

18 J. SHERBER, C.A. MAPCUS, M.
KLEINBERG, S.

FORDHAM HOSP., EFCNX, N.Y.
27166E

RAPID CLEARANCE OF ISOSCRIBE DINITRATE FROM RABBIT BLOOD—
DETERMINATION BY GAS CHROMATOGRAPHY.

BIOCHEM. PHARMACOL.
15 667- 612(1970)

BIOCHEMICAL METHODS
VASCULATORS, CECARY
NITRATES AND NITRITES
ISOSCRIBE DINITRATE
SORBITRATE
ADMINISTRATION, IV
ANIMAL STUDY
RABBITS
BLOOD
CRUG CLEARANCE
CIPHENYLAMINE
DETECTION
CHROMATOGRAPHY (GLC)
BIOCHEMISTRY

Figure 9. Page of bibliography containing all descriptors

CONCLUSION

The Atlas Biomedical Literature System (ABLS) has been in operation for three years. It presently contains references to more than 7000 articles and is growing at the rate of about 4000 articles per year. There is no limit to the number of condensates the ABLS can contain. It is a flexible and useful information tool capable of producing bibliographies in various forms and of selecting citations from broadly or narrowly defined areas.

ACKNOWLEDGMENT

The authors are indebted to the following members of the Atlas Organization and Systems Planning Department: G. Ward Keever, III, for the initial work in planning the

computer system and for adapting the Mark IV Management System to the special requirements of the ABLS; Charles W. Audette for expanding the system and providing additional computer programs for use in printing the semi-monthly bulletins; Barbara K. Casey for writing the computer programs used for editing the proof copy of the condensates and validating the descriptors; and David L. Freese for help in preparing some of the retrieval programs. We are also deeply indebted to Frederic R. Benson for his support and encouragement.

LITERATURE CITED

- (1) Starke, A. C., Whaley, F. R., Carson, E. C., and Thompson, W. B., "GAF Document Storage and Retrieval System," *Amer. Doc.* **19**, 173-80 (1968).
- (2) Ackermann, H. J., Haglund, J. B., Lindwall, H. G., and Maizell, R. E., "SWIFT: Computerized Storage and Retrieval of Technical Information," *J. Chem. Doc.* **8**, 14-9 (1968).
- (3) Gillis, C. N., "Biomedical Information Retrieval: A Computer-Based System for Individual Use," *J. Chem. Doc.* **7**, 98-100 (1967).
- (4) Horvath, P. J., Chamis, A. Y., Carroll, R. F., and Dlugos, J., "The B. F. Goodrich Information Retrieval System and Automatic Information Distribution Using Computer-Compiled Thesaurus and Dual Dictionary," *J. Chem. Doc.* **7**, 124-30 (1967).
- (5) Teal, T. W., and Greenberg, S. M., "Managing Literature in the Pharmaceutical Industry," *Drug Inform. Bull.* **2**, 136-43 (Oct./Dec. 1968).
- (6) Bennett, R. E., and Frycki, S. J., "Internal Processing of External Reference Services," *J. Chem. Doc.* **11**, 76-83 (1971).
- (7) Clark, G. M., "Co-ordinate Indexing of Pharmaceutical Information," *Bull. Med. Lib. Ass.* **50**, 422-8 (July 1962).
- (8) Kreiter, C. S., and Galindo, J. G., "Filing and Retrieving Pharmaceutical Information," *Drug Stand.* **26**, 151-2 (1958).
- (9) Mark IV Management System supplied by Informatics, Inc., Sherman Oaks, Calif.

An Experimental Computerized CBAC Search Project

STEPHEN R. HELLER,* RICHARD J. FELDMANN, and KENNETH P. SHAPIRO
Heuristics Laboratory, Computer Center Branch and Data Management Branch,
Division of Computer Research and Technology, National Institutes of Health,
Department of Health, Education and Welfare, Bethesda, Md. 20014

Received September 17, 1971

A literature search system for CBAC is described. Both SDI and retrospective searching techniques are available to the user. The system relies heavily on searching keywords from the CBAC text rather than the text itself and employs batch/tape and time-sharing/disk hardware.

