as a consultant to Smith Kline & French Co. (now Smithkline Corp.) and became one of my closest friends; the first issue of *Index Chemicus* was dedicated to his memory. I was always encouraged by Madeline Berry, Hannah Friedenstein, Aaron Addelston, Al Gelberg, Bill Longenecker, and other Division members too numerous to mention. I was going to mention more names but as I reviewed some old correspondence, I realized how fallible my memory is. For example, if I were to name one member of the CNA I would have to name a dozen or more. But certainly Bill Wiswesser and Al Smith have played a key role in the development and use of WLN by ISI. So did Howard Bonnett.

As many of you know, the *Index Chemicus* was started with the support of approximately twelve drug companies. Joe Clark of Lederle, Bill Sullivan of Hoffmann-La Roche, and Alex Moore of Parke-Davis were especially helpful to me. Others who helped IC were Walt Southern of Abbott Labs, Howard Nutting of Dow, George McCarthy of Geigy, Charles Rice at Lilly, Evelyn Armstrong and Bob Harte at Merck, Rita Goodemote at Schering, Max Gordon at Smithkline, Doug Remsen at Squibb, Fred Bassett of Upjohn, Eliot Steinberg and Lee Starker at Warner-Lambert, and Ernie Hyde of Imperial Chemical Industries.

My own co-workers at ISI, including Gaby Revesz, Bonnie Lawlor, and Charlie Granito, have made it possible for some of these ideas to persevere in the face of tremendous odds. Not the least of my friends have come from the ranks of CA. I will not embarrass those who still work there by naming them. But for the man who has everything, it is perhaps most gratifying of all to have respected competitors as friends.

REFERENCES AND NOTES

(1) R. K. Merton, "The Matthew Effect in Science: the Reward and Communications Systems of Science Are Considered", *Science*, **199**, 55 (1968).
(2) E. Garfield, "Citation Analyses, Mechanical Translation of Chemical Nomenclature, and the Macrostructure of Science", *J. Chem. Inf. Comput. Sci.*, **15**, 153 (1975).
(3) E. Garfield, "Proposal for a New Profession: Scientific Reviewer", *Curr. Contents*, No. 14, 5–8 (4 April 1977).
(4) H. Brooks, "U.S. Science and Technology: a Prescription for "Health", *Science*, **195** 536 (1977).
(5) E. Garfield, "Putting Our Money Where Our Needs Are", *Bull. Am. Soc. Inf. Sci.*, **1**, 10, 32 (1974).
(6) E. Garfield, "Significant Journals of Science", *Nature (London)*, **264**, 609 (1976).
(7) A. Mazella, "Bibliometric Study of the Review Literature", Final Report to the National Science Foundation on Grant No. DSI-76-0534. Philadelphia: Institute for Scientific Information, 1977.
(8) A. M. Woodard "Review Literature: Characteristics, Sources and Output in 1972", *ASLIB Proc.* **26**, 367 (1974).
(9) J. Cole and S. Cole, "The Ortega Hypothesis", *Science*, **178**, 368 (1972).
(10) S. T. Turner and D. Chubin, "Another Appraisal of Ortega, the Coles, and Science Policy: the Ecclesiastes Hypothesis", *Social Sci. Inf.*, **15**, 657 (1976).
(11) E. Garfield "Introducing Citation Classics: the Human Side of Scientific Reports", *Curr. Contents*, No. 1, 5–6 (3 Jan 1977).
(12) E. Garfield, "Who Are the Information Scientists?" *Curr. Contents*, No. 5, 32 (7 Aug 1962); reprinted in "Essays of an Information Scientist, Vol. I, ISI Press, Philadelphia, 1977 p 2.
(13) J. Martyn "Unintentional Duplication of Research", *New Sci.*, **21**, 388 (1964).
(14) H. Skolnik, "Milestones in Chemical Information Science", *J. Chem. Inf. Comput. Sci.*, **16**, 187 (1976).
(15) E. Garfield "Let's Erect a New Tower of Babel", *Current Contents*, No. 45, 5–7 (6 Nov. 1974); reprinted in "Essay of an Information Scientist", Vol. 11, ISI Press, Philadelphia, Pa., 1977, pp 172–174.
(16) E. Garfield, "La Science Francaise est-elle si provinciale?" (Is French Science Too Provincial?), *Recherche*, **7**, 757 (1976).
(17) E. Garfield, "Le noveau defi American I", *Curr. Contents*, No. 15, 5–11 (11 April 1977).
(18) M. Debre, "La langue Francaise et la Science universelle" (The French Language and Universal Science), *Recherche*, **7**, 956 (1976).
(19) E. Garfield, "Highly Cited Articles. 33. Articles from Italian Journal and from Italian Laboratories", *Curr. Contents*, No. 6, 5–12 (7 Feb 1977).
(20) E. Garfield, "Can Italian Biomedical Journals Survive?", presented at the First Conference of the Biomedical Press in Italy, Feb 16, 1976, Rome.

