and testing capabilities of the CAIN citation data base. The service was made available during the test period to university graduate students, staff, and other user groups, including government agencies, associations, and agri-business.

During the test period, 137 searches were made returning a total of 13,404 citations in a total of 2,762 minutes of connect time. A median search took 21 command steps using 20 descriptors and retrieved 54 citations in 18.47 minutes. The median cost, exclusive of terminal rental and personnel was $14.14. These costs were divided as follows: computer and terminal time, 53.3%; off-line printing at 5¢ per citation, 31.3%; telephone, 15.4%. User reaction solicited through questionnaires indicated that 42.3% of the citations retrieved were relevant to research needs, 30.5% were of peripheral value, and 27.3% were not useful.

The most impressive factor to users in all studies is the time saved by computer searching as opposed to manual searching. It is equally apparent that CAIN must be supplemented by alternative services in order to secure comprehensive coverage of agricultural literature. Services covering agricultural literature within other scientific disciplines are especially important.

LITERATURE CITED

(1) Campey, L. H., "User Reactions to CAIN" (cataloging and indexing data base of the "Bibliography of Agriculture"), Luxenbourg, Commission of the European Communities, 1974, 51 pp.
(2) Jestes, E. C., "A Comparison of SDI-Current-Awareness and On-Line Retrospective Bibliographic Search Services from the CAIN Data Base," Davis, University of California Library, 1975, 13 pp.
(3) Oyler, D., "CAIN On-Line at the University of Wisconsin, in "Agricultural Libraries Information Notes," Vol 1, No. 1, Jan. 1975, National Agricultural Library, Beltsville, Md.

Planning for Effective Use of On-Line Systems†

JAY S. BUCKLEY, JR.

Central Research, Pfizer Inc., Groton, Connecticut 06340

Received May 12, 1975

A perspective of developments in the field of computerized information retrieval is presented, and policies are suggested for the efficient use of on-line systems. Examples of MEDLINE, CHEMCON, and TOXLINE searches are given with cost data and compared, where possible, with manual searches. The lowest costs were obtained with on-line searching by information scientists. MEDLINE is superior in quality to manual searching of *Index Medicus*, but on-line searching of *Chemical Abstracts* keywords is not an effective substitute for manual searching of *Chemical Abstracts* Substance Indexes.

The rapid advances in the field of computerized information retrieval reflect the fast pace of developments in computer hardware and software. A perspective of these developments must serve as a basis for any discussion of planning for the use of on-line systems. I would like to offer a perspective in terms of the three decades; the 1960's, the 1970's, and the 1980's. The 1960's can be viewed as the period of batch systems. Using tapes to sequentially store the main body of information and storing search programs in relatively small computer cores, very effective searching systems were developed. Examples are Medlars, Excerpta Medica, and ASCA. These batch systems are still very useful today particularly for information prior to 1970, patent data, SDI's, and substructure searches.

Toward the end of the 1960's, great progress was made toward increasing the capacity of computers to store data both in core and in accessory disk packs. This progress paved the way for on-line systems of the 1970's, and the present decade can be viewed as the period of rapid introduction and development of on-line systems. We are only halfway through this decade; and these systems are encountered in our daily life at banks, airline terminals, stockbrokages, etc. In our field, large data bases such as *Chemical Abstracts* and MEDLINE first became accessible on-line in 1972. Today, the Lockheed Information Systems has a choice of 16 data bases, and the System Development Corporation offers 13.

Advances being made today are paving the way for the systems of the 1980's. Of particular interest are the anticipated changes in cost for the major components of on-line systems. A study by Vincent Giuliano and Robert Kvaal of Arthur D. Little, Inc.,[1] indicates that by 1983 costs for central processing will be less than half of today's. By this date, communications costs will have decreased to one-third or less, and the cost of an "intelligent" CRT terminal will have decreased by a factor of 5. Their predictions with regard to the cost of random-access memory hardware are particularly interesting. Today this hardware costs about 30¢ for the storage of a page of printed text. They predict that this cost will be 1¢ or less in the 1980's and that it will be cheaper to store information in random-access memory than to print, distribute, and store hard-copy publications. With these changes in costs, we can view the 1980's as the period of widespread development of CRT terminals to access remote data bases. They will be a standard tool in research laboratories to be used by the scientists for many purposes, including literature searching and reading scientific publications.

