A HYBRID INTERACTIVE SEARCH SYSTEM (UN SYSTEME HYBRIDE INTERDEPENDANT DE LA RECHERCHE DOCUMENTAIRE)

John D. Currie
Digital Methods Limited
1716 Woodward Drive
Ottawa, Ontario K2C OP8

ABSTRACT

Independent storage and retrieval systems based on either micrographic or computer techniques have been extensively used. A different approach is to assemble a flexible hybrid system in which bulk storage is in microform, while dictionaries used in searching are stored in a computerized data bank. The economics, response time and search methods of the Defence Scientific Information Service (DSIS) retrospective search system are examined. (Les systèmes de stockage indépendants et de recherche documentaire basés sur la micrographie ou sur les techniques de l'ordinateur ont été très employés. Un autre point de vue est de rassembler un système hybride flexible dont le stockage consiste en microcopies tandis que les questions formulées dans la recherche documentaire sont rassemblées dans une banque de données. L'économie, la vitesse de réponse et les méthodes de la recherche documentaire du Service d'Information Scientifique pour la Défense (SISD) sont ici examinées.)

INTRODUCTION

Independent storage and retrieval systems based on either micrographic or computer techniques have been successfully used for some time. This paper examines the merits of a flexible hybrid system that is being developed in the Defence Scientific Information Service (DSIS) of the Defence Research Board (DRB).

Since 1946 DSIS has been responsible for collecting and cataloguing Canadian and foreign scientific and technical material of defence origin and disseminating this material to those who participate in Canadian defence and development. Much of this material is of a classified nature.

Since 1968 DSIS has recorded document accessioning data in machine readable format. These data are then used to generate various types of announcement, including KWOC, Digests and SDI and are also used to produce conventional type catalogue cards.

BACKGROUND

Retrospective searches until recently have been conducted using manual methods primarily centered around catalogue cards. With the availability of machine readable data retrospective searching can be automated.

The total DSIS document collection includes approximately 350,000 titles of which only 40,000 are in machine readable form, i.e. accessions since 1968.

The costs of entering data in the same readable form, i.e. full citation and abstract have precluded adding older documents (pre 1969) to the computer files. However, to improve searching efficiency, to reduce storage requirements and to allow the older document collection to be more accessible to DSIS customers it was decided to microfilm all citations (catalogue cards) relating to the older documents. This has allowed the older material to be subsequently included in the computer files.

Microfilming has been completed for some files (subject) and involved re-ordering of the cards to suit current DSIS indexing practice. A further film ordered by accession number has also been filmed.

Data from the subject films has allowed DSIS to provide machine retrospective searching for its total collection. For the older documents (pre 1969) the accession numbers associated with a particular subject heading have been manually read from the subject ordered film and input to a computer file. Thus the total collection can now be searched by subject and more recent documents by other surrogates. A computer search can be used to output actual document citations for recent accessions and a microfilm cross reference for older material.

The original plan for the machine readable data was to store the complete bibliographic data on files with controlled access to users via secure telephone lines. Because of the current unavailability of a secure communications network and of Department of National Defence computing facilities which could provide the type of interactive query service we require, an interim hybrid system has been developed. An unclassified data base containing the data is stored in a public computer in Toronto. Each user terminal of the system has a microfilm reader associated with it, together with a set of microfilm cartridges containing catalogue card images of all accessions in the DSIS document collection. The images are in accession number order. The results of a computer search indicate the frame numbers at which the full bibliographic data is stored on the film. Viewing the abstract allows the user to determine whether he wishes to see the document.

SYSTEM ORGANIZATION

Use of this system provides almost the same search capability previously available only at DSIS, Ottawa. The use of the microfilm allows citations associated with classified material to be made available to the end user even though the index system employs a public utility and normal public communications facilities (voice-grade telephone).

SYSTEM ORGANIZATION

The DSIS retrospective search system consists of three sub-systems:

(a) The on-line terminal connected via a telephone line to the computer based index material.

(b) The microfilm sub-system consisting of a viewer and microfilm cartridges which are referenced by the computer data.

(c) The actual documents which will in many instances be available in microfiche form in a location convenient to the on-line terminal.

Terminal and Computer Sub-System

This sub-system consists of three major components: the terminal; the computer system and associated programs; and the data base.

Terminals. Currently the Bell VUCOM I and the IBM 2741 terminals are used in the DSIS system. The VUCOM I terminal consists of a keyboard attached to a Cathode Ray tube (CRT) device which displays the keyboard entries and the computer responses. An optional printer is available which can be used to print out (hard copy) of selected items which are also displayed on the CRT. The IBM 2741 terminal uses a standard IBM Selectric typewriter. Keyboard entries and computer responses are recorded on normal fan-fold continuous stationery. The VUCOM system operates at a higher speed (30 characters/sec versus 15 for the 2741) and is noiseless. Its major disadvantage is limited screen size. This can be overcome by using the printer, but at increased cost.

Computer Systems and Software. The DSIS system uses the computing system available from IP Sharp Ltd. Their computer is located in Toronto and can be accessed via the public telephone systems from specific locations across Canada and the USA with IP Sharp absorbing the long distance telephone charges from these cities. (Vancouver, Calgary, Toronto, Montreal, Quebec, Halifax, Washington D.C., etc.)

SYSTEM ORGANIZATION

The IP Sharp system employs an IBM computer, Model 370/145 which includes extensive on-line storage. The computer system operates in a timesharing mode and provides a service known as 'APL-PLUS'. This service allows the high level programming language APL to be used simultaneously by many users. Each user is serviced so rapidly by the system that under most operating conditions the system appears to be dedicated to the individual user.

