Terse Literatures. III. The ARTEC Computer Program

TULL T. GLAZENER and CHARLES L. BERNIER*

School of Library and Information Studies, State University of New York at Buffalo, Buffalo, New York 14260

Received March 23, 1977

Boolean search, printout, updating, and bibliographic citations for collectaneas of Terse Conclusions (TCs) are programmed for the UNIVAC 1106 computer. Three data files (index, TCs, and bibliographic data) and five programs (search and retrieval, updating the TC or bibliographic files, updating the index file, and reordering TCs) are used.

INTRODUCTION

Terse Literatures (TLs) were invented to help searchers to read and use efficiently all data relevant to their interests. Of the different kinds of Tls,[1,2,4] Terse Conclusions (TCs) seem likely to be of the greatest value because authors' conclusions are major components of works. TCs report conclusions of authors concisely, clearly, and unambiguously; they do not depend upon context to be understandable by a person in the subject field. An example of a TC is, "Chromium(III) is a cofactor with insulin at the cellular level".[3]

The usefulness of TCs is improved by their organization into collectaneas by means of indexing, categorizing the index entries (each consisting of subject heading, TC, and reference), and the ad hoc logical ordering of the TCs under each subject heading. This organization of TCs increases the ability of the reader to assimilate, integrate, and use relevant conclusions of authors.

As new TCs come into the literature, it is desirable that they be merged logically and promptly into the organized collectaneas of TCs at hand.

ARTEC

ARTEC (*A*utomatic *R*etrieval of *TE*rse *C*onclusions) is the computer program (designed and implemented by T.T.G.) for this purpose. The TCs for use with ARTEC were supplied and indexed for a Newsletter.[4,5] ARTEC is operational on a UNIVAC 1106 computer as an on-line retrieval system. The programs were written in COBOL taking advantage of the ASCII character set. Approximately 220 TCs indexed by about 90 subject headings are presently in the database. A separate database of about 80 bibliographic references from the articles from which the TCs were derived is also maintained.

FILE ORGANIZATION

The system utilizes three separate data files: an index file (SYNDETICS), a Terse Conclusion file (TC-FILE), and a reference file (REF-FILE). Since the records on each of these files must be accessed and processed randomly, they were created as Indexed Sequential (ISAM) files. A discussion of the structure of each follows.

SYNDETICS

SYNDETICS is an ISAM file which is an index to the Terse Conclusion database. Each record of the file (SYN-REC) contains 106 characters. The actual key associated with each record (INDEX-KEY) consists of a 24-character subject heading (KEY-WORD) and a two-digit record suffix number (SYN-SUFFIX). Another two-digit field in the first record for each subject heading gives the number of additional suffix records (NO-OF-SUFFIX).

The remaining 78 characters of each record are divided into three fields of 26 characters each (TC-OR-CROSS). Each index entry may be one of three types: Terse Conclusion numbers ("TC"), "See" cross references ("SE"), or "See also" cross references ("SA"). The first two characters of each of these three fields is one of the above two-letter codes (REF-TYPE). If "SE" or "SA", the rest of the field contains a 24-character subject heading (CROSS-REF). If "TC", the remainder of the field is divided into eight three-digit TC numbers (TC-NO).

TC-FILE AND REF-FILE

TC-FILE and REF-FILE are ISAM files which contain the actual Terse Conclusions and bibliographic references, respectively. Both files have identical structures.

Each record on these files (DATA-REC) contains 167 characters. The actual key associated with each record (DATA-KEY) consists of a three-digit TC or reference number (DATA-NUM) and a two-digit record suffix number (DATA-SUFFIX).

The remaining 160 characters are divided into two fields of 80 characters each (INFO). These fields contain the text for the Terse Conclusions or the bibliographic references.

PROGRAM DESCRIPTIONS

The ARTEC system consists of five independent COBOL programs. One program performs search and retrieval operations, while the other four are "maintenance" programs. These programs are stored in absolute as well as source-code modules to facilitate on-line execution from a terminal.

LIPROJ

LIPROJ is the ARTEC program which performs the retrieval operations of the system. It is executed via the procedure file RUNRETRIEVE. It accepts input from the terminal and the three ISAM files and sends output to the terminal.

The program is capable of searching for Terse Conclusions indexed by a single or by any number of subject headings combined by the Boolean "AND", "OR", or "NOT" operations (*, +, and −, respectively). At present, the system is capable of parsing an expression with up to 28 levels of nesting.

The searcher enters the expression through the terminal in response to a system request. Each expression must be terminated by a semicolon (;). If the user enters an invalid subject heading, (i.e., one not recognized by the system), the system responds with an error message and asks the user to enter it again. Syntax errors (e.g., unbalanced parentheses, illegal operators, missing terminators, etc.) also generate specific error messages to the user.