With this perspective of the three decades, I would now like to turn to the subject of the present use of on-line systems at Pfizer Central Research and how we have planned to use them.

Figure 1 shows the systems routinely used, the supplier, the cost per connect hour including Tymshare costs, and

† Presented at the first meeting of the Association of Information Officers in the Pharmaceutical Industry, Canterbury, England, July 9–11, 1974, and before the Division of Chemical Literature, 168th National Meeting of the American Chemical Society, Atlantic City, N.J., Sept 10, 1974.

DATA BASE	SUPPLIER	COST/HOUR	OFF-LINE COST PER CITATION
CAIN	SDC	$40	$.08
CHEM7071	SDC	55	.08
CHEMCON	SDC	55	.08
MEDCOMP	SDC	44	.03
MEDLINE	SDC	55	.03
PREDICASTS	LOCK.	70	.05
TOXLINE	NLM	15	.10/PAGE

Figure 1. Use of on-line data bases at Pfizer Central Research.

DATE	TIME			*SYSTEM			**OFF-LINE PRINT	USER	REQUESTER	SEARCH PARAMETERS
	ON	OFF	TOTAL	M	CC	C				
4/29	2:25	2:31	6'	√			√22	JSB	T. DEVON	DIAZEPAM & STIFF-MAN
4/30	1:00	1:05	5'			√	√19	JBH	N. SCOLLICK	MANUF.-TERRAMYCIN

Figure 2. Logbook format.

```
SS 1 /C?
USER:
DIAZEPAM AND HUMAN

PROG:
PSTG (790)

SS 2 /C?
USER:
1 AND MUSCULAR DISEASES

PROG:
PSTG (5)

SS 3 /C?
USER:
1 AND MUSCLE CRAMP

PROG:
PSTG (3)

SS 4 /C?
USER:
1 AND SPASM

PROG:
PSTG (17)

SS 5 /C?
USER:
2 OR 3 OR 4

PROG:
PSTG (22)

SS 6 /C?
USER:
STRS 5 (TI) :STIFF:

PROG:
PSTG (3)

SS 7 /C?
USER:
"PRT 3,TI,SO"

PROG:

TI- (THE STIFF MAN SYNDROME)
SO- REV RHUM MAL OSTEOARTIC 38 253-63 APR 71

TI- GENERALIZED MUSCULAR STIFFNESS, FASCICULATIONS, AND MYOKYMIA OF
    PERIPHERAL NERVE ORIGIN.
SO- ARCH NEUROL (CHICAGO) 22 430-9 MAY 70

TI- (PSYCHOGENIC ASPECTS IN THE STIFF-MAN SYNDROME.  PRESENTATION OF A
    CLINICAL CASE)
SO- ENCEPHALE 58 349-59 JUL-AUG 69
```

Figure 3. MEDLINE search for Diazepam and the stiff-man syndrome.

the cost for off-line printing of search results. With the exception of TOXLINE, these costs must appear quite high, but with proper planning, we are convinced that these systems are the least expensive methods for searching the recent literature today. At our laboratories there are four information scientists who perform searches. Each knows how to operate all systems, but individuals specialize in certain systems to capitalize on experience and to batch searches in a given system. The following five policies have been established on the basis of experience and cost data.

(1) To notify all information scientists before a system is used, to increase search batching.

(2) To encourage laboratory scientists to sit with the information scientist at the terminal, particularly for their first query in a system. This develops understanding and encourages suggestions of synonyms and search strategy on future questions.

(3) For the information scientist to prepare a handwritten search strategy for each question before a terminal session.

Telephone	$ .66
Tymshare	1.00
Terminal	.30
Search Preparation	1.60
Searching Time	1.60
Data Base	4.50
Off-Line Print	.66
	TOTAL $10.32

Figure 4. MEDLINE, search cost.

(4) When possible, to emphasize off-line printing rather than printing at the terminal. If a search yields more than 15 to 20 hits, try to select the most pertinent references or print Author, Title, and Source for the 10 most recent articles on-line and print the full information off-line. The terminal sheet with the search strategy is immediately given to the requester with an explanation.

(5) To enter all searches into a logbook. This has been most helpful in avoiding duplicate searches and in solving problems with regard to bills. The form of the log is illustrated in Figure 2. The date, the time, the system used, ordering of off-line prints, the searcher, requester, and the search subject are recorded. The two searches shown will be considered in detail.

