DEVELOPMENT OF AN APPLICATION SOFTWARE FOR ON-LINE INFORMATION RETRIEVAL

R. OJHA
D.K. BANDYOPADHYAY
S.R. KUNDU
INSDOC
Hillside Road
New Delhi-110012

A software package has been developed for on-line searching the INSPEC database. Care has been taken to maintain the portability of the package so that it can be run on any third generation computer system. Details of the package and its utilization for SDI service have been described.

INTRODUCTION

Research and development in the field of science and technology can be sustained only by regular flow of information which helps the scientists and technologists to keep track of development in their own fields all over the world and thus the chance of duplication of research is eliminated. Further, in an over-used cliche of information explosion, in the world of to-day, it is difficult for an individual scientist/technologist to go through a large number of scientific journals or other publications for the information he needs in his field of endeavour. Thus came into vogue the computerised Selective Dissemination of Information (SDI) service from the bibliographic data bases. After constructing the user scientist's/technologist's profile (knowing his field of endeavour) a suitable data base is searched at the SDI centre and searched result is dispatched to the concerned person. The computer operation associated with this type of searching is essentially a batch-mode operation. In the developed countries this type of information service has received well appreciation and acceptance from the user scientists. Again with the advent of the network operating computer system, the searching of a data base for giving SDI service can be taken away from the batch-mode operation and adapted to that of on-line searching which is now in vogue in the developed countries. In this paper the status of the SDI service (from INSDOC) in India and the feasibility for on-line searching the data bases has been briefly touched before discussing in detail the application package, developed for on-line searching the INSPEC data base.

2. SDI SERVICE IN INDIA (FROM INSDOC)

Canadian Institute for Scientific and Technical Information (CISTI) made a valuable contribution, by developing an application package named as CAN/SDI for giving SDI services in the early seventies. By this package different bibliographic data bases covering different subjects can be searched. Some of them are: CAS (Chemical Abstract Service), INSPEC (Information Services in Physics, Electrotechnology, Computer and Control), COMPENDEX (Computerised Engineering Index), NTIS (Government Report Announcements of the National Technical Information Service), ISI (Institute for Scientific Information) (i) Science Citation Index; (ii) Social Science Citation Index), etc., etc. Of these data bases INSDOC is getting the first three and SDI Centre of INSDOC is giving SDI service from CAS & INSPEC to the nation (about 450 user scientists) using CAN/SDI package, gifted by UNESCO. For rendering SDI service using CAN/SDI package there are five phases involved in the computer operation

1) Conversion phase
2) Compile profile phase
3) Search phase
4) Sort phase
5) Print phase

The package accepts two inputs i.e. (1) Data base tape, and (2) Profile information. The
profile is constructed according to the need of the user scientist/technologist. So, apart from installing the package (package has been installed at IIT, Madras, Computer Centre) and constructing the profile, we (INSDOC) are depending on the foreign package for giving SDI services.

3. THE CONCEPT OF ON-LINE SEARCHING

It is quite natural that one will be tempted to ask the question, can we not develop our own application software to search the bibliographic data bases for giving services? And again when India is already having network operating system (like Cyber 170 model 720, housed at NIC, New Delhi) can we not explore the possibility of utilizing the available system to render SDI service, on-line, by developing a suitable package? To answer these questions, we have come out with a project and as a first mission, have developed an application software which can search the INSPEC data base, on-line, according to the user's query. What an information seeker has to do is to enter his query at RJE (remote job entry) terminal. This query goes as data to the program package. The package program searches an issue of INSPEC data base tape, stored on a disk volume and displays one by one the 'hits' i.e., the records in the INSPEC data base tape that satisfy the query. One can also get the hard copy of the retrieved information, if desired.

4. DATA STRUCTURES OF THE INSPEC DATA BASE

INSPEC data base (1) comes in MARC format and is of variable length record format. Maximum record length is 6200 bytes (=characters). The tapes are unlabelled and in BCD character representation having physical record length of 2000 bytes (=characters). In MARC representation each bibliographic record is made up of three parts as given below.

1) First part is a fixed length leader (24 characters)
2) Second part is a variable length directory
3) Third part is the variable length data.

The leader: First five characters represent a number indicating the length of the record. Four characters (14th to 17th) represent a number indicating the length of the leader plus directory.

The directory: The directory is made up of a number of entries of equal length, viz, 12 characters each. The number of entries may vary from 1 to 49, depending upon how many types of bibliographic description the record contains.

One entry in the directory is made up of three parts

1) first three characters of the directory represent tag no.
2) next four characters represent the length of data corresponding to the tag no.
3) and last five characters indicate relative position of information corresponding to the tag no. from the end of the directory.

Data: This portion contains the actual bibliographic information. 1$, 2$, 3$ act as separators for different kind of bibliographic information, and within one kind of information a $ acts as a separator if there are more than one information of the same kind.