On-Line Searching—Specialist Required[†]

DAVID M. KRENTZ

Information Systems Department, E. I. du Pont de Nemours & Company, Inc., Wilmington, Delaware 19898

Received September 22, 1977

On-line searching has expanded into a complex specialty. Efficient use of computer-searchable files depends increasingly on the expertise of the information specialist whose skill in interviewing the client reflects a full understanding of the technical limitations of computer search systems and available databases. For the search, the specialist supplies experience in such areas as differences between on-line files and their hard-copy equivalents, the various indexing patterns of abstracting services, and the application of Boolean logic. The value of this expertise is illustrated here in terms of problems found in searching *Chemical Abstracts Condensates* relative to vocabulary, chemical nomenclature, search logic, and the search system used.

On-line bibliographic searching is alive and well. Today, on-line searching is a multi-million dollar per year business and is growing. It is estimated that 365 000 conversational terminals will be in operation by the end of 1977, as compared with 243 000 only three years ago.[1] Last year over 1 200 000 on-line searches were run. This number is expected to grow by 10 to 15% by the end of 1977.[2]

During the development and growth of on-line searching, the information specialist has played a key role by testing new systems and convincing the end user of their usefulness and value. But, in today's advanced systems, what part should and do the information specialists continue to play in on-line searching? How important is their skill in using the terminal to assure speed, reliability, and completeness at a minimum cost? To help determine these answers, we must review briefly how bibliographic data banks, the computer programs, and the interactive terminal systems to search them have developed.

BATCH COMPUTER SEARCHING

By the late 1960s, refinements in computerized bibliographic information retrieval techniques opened the door to economical searching of many major data banks: *Chemical Abstracts*, ERIC, *Nuclear Science Abstracts*, NASA, *Biological Abstracts*, and others. During this period, governmental financial aid was supporting computerized information development groups at such organizations as Massachusetts Institute of

†Presented at the 11th Middle Atlantic Regional Meeting of the American Chemical Society, Newark, Del., April 21, 1977.

0095-2338/78/1618-0004$01.00/0

Technology, the University of Georgia, the University of Pittsburgh, and the Illinois Institute of Technology Research Institute,[3,4,7] in their efforts to develop computer programs capable of batch searching commercially available databases. Once the costs became acceptable, user demand for computerized literature searching and its benefits increased dramatically.[3]

Literature updating programs, generally described as Current Awareness or Selective Dissemination of Information (SDI), allowed a scientist or researcher to define his specific "profile of interest" using a series of keywords placed in a Boolean logic expression or strategy. The programs would then deliver weekly, biweekly, or monthly—as the case may be—lists of computer-printed citations, making it possible for the client to keep up to date at a cost many were willing to pay. Users were quick to credit searches for a share of their increased productivity. Almost all pointed to substantial time savings in literature searching.[6,8] Many said information located through these searches was valuable to their work.[5]

Up to this point, searchers had the flexibility of changing on-going SDI search strategies to fit user needs more closely. As SDI tapes grew and information centers accumulated them into retrospective search tools, the luxury of search strategy modification grew costly. This put a sharp focus on the need to design retrospective search strategies correctly the first time. Searchers became acutely aware that reruns on these large files added several hundred dollars to the cost and added several weeks to the completion schedule of a search. Consequently, retrospective searchers had to enhance their profiling expertise.

ON-LINE SEARCHING ADVANTAGES

It was not long before the files and search costs grew beyond the economical limits of even expert profiling. By the early 70s, a combination of factors (increasing costs; demand for fast, cost-acceptable retrospective searching; more timely answers; and the existence and maturation of the "profiling specialist") created demand for new on-line searching[22] systems and techniques as developed by the National Library of Medicine (NLM), the System Development Corporation (SDC), and others.[3]

On-line searching has several advantages over batch searching. First, search response time is fast, 10–15 min for an answer vs. several weeks required for batch searching. Second, it is less expensive, \$50–\$75 or less, vs. \$100–\$300 or more.[9–16] Third, search strategies may be tested, modified, and reviewed in a matter of seconds in a variety of databases, to give the best possible retrieval from each base. These advantages stimulated many to change from batch to on-line searching.