In the case of *one* subject heading, the program searches the index file (via a random access read) and gets all TC numbers that are indexed by that heading. It then reads the TC file for those numbers and prints out the related TCs at the terminal. A reference number for the bibliographic

0095-2338/78/1618-0016$01.00/0

reference appears at the end of each TC. If the subject heading is recognized as a synonym for another heading, a "See" cross reference message is printed. After the TCs have been printed, a list of all "See also" cross references is also printed.

In the case of *more than one* subject heading, the program automatically parses the original expression into "post-fix" notation.[6] The expression is then evaluated as a series of binary expressions. The TCs from each operand pair are retrieved and placed on a temporary file. A message is printed out at the terminal indicating how many TCs were found for each heading. The program automatically retrieves any TCs from "See" cross references; "See also" cross references are ignored. The temporary file is then sorted according to increasing TC numbers (via the COBOL sort routine). This file is then read sequentially, checking each TC number for the specified Boolean relationship. TC numbers which meet the conditions are written to a "result file". When the expression has been completely evaluated, the TCs corresponding to the numbers on the final result file are printed at the terminal.

When the user is finished entering subject headings, he enters "@EOF" in response to the system's request for a query. The system will then respond with a request for any reference numbers. The user may then enter the reference numbers of any TCs for which he would like a bibliographic reference. The program then searches the reference file (via a random access read) and the reference is printed out at the terminal. The program terminates when the user enters a zero (∅) for the reference number.

FILE MAINTENANCE PROGRAMS

Maintenance of the files in the ARTEC system is handled by four programs. These programs allow the indexer to add new records to any of the files and to make corrections on existing records.

NEWDATA is a program which accepts as input either Terse Conclusions or bibliographic references and creates a properly formatted temporary file (TCDATA or REFDATA) which is then used as input for the corresponding permanent file. It is executed by the procedure files NEWTC or NEWREF, depending on whether the Terse Conclusion file or the reference file is being updated. The indexer types in the record number and the number of suffix records in response to the system requests. He then enters as many 80-character lines as necessary for those records. To change an existing record, the indexer types in that record number and the corrected lines. New records are created by entering a new record number. When all modifications and additions have been entered, the indexer types in a negative number in response to the system's request for the record number. A message advising the indexer that a temporary file has been created and to run the proper "loader" program is printed out at the terminal.

LOADATA is a program which reads records from a temporary file (TCDATA or REFDATA) and writes them to one of the ISAM files. It is executed by the procedure files LOADTC or LOADREF, depending on whether the Terse Conclusion file or the reference file is being updated. No input from the indexer is needed. The permanent file is "opened I-O", so that the writing of an existing record automatically becomes a "rewrite", and new records are inserted into their proper position. When all the records on the temporary file have been written on the permanent file, the total number of records written and the last record that was written is printed out at the terminal.

SYNLODER is a program which reads records from a temporary file (INDEXDATA) and writes them to the SYNDETICS ISAM file. It functions in the same manner as LOADATA (see above). At present, there is no program in the ARTEC system to create this temporary file. However, this can be easily accomplished by one of the UNIVAC 1106 system text editors.

REORDER is a program used to organize the TC's into logical collectaneas by reordering the TC's on the index file (SYNDETICS) into any order specified by the indexer. The indexer specifies any subject heading which he wishes to reorganize. The program retrieves all TC's indexed by that heading and prints them out at the terminal, each with a corresponding number. The indexer then enters the new order that he wishes by assigning new numbers to each TC in response to system requests. The program reorders the TC's and rewrites the record(s) onto the SYNDETICS file. The program terminates when the indexer types in "END" in response to the request for a subject heading.

REFERENCES AND NOTES

(1) C. L. Bernier, "Terse Literatures. II. Ultraterse Literatures", *J. Chem. Inf. Comput. Sci.*, **15**, 189–92, (1975).
(2) C. L. Bernier, "Terse Literatures", manuscript accepted for publication in the "Encyclopedia of Library and Information Science", 1977.
(3) W. Mertz, "Biological Role of Chromium", *Fed. Proc.*, **26**, 186 (1967).
(4) C. L. Bernier, "A Sample Newsletter of Biomedical Terse Conclusions", unpublished (Aug 9, 1971).
(5) D. C. Woodson update of the above "Newsletter" in 1976.
(6) J. R. Wirth, "Algorithms + Data Structures = Programming", Prentice-Hall, Englewood Cliffs, N.J., 1976, pp 198–9.