The MEDLINE search was for references to the use of Diazepam in the stiff-man syndrome. The benzodiazepins are good muscle relaxants as well as psychotherapeutic agents. Details of the actual search are illustrated in Figure 3. The first search statement is straightforward since Diazepam is a MESH term. However, the stiff-man syndrome is not in MESH, so it was necessary to combine the first statement with several MESH terms that might be used for this concept. Combination of the hits resulted in 22 references that were considered to be too many to print at the terminal. The decision was made to print the references off-line and to look for the most pertinent references by stringsearching the titles for the character sequence STIFF. This yielded the three references shown. It is interesting to note that two of the three references were to the stiff-man syndrome, and the other one resulted from the word "stiffness" in the title. A better strategy would have been to use the variable character key between STIFF and MAN. A most important point is that none of these three references can be found in a hand search of *Index Medicus* under the heading Diazepam. Diazepam was not selected as a "starred" or indexing term for these three articles, and, in fact, it was selected for indexing in only 10 of the 22 off-line references. This illustrates the greater power of on-line searching where all MESH terms are available. This search required 6 min of connect time, and an estimate of the total cost of the search is presented in Figure 4. In this estimate, the weighted average cost for the time of an information scientist was used, and it was assumed that the time required for search preparation was equal to the search time. To determine whether or not this cost of $10.32 was reasonable, a comparison was made with other methods for accomplishing this search.

Figure 5 shows the costs of the alternate search methods. A MEDLINE search can be purchased for $50 from the PMA Science Information Services. To obtain the cost of manual searches, the Cumulated *Index Medicus* was scanned for 1970–1973 under the therapeutic use of Diazepam. Pertinent references were handwritten and submitted to a typing pool. The searching time was 36 min, and using weighted average costs for a research scientist and an information scientist plus typing costs the $19.25

SEARCH METHOD	COST
Purchased MEDLINE	$50.00
Manual, Research Scientist	19.25
Manual, Information Scientist	16.25
On-Line, Research Scientist	19.54
On-Line, Information Scientist	10.32

Figure 5. Cost comparisons for the MEDLINE search.

```
SS 1 /C?
USER:
ALL MANUF# OR ALL PREPN# OR ALL SYNTH# OR ALL PROD#
PROG:
PSTG (33729)
SS 2 /C?
USER:
ALL PROCESS# OR ALL RECOV# OR ALL ISOL# OR ALL PURIF# OR ALL FERMN#
PROG:
PSTG (15186)
SS 3 /C?
USER:
TERRAMYCIN OR OXYTETRACYCLINE
PROG:
PSTG (206)
SS 4 /C?
USER:
3 AND ALL CA016#
PROG:
PSTG (11)
SS 5 /C?
USER:
3 AND 1
PROG:
PSTG (10)
SS 6 /C?
USER:
3 AND 2
PROG:
PSTG (7)
SS 7 /C?
USER:
4 OR 5 OR 6
PROG:
PSTG (19)
SS 8 /C?
USER:
"PRT FULL,OFF-LINE STOPAD"
```

Figure 6. CHEMCON search for the manufacture of terramycin.

```
TI- PRODUCTION OF OXYTETRACYCLINE USING STREPTOMYCES ALBOFLAVUS (ATCC
    15388)
KW- OXYTETRACYCLINE STREPTOMYCES
CS- CA016000
JC- USXXAM
AU- VILLAX, IVAN
SO- U.S. 6 PP.
PY- 250172
UC- P
CP- CA07625152033V
CV- 100000
PN- 3637463
PA- INTERNATIONAL RECTIFIER CORP.
PC- 195/80; C 12D
PD- 191062
PP- 40,140
PL- PORTUGAL
```

Figure 7. Pertinent reference by searching Section 16.

Telephone	$ .66
Tymshare	.83
Terminal	.25
Search Preparation	1.33
Searching Time	1.33
Data Base	3.75
Off-Line Print	1.52
TOTAL	$9.67

Figure 8. CHEMCON search cost.

SEARCH METHOD	COST
Purchased CHEMCON	$125-300
Manual, Research Scientist	55.16
Manual, Information Scientist	43.66
On-Line, Research Scientist	17.41
On-Line, Information Scientist	9.67

Figure 9. Costs for searching chemical abstracts.