Specialist Needed. Did the on-line capability make searching specialists obsolete? During the development period, demonstrations at technical meetings (and there were lots of them) may have led the casual observer to believe that almost "anyone" could tap the keys on a terminal and extract the information he wanted. Such "casual observers" soon learned that it hardly ever works that way. Today, about five years or so later, refinements, including newly developed on-line systems, the rapid expansion of reliable on-line phone networks, and the conversion of formerly batch searchable bibliographic files, have opened a new era of fast information retrieval. Now, more than ever, the information specialist's knowledge and skill in profiling and search techniques make the function at the interface between the client and the databases the key to efficient cost-effective searching.

On-Line Searching Pitfalls. What are some of the pitfalls searchers may find? Here are examples based on working experience with the SDC and Lockheed on-line systems using

```
PREPN              ]       [ANISIDINE
  +                ]       [  +
PREPARATION        ]       [AMINOANISOLE
  +                ]   *   [  +
PREPG              ]       [1-AMINO-4-METHOXYBENZENE
  +                ]       [  +
PRODN              ]       [METHOXYANILINE
  +                ]       [  +
SYNTHESIS          ]       [C7H9NO
  +                ]       [  +
MANUFACTURE        ]       [105-13-5
  +                ]
MANUF              ]
  +                ]
MANUFG             ]        ( + = OR; * = AND)
  +                ]
MANUFD             ]
```

Figure 1. Terms for the CAS anisidine search.[23]

```
SS 1/C?
USER:
ANISIDINE

PROG:
SS 1 PSTG (126)

SS 2/C?
USER:
AMINOANISOLE

PROG:
SS 2 PSTG (5)

SS 3/C?
USER:
AMINO AND METHOXYBENZENE

PROG:
SS 3 PSTG (1)

SS 4/C?
USER:
METHOXYANILINE

PROG:
SS 4 PSTG (20)
```

Figure 2. SDC postings for CAC anisidine search.

the *Chemical Abstracts Condensates* (CAC) bibliographic database.

There are three basic parts of any on-line system. The bibliographic data bank, the search system, and the hardware, i.e., terminal and printer. There are problems peculiar to each of them.

BIBLIOGRAPHIC DATA BANK

The computer-searchable counterpart of the weekly issues of *Chemical Abstracts* (CA) is *Chemical Abstracts Condensates*. CAC is equivalent in all respects to its hard-copy counterpart, except that it does not contain abstracts.

Searching CAC via computer is accomplished by formulating a search strategy composed of keywords describing the subject matter desired[23] (Figure 1). In this example, the searcher wanted all information on the preparation or manufacture of anisidine.

The keywords are entered, one at a time, and combined using the appropriate Boolean logic operators AND, OR, or AND NOT (described below). The computer then responds, telling the searcher the number of citations in the file which meet the requirements of each keyword (Figure 2).

However, such numbers can be misleading. Why? Because the indexing input often reflects only the words actually seen in the *title* and the *abstract* of the article. The whole article, in many cases, is not indexed, nor is there any enrichment of the index by added synonyms or terms.

Indexing. In many cases, keywords entered in CAC are taken from abstracts rather than the full article. When

```
SS 1/C?
USER:
PREPN

PROG:
SS 1 PSTG (14234)

SS 2/C?
USER:
PREPARATION

PROG:
SS 2 PSTG (19621)

SS 3/C?
USER:

2 AND NOT 1
                         17067 DOCUMENTS
PROG:                    NOT POSTED TO
SS 3 PSTG (17067)        PREPN

SS 4/C?
USER:

1 AND NOT 2
                         11630 DOCUMENTS NOT
PROG:                    POSTED TO PREPARATION
SS 4 PSTG (11630)

SS 5/C?
USER:

1 OR 2
                         TOTAL
                         UNIQUE DOCUMENTS
PROG:                    POSTED TO BOTH PREPN
SS 5 PSTG (31301)        AND PREPARATION
```

Figure 3. Postings for the abbreviated and fully spelled search terms for preparation, illustrating little document overlap in indexing between the two terms.

Chemical Abstracts Service (CAS) prints the cumulative Subject Volume Indexes, expanded chemical names, i.e., preferred nomenclature, comprehensive coverage of trade names and synonyms, and terms from in-depth indexing, are available for searching. CAC tapes, however, are never updated with the expanded indexing terms included in the indexes. SDC and Lockheed are both presently considering a merger of CAC and CASIA (Chemical Abstracts Subject Index Alert) tapes to produce a merged file.

Abbreviations. CAS uses many abbreviations in indexing. These abbreviations appear in the keyword field of CAC. However, Lockheed and SDC systems search both the document title as well as the keyword fields of the CAC file. A complete search, therefore, requires entering all applicable full words since they may appear in the titles, and all proper abbreviations thereof, as shown in the keyword field.

To illustrate: PREPARATION and the abbreviation PREPN are both search terms. Figure 3 shows posting levels for the two terms. If a searcher were to use only PREPARATION, he would retrieve 19 621 documents posted to that term (CAC 1/72-1/77). If the abbreviation PREPN were added, he would pick up 11 680 more documents (SS5 minus SS2). It is noteworthy that even though there is a difference of only 2833 (SS3 minus SS1) in number of documents between the postings of the two terms, there are actually 11 680 (SS5 minus SS2) more unique documents retrieved when both terms are used. This indicates there is little overlap between fully spelled out and abbreviated search terms. Figure 4 lists other preparation-type indexing terms for the relatively simple anisidine search. The wide spread between the number of postings shown for each word and the number for its abbreviation illustrates how important it is for the searcher to know the database pattern.

The indexing input pattern for *Chemical Abstracts Condensates* presents many more challenges to the specialist who wants to get the correct retrieval on the first run. These include (1) entering abbreviations and nomenclature without punctuation, as the computerized file will not accept punctuation and will respond with a negative answer when it is used; (2) using prefixed abbreviations, such as electroprepn and photoredn, as they are missing from the published list of accepted terms; (3) using chemical prefixes, such as alpha, para, etc., and numerical substitution position notation, but only as they appear in titles, and they must be entered without punctuation; (4) avoiding inverted chemical names, such as Benzene, dichloro, which are not searchable in that form and must be searched as individual keywords or by proximity or String Search[24] techniques; (5) avoiding empirical formulas, molecular formulas, and chemical registry numbers, as they cannot be searched in CAC, but can be extracted from CASIA.

ABBREVIATION		FULL WORD	
PREPG	3	PREPARING	1633
PRODN	19695	PRODUCTION	25115
PREPN	12312	PREPARATION	19756
PREPD	19	PREPARED	2120
MANUF	13344	MANUFACTURE	5256
MANUFG	316	MANUFACTURING	1737
MANUFD	57	MANUFACTURED	251

Figure 4. Comparison of postings: abbreviations vs. full words for preparation, manufacture, and production search terms (CAC Jan 1972–Jan 1977).

```
SDC - ORBIT

   1. DOW/CS AND CHEMICAL/CS

LOCKHEED - DIALOG

   1. DOW/CS

   2. CHEMICAL/CS

   3. PA=DOW CHEMICAL

   4. (1 * 2) + 3
```

Figure 5. SDC vs. Lockheed corporate author search.

Most of the factors discussed up to this point have dealt with chemical nomenclature. There are other built-in problems which can affect search results in an important way. Corporate authors and Chemical Abstracts Service section numbers are two.

Corporate Authors. While it is possible to do a corporate author search in the CAC file, there is an important limitation to remember. The corporate author, in the nonpatent literature, is cited as such by CAS only if no personal authors or editors are identified.[17] In the patent literature, the inventor and patent assignee are both listed in the author field and are searchable,[18] but they are searchable only if the database vendor chooses to make those fields searchable by their search systems (i.e., ORBIT and DIALOG) (Figure 5).

In SDC-ORBIT, a corporate author search is conducted by searching the corporate author term(s) with the limiter "/CS". This is because SDC has combined the three fields containing corporate author information (corporate author name, work address, and patent assignee) into the same searchable field.

In Lockheed DIALOG, the keywords plus the patent assignee field must be seached, e.g., PA=term and S term. Since Lockheed has not combined the patent and literature citations into one searchable field, the preferred search method is to expand the PA=field, select the appropriate E numbers, and OR these postings with those resulting from the keyword SELECT search statements.

