (Y/N)

USER:
Y

PROG:
(2012) SCHD (17) QUAL; CONT? (Y/N)

USER:
Y

PROG:
SS 2 PSTG (17)
```

Figure 7.

The use of Chemical Aspects in searches was mentioned, but the use of this type of searching for the generic or Markush claims in patents must be emphasized. The searching may become a little complex, as one may have to specify each carbon count of a desired range as well as UNKNOWN CARBON COUNT with the appropriate reactive constituents, e.g., MONOHYDROXY for monoalcohols or HYDROCARBON for carbon-hydrogen compounds, but these searches do work.

One thing not covered so far is the selection of print options. The PRINT command produces the identification number (AN), title (TI), authors and affiliation (AA), source reference (SO), and section headings (CC). PRINT TRIAL produces AN, TI, CC, and index term (IT) for the first two records found. In on-line searching of a problem question, the display of the IT and CC fields can lead to clues for weighting; the PRINT FULL command provides a display of all categories of the unit record. The addition of EXCLUDE to the PRINT FULL command can be used to eliminate one or more categories as desired. The PRINT command can also be tailored to give any combination of categories. Fast off-line printing is available. Turnaround time has been good.

UOP's experience with the P/E NEWS has been somewhat limited. This is understandable because the Research Library users generally request technical data. However, it has been used occasionally. Sample queries included locating announcements on specific companies, air pollution legislation

```
SS 1 /C?
USER:
ALL DOW CHEMICAL: AND ETHYLENE AND PIPELINE AND
ALL ALBERTA:

PROG:
SS 1 PSTG (6)

SS 2 /C?
USER:
"P 6"

PROG:

-1-
AN  - 88017046
TI  - EDMONTON-TO-SARNIA PRODUCTS PIPELINE TO
      BE FINISHED BY 1977
SO  - THE OIL DAILY N.6213 P6 (8/20/76)
```

Figure 8.

concerning the petroleum industry, Dow's ethylene pipeline in Alberta, and statistics on leaded, unleaded, and premium gasolines. Figure 8 illustrates the Dow ethylene pipeline search. Note that the strategy called for ALL DOW CHEMICAL: AND ETHYLENE AND PIPELINE AND ALL ALBERTA:. The P/E NEWS does not use a controlled vocabulary, hence the need for truncation for terms such as DOW CHEMICAL and ALBERTA. Seven hits were produced.

CONCLUSIONS

The searches displayed here may seem somewhat simplistic. It would be easy to choose others that are much more complex requiring many OR and AND operators. However, they would not go any further toward explaining the system's characteristics than the ones used and would have been much harder to follow.

Moreover, in these files the searcher rarely, if ever, needs to use as many alternative terms as are needed in a natural language system, and has access to such searching techniques as linking, roles, chemical aspects, and weighting that are not always available in other systems, and certainly not all in one system. Degrees of specificity are possible. For example, searches can be made for Anti-Wear Additives in Gear Lubricants, for All Additives in Gear Lubricants, or for All Lubricant Additives.

The times for the searches have varied from a few minutes of computer time to more than an hour. One does not always succeed on the first try. The Hydrocracking search took two preliminary searches with broader concepts before the requestor was convinced that a delineation of the particular aspects of the subjects that were of interest was required for the search to produce reasonable results.

In many of the searches shown, the purpose of using the various special features of the system besides assuring pertinency was to curtail the number of hits to a "reasonable" number—reasonable depending upon whether the need for the information was immediate (necessitating on-line printing), for background information, and/or for a truly comprehensive search. Generally, System Development Corporation has been prompt with their return of off-line prints. As much on-line printing as possible is done, for nothing impresses the client more than having the specific article or a list of pertinent references put into his hands with a turnaround time of an hour or less. However, as said earlier, the use of the 300 BAUD is slow as modern computer printers go. Searching and printing with 1200 BAUD terminals has been made available. Reportedly the 1200 BAUD terminals have been enthusiastically received by searchers for they give them the ability to supply requestors with immediate answers even when a large number of hits are encountered.

LITERATURE CITED

(1) I. Zarember, private communication; paper presented before the 66th Conference of the Special Library Association, June 8–12, 1975, Chicago, Ill.