The DSIS retrospective search system uses the APL-PLUS system at the same time other users are carrying out their own computations. In addition, many users may be using the retrospective search system at the same time, each user having his own copy of the programs associated with the system, but all sharing the same data base. (Index material). When a user signs onto a system he is given a copy of the programs in an area of computer storage known as workspace. The workspace and the copy of the programs are then dedicated to his use all time he is signed onto the system.

The search system can be used by library clerical staff or scientists with a minimum of training since the query language involves simple English questions and answers. The system is also being developed in a bilingual version. Various levels of user expertise in the system are accommodated by providing a conversational (prompted) mode and a command (unprompted) mode.

The retrospective search system has been designed and developed by Digital Methods Limited for DSIS.

Data Base. The data base used in the retrospective search system consists of a series of files which are used by the system to select documents which meet the requirements of the user's search questions. The files are organized in a similar manner to the conventional library catalogues and allow searches to be conducted alphabetically by such terms as personal author, corporate author, descriptors and keywords, and COSATI code.

Data relating to documents included in the system are organized in the form of indices. These indices are searched separately, e.g. subject or author, but the results of an author search can be logically combined with the results of a subject search to find documents written by a particular author on a particular subject. The logical functions AND, NOT and OR are available.

The results of a search on the data base are displayed to the user in stages, first he is told how many documents have been found which meet his requirements, e.g. a certain subject or author. The actual data base in the system relating to the documents can then be requested.

Thus by refining the search the user can limit the final results to a small set of documents which meet his particular requirements.

The actual data about a particular document which is displayed to the user is limited to the title and various cross references. This approach has been taken to allow all DSIS material including classified documents to be available from the system. The cross reference material relates to:

(a) The original DSIS accession number.

(b) A microfilm number which allows the full catalogue and indexing material to be selected from the microfilm.

(c) A microfiche number which will eventually allow the document microfiche to be selected from the library file.

Microfilm Sub-System

Microfilm Cartridges. Cartridges provided by DSIS for the retrospective search system have been produced by filming existing catalogue cards for the years 1946 to 1968 inclusive and for the years 1969 to date by computer output microfilm (COM). In both the manually produced film and the 'COM' there is a small square 'blip' encoded opposite each frame of film for ease of retrieval. The cross references provided by the computer system identifies the cartridge number, the frame number, and the accession number of the document.

Microfilm Reader. In order to take full advantage of the computer indexing and the film format in the microfilm cartridge it is essential to have a good reader-printer. DSIS use a page search reader-printer which functions by counting the blips encoded on the film. If the proper cartridge is inserted in the frame and the correct frame number is keyed in the reader automatically stops at the selected frame so that the desired accession number as specified by the computer cross reference data is available for viewing. Retrieval of a particular document catalogue card image including loading the cartridge is of the order of seconds.

Microfiche Sub-System

The ideal microfiche support for the DSIS retrospective search system is a microfiche library of all documents accessioned by DSIS at locations of customers that have terminals. As an interim goal master microfiche of all documents accessioned since 1969 are being made available at DSIS. Duplicate copies for customer's requests will be provided. At a later stage complete or partial sets of these fiche may be provided at the customers terminals.

SYSTEMS COSTS

This time-sharing system provides the user with a very powerful system at low cost since the user's cost in addition to terminal rental and data base storage reflect only the actual individual use made of the computer processing time. The costs associated with the individual user are made up of two components:

(a) The time connected to the system.

(b) The cost associated with the actual computing time, i.e. the total of all the small increments of processing time during which the user actually has the computer working for him.

Service Charges. These are:

connect time	- 20 cents per minute
computer time (CPU)	- 30 cents per second
terminal rental (VUCOM + printer + acoustic coupler)	- $200 per month
data base storage	- $20,000 per year

Assuming three searches per day at each of 10 terminals with the average search being fifteen minutes connect time and fifteen seconds computer time. The estimate cost per search is:

connect and CPU time	- $7.50
terminal rental and data base storage	- $5.50
Total cost per search	$13.00

The estimates for search time are based upon our limited experience and will no doubt be revised.

The total cost of producing the microfilm used in the retrospective search system for DSIS and all Establishment libraries was approximately $7000. Microfiche readers vary in price from $1,500 to $5,500. We use the more expensive page search reader-printer, however this equipment was already in use in a stand alone microfilm catalogue system made available to our major customers. This system allowed manual search of microfilm after identifying relevant subjects in a hard copy index which provides reel and frame number references.

CONCLUSION

To include the bibliographic data in the computer file rather than on microfilm would more than triple the actual storage costs, i.e. $65,000 rather than $20,000 and in addition would substantially increase the computer costs for each search. To this must also be added to cost of encrypting equipment at both the computer and the remote user. (Obviously a secure classified computer installation would also be required.)

CONCLUSION

What we have achieved is not our ultimate goal, but it has permitted us to provide an improved service to our customers in a relatively short time frame at reasonable cost and at the same time overcome problems of security that could have delayed our system for an unacceptable length of time. We are, in effect, using the media to their best advantage: the computer for the manipulation of information and the microfilm for storage of information.

REFERENCES

AMEY, G.X., 1971, "Use of Simulation in Design of an Online Information Retrieval System." IEEC Conference Digest, October 1971, p. 60-61.

_____, "Evolutionary Transition to Automation of Information Management." The Fifth Annual Colloguium on Information Retrieval: Ramey J.A., Information Interscience Incorporated, p. 63-70.

_____, 1969, "A National Information Utility for Communication of Scientific Information." IEEC Pre-Conference Digest, October 1969, p. 108-109.

_____, 1968, "SOCRATES - An Information System for the Canadian Defence Community." Proceedings of Session on Information Retrieval at the 1968 Defence Research Board Symposium.

CURRIE, J.D., 1972, "Microforms in DSIS: An Information Center's Experience with Microforms." DSIS Technical Note No. 9, July 1972.