Search/Subsection Searching. CAS categorizes its subject coverage into 80 different sections, each of which is further divided into subsections.[19] Section and subsection *numbers* are searchable, and, in some cases, the numbers should be used as search terms if a comprehensive or restrictive search of the database is required. This is because of the way Chemical

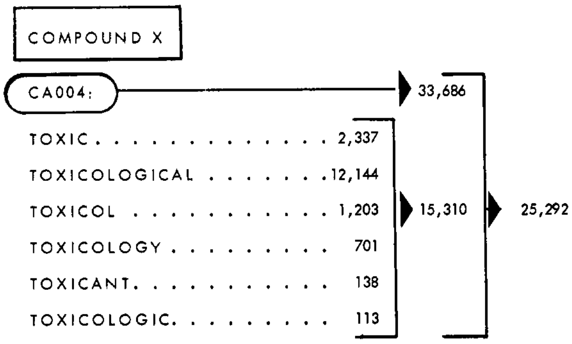

Figure 6. Postings: CAC section number vs. keywords.

Abstracts Service assigns keywords to articles.

The weekly issues of *Chemical Abstracts* (and CAC) announce the most recent advances in chemistry as soon as possible. Comprehensive indexing is not performed for each article abstracted in these issues but is done later when the more complete Subject Volume Indexes are issued.

CAS classifies each article in one or more of the 80 major sections which best describe the main thrust of the article; then a series of keywords is assigned to identify secondary material in the article which is not comprehensive enough to warrant classification by section. These are the keywords found in the pink index pages of the weekly issues and the keyword field of CAC. As a general rule, keywords are not assigned to describe the section to which an article is assigned as it is assumed that searchers would cover that section in the course of their manual search.

To illustrate, if a document deals with the *toxicity* or *toxicology* of *compound X*, all documents will not be retrieved using these terms. To retrieve all documents, the searcher has two options: (1) search for all postings to compound X, and screen the results, or (2) add Section CA004 titled "Toxicology" to a selection of toxicity terms and search on that combination of terms. Figure 6 shows term postings for toxicological terms as well as for Section CA004. Using the computer to combine the toxicity terms, 15 310 unique documents are found compared to 33 686 found in Section CA004. Combining the section number and toxicity term yields 25 292 unique documents, or 9982 more than found using only toxicological terms.

Additional advantages afforded by section searching include elimination of unwanted material through section number negation, or search specificity by limiting a search to one or more section categories.

However, caution should be exercised when searching by section numbers. The specialist should know that beginning with *Chemical Abstracts*, Volume No. 81 (July–Dec 1974), CAS reorganized CA sections' contents to consolidate information on energy. This consolidation created a new Section 51, composed of all the previous contents of Sections 51 and 52, as well as *part* of Section 53.[20,21] In addition, a new Section 52 was created from parts of six other sections: 16, 48, 71, 73, 75, and 77.[20,21] This break in the series pattern makes it necessary to conduct dual section searches of CAC: one for January 1972 through June 1974, and a second for July 1974 to date, to ensure complete coverage on searches involving energy. Additionally, searches of the 1970–1971 file require the use of another section listing since some section numbers are different from current numbers (Figure 7).

SYSTEM PROBLEMS

There are other problems for users of on-line information systems specific to logical operators used in computer searching of a bibliographic database. Assistance from a specialist who understands the use of the Boolean logical operators AND, OR, and AND NOT, facilitates getting the correct answer from the system. These operators are used to create sets of keywords which will expand or limit a given search within the parameters of the question at hand. The OR operator is the expander; it adds documents by telling the computer to select all documents containing the keyword(s) specified. The AND operator is the search limiter; it imposes the demand that all documents found must have at least one keyword from each of the OR sets previously created.