An experimental Chemical Information System, consisting of the literature, substructure search, and property files is being tested at the Division of Computer Research and Technology of the National Institutes of Health.¹ One part of this experimental system, the literature project, is the topic of this paper.

*To whom inquiries should be addressed.

THE DATA BASE

The major problem that confronts a new information center is what data base to use. Some data bases cover a limited area, some are broad, some contain only titles, some contain full abstracts, some are highly indexed with complex dictionary and index terms, some provide rapid literature indexing and abstracting, and some are slower but more thorough. One, or even two, data bases will not cover

AN EXPERIMENTAL COMPUTERIZED CBAC SEARCH PROJECT

the needs of the broad biochemical/medical research complex at NIH. MEDLARS is already used extensively at NIH, but is highly medically oriented.

Among all the relevant sources of chemical/biological data bases, Chemical Abstracts Service has the largest computer-readable files as well as plans for greatly expanding its services in computer-readable data bases. Chemical-Biological Activities (CBAC), one of the journals published by Chemical Abstracts Service, was chosen as the first experimental data base. While CBAC does not cover as many journals as one might prefer, it does cover a wide range of chemical and biochemical, and some clinical, journals. Its outstanding features from the user viewpoint appeared to be the availability of the full abstract on computer tape and the large number of searchable data elements which include author, title, location of work, CODEN, abstract, molecular formula and REGN (Chemical Abstracts Registry Number).

THE SEARCH PROGRAMS

Computer text-searching programs run either in batch mode or are interactive. The frequency of appearance of CBAC (fortnightly) makes it reasonable to do current awareness or SDI (Selective Dissemination of Information) searching in the batch mode.

Retrospective searching is a different matter. If a scientist is writing a review or needs to find references for a broad study, a batch search is probably quite acceptable since a good deal of extraneous information can be tolerated (or may even be desirable to be sure of complete coverage). If he wants to explore a few references to a particular research topic, however, a narrow request is probably served best by interactive searching.

Another valuable use of an interactive retrospective search system is to establish a properly worded SDI profile for the particular data base. This is especially true for an open-vocabulary system such as CBAC.

At DCRT, experiments are being undertaken with three search systems: Batch SDI, Batch Retrospective, and Interactive Retrospective.

Batch SDI. Through the generosity of the National Science Library/National Research Council of Canada, DCRT has obtained and substantially modified the CAN/SDI system. The CAN/SDI system consists of four programs: reformatting, profile generation, searching, and printing.

The first and last parts have been completely rewritten and considerable changes have been made to the middle two.

While the CAN/SDI reformatting program would leave a given data base essentially in the original form, the DCRT modification using the CBAC data base makes a very substantial change and thus it may be preferable to call this first program a reprocessing program.

The CBAC tapes are read in and one record is built for each abstract (originally CBAC had about 40 to 50 records) which consists of tags of data and pointers to these tags within the record. For example, the tag for REGN contains a list of all the registry numbers in this abstract separated by delimiters. The advantage of this is that when searching for REGN, one need only look at the tag containing the REGN and not at any other part of the record. The same is true, of course, for the other data tags in the abstract record (keywords, authors, etc.).

In the reprocessing, two operations are performed on the title and abstract sections of CBAC. First a word is read in and if it is a "trash" word (words that are not of value for searching purposes), it is thrown out. If it is not a trash word, then a check is made for duplication in that abstract

6299 MECHANISM OF D-CYCLOSERINE ACTION: TRANSPORT SYSTEMS FOR D-ALANINE, D-CYCLOSERINE, L-ALANINE, AND GLYCINE. JOBAAY, 103,3,70,778-88
Wargel RJ, Shadur CA, Neuhaus FC; Dep. of Chem., Northwestern Univ., Evanston, Ill.

