ALBERTA COMPUTERIZED ENQUIRY SYSTEM
A FEASIBILITY STUDY

W.J. Neilson, Research Officer
V.S. Townsend, Research Officer
Research Council of Alberta
11315 87 Avenue
Edmonton, