

CAIN: A Computerized Literature System for the Agricultural Sciences†

JOSEPH F. CAPONIO*

Environmental Science Information Center, National Oceanic and Space Administration, Washington, D.C. 20235

LEILA MORAN

National Agricultural Library, U.S. Department of Agriculture, Beltsville, Maryland

Received January 28, 1975

CAIN is an on-line document locator and bibliographic control system for the collection of the National Agricultural Library with great potentiality of disseminating on a national and international level useful information on the agricultural sciences.

CAIN, an acronym for CAtaloging-INdexing, is a document locator and bibliographic control system for the collection of the National Agricultural Library. Bibliographic information about new monographic and serial titles and signed journal articles is captured when this material is cataloged and indexed. The Library, as one component of the national library system for the United States (the other two components are the Library of Congress and the National Library of Medicine), is primarily responsible for the collection and dissemination of agricultural information on a national and international basis. For over 112 years, publications in many languages have been acquired through gifts, exchange from foreign correspondents throughout the world, and by direct purchase. This literature is chiefly in monographic and journal format. Since 1970, the bibliographic records for newly acquired publications, including signed articles selected from some 6,500 journals and serials, have been exploited to build a data base of information in economics, rural sociology, animal industry, engineering, entomology, food and human nutrition, forestry, pesticides, soils, fertilizers, botany, and related subject fields.

The National Agricultural Library began in 1967 to develop systems to prepare indexes for the "Bibliography of Agriculture," the "Pesticides Documentation Bulletin," and for the categorical and alphabetical issues of the "Agricultural Biological Vocabulary." These systems were consolidated and expanded in 1969 to process all input data within one set of parameters. This system, inaugurated in 1970, includes all bibliographic records, numbering over 500,000 entered into the CAIN data base since that date. CAIN does contain some earlier records published in the now defunct "Pesticides Document Bulletin" from December 1967 through December 1969.

Bibliographic records from the Food and Nutrition Information and Educational Materials Center of the National Agricultural Library were added to the CAIN data base in 1973. These records differ from the others added to CAIN, in that they contain abstracts. This additional information is required by the Center in order to carry out its mission to support the School Lunch and Breakfast Programs and training programs for school food service personnel of the Food and Nutrition Service, USDA.

Other data bases served by the CAIN system, but not included in the major one, are the Herbicides data base of the Agricultural Research Service and the International Tree Disease data base of the U.S. Forest Service. Access to these data bases is through their respective agencies.

SEARCH AND RETRIEVAL SERVICES

Access to the CAIN data base is available to potential users through several channels. The data base is sold on a world-wide basis in the form of monthly tapes. The cost is \$45.00 per reel. This covers merely reproduction costs. The complete CAIN data base in total (1967-69 Pesticides data only and 1970-1974 full library data) or by year is also available. Prices are available on request.

Literature search services have been primarily in the form of current awareness profiles. These profiles have been made available to U.S. Department of Agriculture researchers through the Department's Agricultural Research Service. General public access has been available through the Universities of Florida and Georgia, and a number of other locations around the world.

On-line interactive bibliographic search and retrieval service was put into operation July 1973 (Figure 1). This on-line service is designed to serve the bibliographic information needs of the personnel of USDA and the agricultural community at large through the National Agricultural Library. Service is available free-of-charge to employees of the Department and to on-site users of the Library's resources. On-line searching may also be used at the discretion of the reference staff in answering queries for information received by mail. Scientists and researchers outside of USDA can get on-line to CAIN through Lockheed Information Systems, Palo Alto, Calif., or Systems Development Corporation, Santa Monica, Calif.

PRINTED PRODUCTS

Publications currently produced from the CAIN tapes include book catalogs, bibliographies, catalog cards, journal title listings, and various management assistance programs. For example, the "Bibliography of Agriculture" is a monthly author/subject index to the literature of agriculture and allied sciences based on the bibliographic records prepared by the Library for its CAIN system. The Bibliography primarily includes citations to journal articles, government documents, special reports, and proceedings of scientific and technical meetings. Each issue contains an average of 10,000 main entries. Bibliographic data elements for citations include an indication of the availability of an English summary of those articles which appear in a foreign language. Each issue contains a standard main entry section, personal author, corporate author, and subject indexes. In 1973 a geographic index, expanded corporate index, and an enhanced and updated subject vocabulary were incorporated into the Bibliography. The Bibliography was published

† Paper presented before the Division of Chemical Literature, American Chemical Society, Sept 10, 1974, Atlantic City, N.J.

COMPUTERIZED LITERATURE SYSTEM FOR AGRICULTURAL SCIENCES

a

```

*****
*                                     *
*                                     *
*      Lockheed Information Retrieval Service      *
*                                     *
*                                     *
*****

```

User:9018 Date:06/06/75 Time:13:30:49 File:10

Search History		Print Summary				
Set	Items Description	No.	File	Accn/Set	Pmt	Item-range
1	0 BEBEE/SHEAFFER	1	10	6	2	1-108
2	348 SLUDGE					
3	102 HEAVY (C) METALS					
4	5271 TEMPERATURE					
5	7071 CORN					
6	108 (2+3+4)*5					

Srch time 8.47 Print count 108 Descs.: 5 Docs.: 0

File:10, NAL/CAIN

User:9018 Page 14 (Item 93 of 108)

b

472 N21 ID NO.- 71-9046109 213434
Role of temperature in the regulation of leaf extension in Zea mays.
.Corn.
Watts, W R
Nature 229 (5279): 46-47. Jan 1, 1971

450 P58 ID NO.- 71-9045517 213060
Aftereffect of temperature on the degree of inhibition of plant
respiration with fluoride and malonate. .Corn, sunflowers.
LANGUAGE- RUS
Artemova, E K; Zemlianukhin, A A
Fiziol Rast 17 (5): 962-969. Sept/Oct 1970

1.9 P69P ID NO.- 71-9010446 211284
Elevation and temperature effects on severity of maize dwarf mosaic
virus in Sorghum in Arizona. .Corn.
Hine, R B; Osborne, W E; Dennis, R E
Plant Dis Rep 54 (12, pt. 1): 1064-1068. Dec 1970

4 AM34P ID NO.- 70-9092992 208163
Effect of mulches and bed configuration. II. Soil temperature and
growth and yield responses of grain sorghum and corn
Adams, J F
Agron J 62 (6): 785-790. Nov/Dec 1970

442.9 B87 ID NO.- 71-9004206 205108
Effect of high temperature on the growth, quantity of nucleic acids
and protein nitrogen of root tip of corn seedlings
LANGUAGE- BUL
Angelova, S T
Bulg Akad Nauk Inst Fiziol Rast Metod Popov Izv 16: 257-274.
1970

QK873.M5 ID NO.- 70-9298418 BOOK CAT- 00000000 176765
The effects of several transpiration retardants on the water
balance, stomatal aperture, and leaf temperature of field corn leaves
Miller, Neil Austin; 1932-
.Carbondale. 112 l. illus. 1968
DESCRIPTORS- Plant Transpiration

SB191.M2P3 ID NO.- 70-9294923 BOOK CAT- 00000000 173686
Effect of root-zone temperature and phosphorus supply on tissue
levels of nonstructural carbohydrates, aconitate, soluble amino acids,
and organic phosphorus compounds in relation to corn growth
Patterson, Robert Preston; 1939-
.Ithaca. 169 l. 1968
DESCRIPTORS- Maize, Growth, Soil Temperature, Effect On Maize,
Fertilizers, Effect On Cereals

Figure 1. (a) Computer print-out of a search strategy; (b) computer print-out of a CAIN on-line search showing citations to journal articles and monographs.

from 1970 to 1974 by Macmillan Information, a division of the Macmillan Publishing Company, Inc., which prepared and published the Bibliography under control of its own computer indexing system. Publication was taken over by

Oryx Press in 1975.

Books, new serial titles, and other types of library materials are announced monthly in the "National Agricultural Library Catalog" published by Rowman and Littlefield



Figure 2. CAIN on-line terminal and teletype printer at the National Agricultural Library.

Publishers. The Catalog contains a listing by main entry and indexes of personal authors, corporate authors, subjects, and titles. The index portions of the book catalog are cumulated semiannually. Multiyear cumulations may also be produced. A quinquennial edition, 1966–1970, in 12 volumes, was published by Rowman and Littlefield in 1973.

EVALUATION

The CAIN on-line system has the ability to handle many terminals with simultaneous interrogation and a high-speed reaction time. The user can carry on a direct dialog with the computer, which responds to the directions and redirections of the user seeking information on a specific subject (see Figure 2). The U.K. Office of Scientific Technical Information commissioned investigations of user reactions to CAIN services. The conclusions of the study indicated that CAIN provides good coverage of the world agricultural literature.¹

Several other studies have been sponsored by NAL on usage of the CAIN on-line system. The University of California Library at Davis, using NAL grant funds, made a comparison of SDI-current-awareness and on-line retrospective bibliographic search services from the CAIN Data Bases.² The CAIN SDI service has been in operation at Davis since January 1972 and the CAIN on-line service started in October 1973. In order to compare the two services seven profiles were run against both services for comparison of content, workload, and cost data. Some of the conclusions drawn were: (1) care should be taken to keep the searches as similar as possible; profile analysts need to be versed in the data base design, the translation and search programs, and the print controls and limitations in order to be able to prepare the best profiles; (2) in comparing the amount of literature searched on-line service is almost three times as productive as SDI service; (3) of the seven profiles that were run against both services, four re-

trieved identical bibliographies; of the other three profiles additional citations were retrieved through on-line search that were not retrieved in the SDI searches; the differences are attributable to differences in the search programs, IPS and DIALOG.

Cost comparisons were difficult to compute as both services were free to university users and the SDI costs and part of the on-line costs were paid through grants or library funds. However, charges to nonuniversity users for four years of SDI service were computed at \$190.00. The cost of on-line service for a four-year retrospective bibliography averaged \$15.19 or about 12.5 times less than four years of SDI services.

Analysis of user reactions seems to indicate that both SDI and on-line retrospective searches are desired by users of the CAIN data base. In May 1972, 90 SDI CAIN bibliographies were evaluated for citation relevancy, which averaged 70%. Relevance of on-line retrieved citations was ascertained to be 57%, based on questionnaires sent out during the first nine months of on-line service. Of 64 CAIN SDI users who also received retrospective on-line bibliographies, 34 (63%) indicated they would replace the SDI service with two or three retrospective searches per year and 30 (47%) indicated they would not want to replace their SDI service.

Of 217 respondents to the on-line questionnaire 193 (89%) indicated they thought the on-line retrospective bibliographies were worth the cost (average cost \$14.00 per bibliography) and 24 (11%) thought the bibliographies were not worth the cost.

NAL also made grant funds available to the Steenbock Library of the University of Wisconsin for a study of forestry literature indexing coverage in the CAIN data base.³ The results are equally applicable to the literature coverage of other scientific disciplines. At the Steenbock Library CAIN on-line accessing was an adjunct to the forestry literature study, providing an opportunity for demonstration

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"), Luxembourg, 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, stockbrokerages, etc. In our field, large data bases such as *Chemical Abstracts* and MEDLINE first became accessible on-line in 1972. Today, the Lockheed Information Sys-

tems 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.