- | | |
|-----------------------------------|--|
| 338692
56417
56406
68417 | The accumulation of D-Alanine, L-Alanine, Glycine, and D-Cycloserine in <i>Escherichia coli</i> was mediated by 2 transport systems. |
| 6299-1 | The systems for D-Alanine and Glycine were related, and were separated from that involved in the accumulation of L-Alanine. |
| 6299-2 | D-Cycloserine appears primarily transported by the D-Alanine-glycine system. |
| 6299-3 | The accumulation of D-Alanine, Glycine, and D-Cycloserine was characterized by 2 line segments in the Lineweaver-Burk analysis, whereas the accumulation of L-Alanine was characterized by a single line segment. |
| 6299-4 | D-Cycloserine was an effective inhibitor of Glycine and D-Alanine accumulation, and L-Cycloserine was an effective inhibitor of L-Alanine transport. |
| 6299-5 | Glycine and D-Alanine at 10^{-3} M antagonized the effect of D-Cycloserine in <i>Escherichia coli</i> , while this concentration of L-Alanine had no effect, as expected since the primary access of D-Cycloserine in <i>Escherichia coli</i> is via the D-Alanine-glycine system. |

The title Digest section of CBAC is converted into keywords. The 140 words in the above digest and title are passed against the "trash" list and duplicates removed. the following 19 keywords remain and appear on the computer tape as follows:

```
MECHANISM
ACTION
DCYCLOSERINE
TRANSPORT
SYSTEMS
DALANINE
LALANINE
GLYCINE
ACCUMULATION
ESCHERICHIA
COLI
TRANSPORTED
DALANINEGLYCINE
SYSTEM
LINEWEAVERBURK
EFFECTIVE
INHIBITOR
ANTAGONIZED
EFFECT
```

Note that all punctuation marks are removed.

Figure 1. CBAC abstract before and after reprocessing step

and if duplication is found, then the duplicate word is not put into the keyword tag field. This keywording of titles and abstracts reduces the size of the field being searched to 15 to 45% of its original size (Figure 1).

The profile generation program is simply a program that "compiles" profiles for the search program. The program checks for syntax and is quite simple to use. The modifications in this program consist of adding two new data elements that can be searched—namely, REGN and molecular formula—and allowing for a wider variety in the number of hits or references to be retrieved for a given search. Normally from one to ninety-nine references can be re-

PROFILE NUMBER 0001 WITH TERMS MECHANISM , ACTION , GLYCINE
 CBAC VOLUME NUMBER 12 AND ISSUE NUMBER 08. ABSTRACT NUMBER 06299
 MECHANISM OF D-CYCLOSERINE ACTION TRANSPORT SYSTEMS FOR D-ALANINE, D-CYCLOSERINE, L-ALANINE, AND GLYCINE.
 MARGEL RJ, SHADUR CA, NEUHAUS FC.
 DEP. OF CHEM., NORTHWESTERN UNIV., EVANSTON, ILL.
 JOBAA, 0103, 0003, 70, 77A-88

THE ACCUMULATION OF
 000338692 D-ALANINE ,
 000056417 L-ALANINE ,
 000056406 GLYCINE , AND
 000068417 D-CYCLOSERINE IN ESCHERICHIA COLI WAS MEDIATED BY 2 TRANSPORT SYSTEMS.

THE SYSTEMS FOR D-ALANINE AND GLYCINE WERE RELATED, AND WERE SEPARATED FROM THAT INVOLVED IN THE ACCUMULATION OF L-ALANINE .

D-CYCLOSERINE APPEARS PRIMARILY TRANSPORTED BY THE D-ALANINE-GLYCINE SYSTEM.

THE ACCUMULATION OF D-ALANINE , GLYCINE , AND D-CYCLOSERINE WAS CHARACTERIZED BY 2 LINE SEGMENTS IN THE LINEWEAVER-BURK ANALYSIS, WHEREAS THE ACCUMULATION OF L-ALANINE WAS CHARACTERIZED BY A SINGLE LINE SEGMENT.

D-CYCLOSERINE WAS AN EFFECTIVE INHIBITOR OF GLYCINE AND D-ALANINE ACCUMULATION, AND L-CYCLOSERINE WAS AN EFFECTIVE INHIBITOR OF L-ALANINE TRANSPORT.