1970 - 1971		1972 - 1977	
001	History, Education and Documentation	001	Pharmacodynamics
002	General Biochemistry	002	Hormone Pharmacology
003	Enzymes	003	Biochemical Interactions
004	Hormones	004	Toxicology
005	Radiation Biochemistry	005	Agrochemicals
006	Biochemical Methods	006	General Biochemistry
007	Plant Biochemistry	007	Enzymes
008	Microbial Biochemistry	008	Radiation Biochemistry
009	Nonmammalian Biochemistry	009	Biochemical Methods
010	Animal Nutrition	010	Microbial Biochemistry
011	Mammalian Biochemistry	011	Plant Biochemistry
012	Mammalian Pathological Biochemistry	012	Nonmammalian Biochemistry
013	Immunochemistry	013	Mammalian Biochemistry
014	Toxicology	014	Mammalian Pathological Biochemistry
015	Pharmacodynamics	015	Immunochemistry
016	Fermentations	016	Fermentations
017	Foods	017	Foods
018	Plant-Growth Regulators	018	Animal Nutrition
019	Pesticides	019	Fertilizers, Soils, and Plant Nutrition
020	Fertilizers, Soils, and Plant Nutrition	020	History, Education, and Documentation

Figure 7. CA numbering for sections, 1970–71 vs. 1972–77.

SET I		SET II		SET III	
A1	TOXICOLOGY	B1	MERCURY	C1	ALGAE
A2	TOXICITY	B2	HEAVY METALS	C2	FISH
A3	TOXIC	B3	METHYL MERCURY	C3	CRUSTACEA
A4	HAZARD	B4	MERCURIC	C4	AQUATIC
A5	HAZARDS	B5	AMALGAM	C5	CA060
A6	POLLUTANT			C6	OCEAN
A7	POLLUTANTS			C7	STREAM
A8	CA004*			C8	STREAMS
A9	WASTE			C9	POND
A10	WASTES			C10	PONDS
A11	DISCHARGE			C11	LAKE
A12	DISCHARGES			C12	LAKES
A13	EFFLUENT			C13	RIVER
A14	EFFLUENTS			C14	RIVERS
				C15	CA061
				C16	FLORA

(*CA0014 IN 1970-71 FILE)

Figure 8. Three parameter search: terms for (I) toxic, toxicity, (II) mercury, (III) aquatic environment.

To illustrate, for the search "All information on the toxicology of mercury wastes on fish, and its effects on the aquatic environment", the strategy must be formulated to have three sets of OR terms—one for toxicity, one for mercury, and one for the aquatic environment (Figure 8).

This strategy illustrates several important points. First, all names, synonyms, plurals, etc., for each word in each parameter must be searched to ensure thorough coverage. Second, CA section numbers must be used to retrieve all relevant documents not obtained by CA keywords, and must reflect differences in the CAC 1970–71 and 72-to-date files. Third, since this is a three-parameter search specific for (I) toxic and (II) mercury or mercuric type compounds in (III) water or aquatic environments, the strategy is not designed to give the searcher assured coverage of all mercury toxicity. If such coverage is needed, the searcher must drop the third parameter and broaden the search to all mercuric wastes, toxicology, or hazards.

System Variations. The selected Boolean algebraic concepts are then used to query the on-line system. Significant differences exist between user manuals for SDC and Lockheed systems with regard to commands and the manner in which the computer accepts and acts on query parameter statements. To illustrate, refer to the terms in Figure 8 by their al-

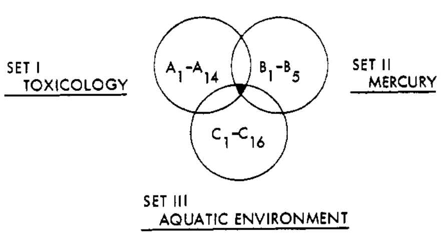

Figure 9. Three-parameter search: Lockheed DIALOG search strategy.

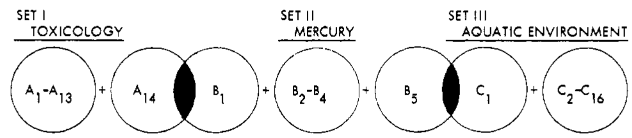

Figure 10. Three-parameter search: SDC-ORBIT nested search strategy.

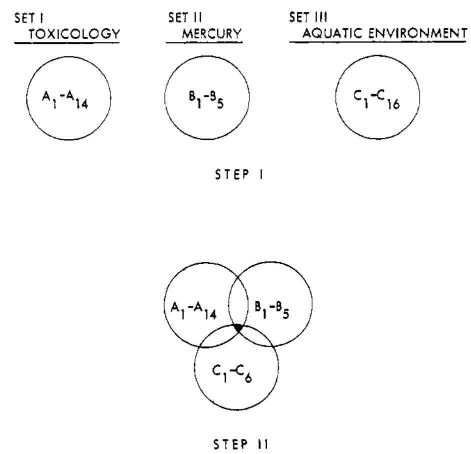

Figure 11. Three-parameter search: SDC-ORBIT, two-step search strategy.