	NUMBER PERTINENT REFERENCES	NUMBER UNIQUE REFERENCES
Manual Chemical Substance Indexes	10	7
CHEMCON	4	1

Figure 10. Comparison of manual and CHEMCON searching of *Chemical Abstracts* for the year 1972.

and $16.25 figures were obtained. The labor figure for the research scientist is higher owing primarily to higher laboratory overheads.

To estimate the cost of the on-line search by a research scientist, we estimated that his preparation time would be the same as that of an information scientist since he would be more familiar with synonyms but less familiar with the searching tools. To estimate his searching time, we went to our logbook for data on experienced users versus inexperienced users. During the first two months of on-line system use, our information scientists required an average of 15.5 min per search. In April of 1974 this had decreased to 7.2 min, so it was assumed that the research scientist would require 12 min for the Diazepam search. From these data we can conclude that the $10.32 cost was indeed reasonable. Also the computer searching was better than the manual, because specific references to the stiff-man syndrome were produced.

Let us now turn to the CHEMCON search that was shown in Figure 2. The search was for references to the manufacture of Terramycin™. This type of search for the manufacture of a product is encountered frequently, and a standard search strategy has been developed by Mr. John Hare from our group as shown in Figure 6. The trick here was to get all of the keyword synonyms for the word manufacture while avoiding an overflow that occurs at about 40,000 postings. These synonyms were carried in the first two statements, and the antibiotic was covered in the third statement. Note that in the fourth statement the antibiotic was combined with the Fermentation Section 16 of *Chemical Abstracts* to catch any references that do not have the keywords of the first two statements. The 19 hits were printed off-line, and the search required 5 min of connect time. Of the 19 references, 10 were concerned with the manufacture of Terramycin™. The false drops were concerned with the preparation of dosage forms or the effect of the drug on the preparation of other materials. The use of Section 16 of *Chemical Abstracts* produced two of the false drops, but was responsible for two of the pertinent references.

One of these references, shown in Figure 7, was a U.S. patent that had none of the search keywords but was assigned to Section 16. The cost of this search was estimated using the same assumptions that were used for the MEDLINE estimate. This cost estimate came to $9.67 (Figure 8), and to determine whether or not this cost was reasonable, the costs for obtaining this search by other available methods were obtained as shown in Figure 9.

The cost range for a purchased *CA Condensates* search was obtained from a commercial supplier. To estimate the cost of a manual search, an actual search was carefully timed in our library and the 2.3 hr required was used for both the research scientist and the information scientist. It is obvious from these data that the on-line system provides the lowest costs. However, the quality of searching is important, and this is illustrated in Figures 10 and 11.

Figure 10 shows the results obtained for the year 1972. This year is presented separately because the manual search was performed using the Chemical Substance Indexes for Volumes 76 and 77. These indexes represent in-depth indexing of an article in contrast to the keywords used in

	NUMBER PERTINENT REFERENCES	NUMBER UNIQUE REFERENCES
Manual Keyword	9	3
CHEMCON	6	0

Figure 11. Comparison of manual and CHEMCON searching of *Chemical Abstracts* from January 1973 to April 22, 1974.

```
"FILE CHEMLINE"
PROG:
THIS TERMINAL IS CONNECTED TO THE CHEMLINE RETRIEVAL FILE SET
SS 1 /C?
USER:
VINYL CHLORIDE
PROG:
PSTG (1)
SS 2 /C?
USER:
"PRT FULL"
CN- 75-01-4
MF- C2-H3-CL
N1- ETHYLENE, CHLORO- (8CI)
N1- ETHENE, CHLORO- (9CI)
SY- CHLOROETHENE
SY- CHLOROETHYLENE
SY- VINYL CHLORIDE
SY- VINYL C MONOMER
SY- TROVIDUR
SY- MONOCHLOROETHYLENE
SY- VINYL CHLORIDE
WL- G1C1
SS 2 /C?
USER:
?
TIME 00:01:26
```

Figure 12. CHEMLINE search for vinyl chloride.

CA Condensates that are derived from the title or context of an abstract. The superiority of the manual search is obvious, and it is important to determine why CHEMCON missed so many references. Four references were missed because the search strategy did not include terms for ion exchange processes. This could be easily corrected, but it emphasizes the necessity for imagination in the selection of search words. The remaining three unique manual references clearly discussed the manufacture of Terramycin™ in the abstract, but the antibiotic was not selected as a keyword. The one unique CHEMCON reference concerned lyophilization for the preservation of particular strains of streptomyces used for Terramycin™ production. These differences are very important to a research organization, and they suggest that the development of on-line searching of the Chemical Substance Indexes should be planned for more efficient searching of *Chemical Abstracts*.