GLYCINE AND D-ALANINE AT 10⁻⁵M ANTAGONIZED THE EFFECT OF D-CYCLOSERINE IN ESCHERICHIA COLI, WHILE THIS CONCENTRATION OF L-ALANINE HAD NO EFFECT, AS EXPECTED SINCE THE PRIMARY ACCESS OF D-CYCLOSERINE IN ESCHERICHIA COLI IS VIA THE D-ALANINE-GLYCINE SYSTEM.

Figure 2. Print program output for the abstract in Figure 1

requested for each profile, but special codes allow the user to get additional references. The program is accessible through a terminal-based text editor system (WYLBUR) and this allows users to code and check their own profiles if they so desire at any of approximately 250 remote terminals serviced by the DCRT/Computer Center Branch of the National Institutes of Health. Normally, one can get 5- to 15-minute turn-around time with this program, and thus it is very convenient to use.

To aid the user in profile formulation, a Profile Design Manual² was written. The 80-page manual, modelled after one written for the CAN/SDI project, contains seven sections. Included in the manual are a description of CBAC, numerous and varied examples of coded profiles, detailed instructions on coding, the "trash" list, the list of journals covered by CBAC, and sample output.

As a further aid to the user, listings of the REGN and corresponding names or molecular formula from CBAC were made available. Of the approximately 54300 unique REGN, about 3500 entries contained the same name with different REGN. The computer program used to pull this list from the data base assumed an error free format. In this case and in many instances in the reprocessing/reformatting of the six years data base, a great variety of errors were encountered. However, it was very pleasing to find the errors fell off sharply as a function of time between 1965 and 1970 CBAC tapes.

There are seven types of terms that can be searched: titles, authors, location of work, CODEN, REGN, Molecular Formula, and keywords. The last is the section that DCRT adds to the data base. Authors and keywords can be right truncated. For economic reasons only titles terms can be left truncated. CODEN and REGN of course, are not truncated. The logic is AND, OR, and NOT.

It is worth noting that of seven term types, keywords account for 95% of the terms used, followed by REGN 3%, all others 2%.

The third program is the search program, and this was modified to search for REGN and molecular formula in the reformatted tapes. The search program has been modified so that it no longer searches for phrases in the abstract section, but rather a keyword or a combination of keywords. For example, to search for "nuclear magnetic resonance" the user must request "nuclear" and "magnetic" and

"resonance." Clearly, combinations out of context are a distinct possibility although it has been surprising that this has not yet been found to have occurred.

The last program is the print phase which prints the abstract essentially as it appears in the hard copy printed version of CBAC (Figure 2).

Batch Retrospective. The same programs and tapes are used for SDI and batch retrospective searching. The reprocessing of the tapes and modifications to the other SDI program enable us to reduce very substantially the cost of retrospective searching of CBAC. In 10.9 CPU minutes on an IBM 370/165, we are able to search an approximately 89,000 document CBAC file for 77 profiles. Because of the efficiency of the batch sequential search in our multi-programming 360/370 system, we feel that development of an inverted file search system on the 360/370 containing full abstracts is unnecessary.

Interactive Retrospective. In addition to the batch 360/370 program previously described, we are experimenting with a streamlined CBAC search system which is interactive and is accessible via teletype terminals from a time-shared DEC PDP-10 computer at DCRT.

The streamlined Interactive Retrospective CBAC search uses an inverted file created by reading in the title, digest author, and REGN sections of the CBAC tapes on the PDP-10, eliminating "trash" words and duplicates as in the 360 reprocessing, collecting all keywords, author names and REGN along with their pointers to the CBAC volume, issue, and abstract numbers that they came from. The file is then sorted, each word on the sorted list is reduced to a 36-bit code by a technique called hash coding. The hash coding involves the use of a 3rd degree polynomial equation based on Fermat's last theorem. This hash word, along with the set of pointers for the word, are stored on disk packs. The search program simply asks the user to type in a subject word, author name, or REGN. The program then hashes the input word, compares the hashed word to those stored on the disk and, usually within 1 to 2 seconds, replies to the user that either the word is not in the file or that there are N references to that word. The user may then request the references immediately or type in further word(s) to use for further searching. The program will automatically search for the Boolean AND conjunctions for all terms typed in.