5. DESCRIPTION OF THE APPLICATION SOFTWARE FOR ON-LINE SEARCHING THE DATA BASE

For on-line searching the data base, the entire data for INSPEC tape is made resident on the disk storage. As given in the description of the leader, 14th to 17th characters represent a number, indicating the length of the leader and directory. However, the leader is of 24 characters in length. So, the length of the directory can be easily calculated. Again, each directory entry is 12 characters in length. So, the length of the directory divided by 12 gives the number of directory entries. From each entry, the tag no. and the relative position of information, corresponding to the tag no. from the end of the directory can be obtained. Though 49 types of bibliographic information can be present in a record, rarely more than half of such bibliographic information occur in the record. If there are n number of bibliographic information present, the access points of the data base can be any of the n items or their combinations. However, most of the queries can be satisfied by searching author, title, journal, abstract, corporate author, institution, year of publication, keywords, language and coden, either singly or with combinations.

The following two-character codes have been adopted to indicate these items

1) AT - Author of the paper
2) TI - Title of the paper
3) JR - Name of the Journal
4) AB - Abstract of the paper
5) CA - Corporate Author
6) IN - Institutions (Author affiliation)
7) YR - Year of publication
8) KW - Keywords
9) LN - Language
10) CD - Coden
In the package these two-character codes are translated to their corresponding tag nos. Boolean logic has been utilized for the purpose of constructing the search expression, according to the need of the user scientist/technologist. Suitable format has been designed for writing the search expression. The following operators have been employed for constructing the search expression.

1) 'AND' (&): This permits to select intersection of two concepts. A record satisfies the search if the terms or phrases connected by 'AND', all appear in one record.

2) 'OR' (+): Permits to select union of two concepts. It allows more than one query to be searched in a single expression. No relationship among the term or phrases connected by 'OR' operator is required. A record satisfies the search if, at least one of the phrases or terms connected by 'OR' appears in the record.

3) 'AND NOT' (\): This means AND followed by NOT. NOT alone is not needed in SDI type of activities because if we do not want to search any term we can simply ignore to mention it. A record satisfies the search if the term or phrase connected by 'AND NOT' does not appear in the record.

4) $ (Truncation): We can have any amount of truncation just by putting $ at the beginning or end of terms.

For illustration a sample search expression is written below

TI MICROPROCESSOR & TI CONTROL\ AT OJHA
+ TI MICROPROCESSOR & TI APPLIC $\ AT DK $
ABBREVIATION - 1
QUERY: TI INFORMATION & TEC. - TI CONTROL + TI MICROPROCESSOR & TI. ENGINE

TITLE: MICROPROCESSOR APPLICATIONS IN ELECTRICAL ENGINEERING, A BIBLIOGRAPHY 1977-1978
ABSTRACT: ELECTRICAL ENGINEERING PROVIDES SCOPE FOR THE USE OF MICROPROCESSORS IN ALL ASPECTS OF CONTROL. IN POWER SYSTEMS, INSTRUMENTATION, MEASUREMENT, TEST EQUIPMENT AND DESIGN, IN COMMUNICATION SYSTEMS, ESPECIALLY, THEY CAN PERFORM A WIDE RANGE OF MONITORING AND CONTROL FUNCTIONS WITH GREAT SAVINGS OVER TRADITIONAL METHODS. THIS BIBLIOGRAPHY BRINGS TOGETHER ABOUT 150 REFERENCES TO MICROPROCESSOR APPLICATIONS IN ELECTRICAL ENGINEERING. THEY WERE TAKEN FROM THE INSPEC DATABASE FOR THE PERIOD 1977 TO 1978.

AUTHOR: MAKAZU, Y.
JOURNAL: J. SOC. INSTRUM. AND CONTROL ENG. JAPAN
VOL. ISSUE: VOL.19, NO.5

TITLE: TIPSY TECHNICAL INFORMATION PROCESSING SYSTEM
AUTHOR: IWABUCHI, Y.
JOURNAL: J. SOC. INSTRUM. AND CONTROL ENG. JAPAN
VOL. ISSUE: VOL.19, NO.5

TITLE: MICROPROCESSOR APPLICATIONS IN THE NEEDS INFORMATION ADAPTIVE SYSTEM
JOURNAL: 12TH ANNUAL SOUTHEASTERN SYMPOSIUM ON SYSTEM THEORY
AUTHOR: HOWLE, W.M., JR.; KELLY, W.M.; IVSHMERIDITH, B.D.

FEED BACK MESSAGE 1-------
1 IS THIS REFERENCE USEFUL... YES... NO... CAN'T TELL...... COMMENT
2 IS THIS REFERENCE USEFUL... YES... NO... CAN'T TELL...... COMMENT