For the period of January 1973, through April 22, 1974, the manual search used the keyword indexes from the individual issues. The results are shown in Figure 11. As one would expect, the difference in results is not as large. Ion exchange was responsible for two of the unique manual references, and the third reference concerned hygienic standards for atmospheric levels of tetracyclines.

TOXLINE was recently transferred from a commercial organization to the National Library of Medicine. This is the first opportunity that industrial users have had to directly access the NLM computers, and the resulting improvements have been appreciated. These improvements include much lower costs, faster response time, fewer system problems, and the use of the ELHILL programs that in our opinion are more convenient than the RECON programs. It should be emphasized that the TOXLINE system is still under development and that many additional files will be included. One search has been chosen to illustrate the use of the Chemical Dictionary as well as TOXLINE.

```
TI- DISTURBANCE OF HUMAN HEALTH BY CHEMICAL SUBSTANCES_TOXICITY OF
    ORGANOCHLORINE COMPOUNDS.)
SS 6 /C?
USER:
"PRT 1,FULL,SKIP 6"
PROG:
SI- HAPAB/73/01170
AU- UEDA K
TI- DISTURBANCE OF HUMAN HEALTH BY CHEMICAL SUBSTANCES_TOXICITY OF
    ORGANOCHLORINE COMPOUNDS.)
LA- JAPANESE
SO- NIPPON ISHIKAI ZASSHI (J. JAP. MED. ASS.); 69(2): 207-216; 1973
AB- HAPAB  GENERAL PHYSIOLOGICAL PROPERTIES OF ORGANOCHLORINE
    (ORGANOHALOGEN) COMPOUNDS ARE REVIEWED.  ORGANOCHLORINES ARE VERY
    SOLUBLE IN FATS, ACCUMULATING IN FAT TISSUES AND IN THE LIVER, ARE
    NARCOTIC, PRODUCE ACNE, AND DAMAGE THE LIVER.  METABOLIC AND - - -
?
TIME 00:07:35
```

Figure 13. Title and full record from the seventh TOXLINE reference.

Telephone	$ .82
Terminal	.43
Searching Time	2.30
Tymshare and System	2.15
	TOTAL $5.70

Figure 14. Costs for the CHEMLINE and TOXLINE searches.

This example is a search for the toxicity of vinyl chloride that has received so much publicity. To enter the Chemical Dictionary, the command "FILE CHEMLINE" is used as shown in Figure 12, and the substance being searched can be entered in a number of ways including the name, CA registry number, or WLN. In this case the name vinyl chloride was entered, and the program responded that it has one dictionary reference with this name. The command was then given to print the dictionary data in full which included the CA registry number, molecular formula, CA nomenclature by the Eighth and Ninth Collective Indexes, Synonyms, and the Wiswesser line notation. The dictionary search required 1.26 min.

Following the dictionary search, the command "FILE TOXLINE" was used to enter the TOXLINE file. The CA registry number obtained from the dictionary produced 14 references. Since CA registry numbers only appear in the CBAC, HEEP, and PESTAB portions of the file, the inverted file was also searched for the terms vinyl and chloride to obtain 140 postings. Combination of these references produced 142 references to vinyl chloride. These were then combined with toxicologic terms (toxicity or toxicology or all carcinogen# or all mutagen#) to obtain 23 postings. The titles of these articles were printed at the terminal, and the seventh title shown at the top of Figure 13 looked particularly interesting. The command was then given to print the full seventh record which included the secondary source, author, title, language, source, and a full abstract. Only part of the abstract is shown in Figure 13. The time for the TOXLINE search was 7.35 min.

The total cost for the CHEMLINE and TOXLINE searches was estimated at $5.70 as shown in Figure 14. Search preparation time is not included because the use of the dictionary obviated prior preparation. Perhaps the costs of this system can be viewed as a preview of the costs that will generally obtain for the on-line systems of the 1980's.

LITERATURE CITED

(1) Giuliano, V. E., and Kvaal, R. J., "Outlook for Data Base Publishing," Arthur D. Little, Inc., Discussion Meeting, Sept 1973.