An example of the PDP-10 program input and output is:

Type word for search specification
DNA

Found 175 references to word DNA
References Y/N?

N

Type word for search specification
RNA

Found 133 references to word RNA
References Y/N?

N

Reference conjunction

Conj No.	No. Abstracts	Words
1	20	DNA, RNA
2	133	RNA
3	175	DNA

Thus, in a matter of minutes of connect time and some 2 to 3 seconds of CPU time, the user identified 20 references that contain DNA and RNA. Even further refinement would have been possible if the user wished to narrow his question. It is important to note that only the CBAC digest number is available here as output, as opposed to the full abstract, title, author, etc.

This program has been very helpful for generating SDI profiles or retrospective search requests for the 360/370 programs because the user can quickly check the existence in the system of a variety of words and their possible synonyms (necessary because CBAC is free text and does not have rigidly limited vocabulary).

USER REACTION TO A BATCH SDI SEARCH SYSTEM

Preliminary studies indicate that the users agree on several points:

The PDP-10 search is impressive, it is useful in its own right and is a strong inducement to try the 360/370 SDI and retrospective searches. Its main flaw is no abstract.

The full abstract (available via the 360/370 programs) is

the best part of CBAC. Users would be pleased if all of CA (and BA) contained full abstracts on tape. (Indeed some even said they would not consider using CA condensates or BA as they currently exist.)

Computer searching of data bases is valuable, but CBAC covers too narrow a range.

A number of users have expressed surprise to find journal articles retrieved that they had not seen previously.

CONCLUSIONS

The testing of CBAC is still in an experimental stage with a limited number of users. It is too early to determine if it will go into production status and, if so, whether it will need further refinements to achieve an optimal balance of the economic and informational needs of the NIH user community.

With the expansion of the size of the CBAC data base beginning in January 1972 (by about 40%), the range of coverage will obviously become greater and should please users. However, it remains to be seen if the intellectual value of increased coverage will offset the financial cost of the larger data base.

Quite clearly when CA with complete abstracts are available in computer readable form, a searchable keyword section would have great economic value.

ACKNOWLEDGMENT

The authors thank Inez Gaffney and James Heilik of the National Science Library, National Research Council, for their assistance. We also thank Keatha K. Krueger, Victoria Ragen, and Billie Mackey for their aid in the Pilot CBAC Study.

LITERATURE CITED

- (1) Feldmann, R. J., Heller, S. R., Shapiro, K. P., and Heller, R.S., "An Application of Interactive Computing—A Chemical Information System," *J. Chem. Doc.*, to be published.
- (2) Heller, S. R., "Profile Design Manual," DCRT/CIS, CBAC Literature Project.

Chemical Substructure Index (CSI)—A New Research Tool

CHARLES E. GRANITO* and MURRAY D. ROSENBERG

Institute for Scientific Information
325 Chestnut St., Philadelphia, Pa. 19106

Received June 9, 1971

In recent years, the extremely limited value of traditional indexes for substructure searching has become well recognized. As a result, considerable research effort has been financed for substructure search systems. However, with few exceptions, this research effort has been aimed at developing computer systems, even though most chemists do not have immediate access to computers. Manual substructure searches remain laborious or impossible to perform. The development of "desk-top" manually search-

able indexes by permuting line notations was first reported in 1963.¹ These indexes permit manual substructure searching that is impossible or impractical with traditional manual indexes such as the subject or formula indexes to *Chemical Abstracts*, *Index Chemicus*, or *Beilstein*. Various versions of permuted line notation indexes are now used by many large organizations for internal files. However, access to the open literature for the chemist interested in finding new compounds that contain a particular substructure was unachievable prior to the introduction of the *Chemical Substructure Index (CSI)*.

* To whom inquiries should be addressed.