phanumeric designations: Set 1 terms, A_1 through A_{14}; Set 2 terms, B_1 through B_5; and Set 3 terms, C_1 through C_{16}. The algebraic equation for this three-parameter nested[25] search is (+ = OR; * = AND):

$$(A_1 + A_2 + A_3 + A_4 + \cdots + A_{14}) * (B_1 + B_2 + B_3 + \cdots + B_5) * (C_1 + C_2 + C_3 + \cdots + C_{16})$$

Depending on the computer system being searched, this may be handled in one of two ways. The first, represented by the Lockheed system, DIALOG, will process and collect all "OR'd" terms within a set first, then combine the "OR'd" sets using the AND operator (Figure 9). This results in a one-step search retrieving each document containing at least one keyword from each set.

The second, represented by the SDC system, ORBIT, will process the AND operators first, and then follow by processing the "OR'd" term collection. This is illustrated by the following two algebraic equations.

Step I: entered by the searcher

$$(A_1 + A_2 + A_3 + \cdots + A_{14}) * (B_1 + B_2 + B_3 + \cdots + B_5) * (C_1 + C_2 + C_3 + \cdots + C_{16})$$

Step II: answer

$$(A_1 + A_2 + A_3 + \cdots + A_{13}) + (A_{14} * B_1) + (B_2 + B_3 + B_4) + (B_5 * C_1) + (C_2 + C_3 + \cdots + C_{16})$$

The final answer gives all documents posted to the single terms A_1 through A_{13}, B_2 through B_4, and C_2 through C_{16}, plus all documents containing terms from the Boolean "AND'ed" ($A_{14} * B_1$) and ($B_5 * C_1$) terms (Figure 10).

This illustrates how the SDC system turns the three-set search into a combination of one- and two-parameter searches, giving an answer of largely false drops. The false drops are eliminated by creating four different search statements—one for each set of terms, and a fourth combining them with the Boolean AND operator (Figure 11).

HARDWARE PROBLEMS

The third trouble spot for on-line searchers is the computer terminal. There are many possibilities of error at the input stage of searching, but some are more prevalent than others. These include breakdown, system failure, incorrect terminal settings, computer failure, incorrect log-in ID codes, passwords or user name, inclusion or exclusion of special characters, i.e., punctuation or spaces, noise, transmission errors, or telephone network problems. Many of these errors are easily corrected by checking and reentering the input, relogging in, delaying the search until system problems are resolved, or rechecking the terminal settings. However, the fact remains that such problems waste technical time and consequently reduce productivity and efficiency, thereby increasing the total cost of a search.

These problems can be minimized if a specialist, familiar with the search system conventions and terminal, is on hand and can quickly recognize and diagnose them when they occur.

SUMMARY

The information specialist is a key factor in the successful operation of on-line searching services. At this point in searching technology, an individual who acts as a skillful interface between the client and the data banks makes the most effective use of both the client's time and money and the data banks' resources.

This discussion centered on the technique that a specialist has learned to use in searching one data bank (CAC) with two systems (SDC, Lockheed). In a given year a specialist may be called on to search up to 50 databases and to use at least as many systems. Each one multiplies the instances in which skill and experience can keep costs down and false drop to a minimum.

Demand for on-line searching is growing. It can be a cost acceptable way to obtain technical information from major bibliographic data bank files. Efficiency requires, however, that the searcher (1) know the data bank conventions, (2) understand profile techniques and the application of Boolean logic, and (3) is thoroughly familiar with the equipment. To minimize costs, the searcher must evaluate and compensate to the extent possible for problems as they arise. Substantial dividends in time saved, cost reduction, and improved results will accrue to the client when aided by a trained and experienced information specialist.

ACKNOWLEDGMENT

The author wishes to acknowledge the technical assistance provided by H. D. Marvel, Jr., during the preparation of this paper.

REFERENCES AND NOTES

(1) International Data Corporation, "Statistical Reference Book-Data Entry/Communication Equipment", July 1976.
(2) M. E. Williams, "Education and Training for On-Line Use of Data Bases", EUSIDIC Conference, Graz, Austria, Dec 1, 1976, p 5.
(3) D. B. McCarn and J. Leiter, "On-Line Services in Medicine and Beyond", *Science*, **181**, 318–324 (July 27, 1973).

(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]

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.

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

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

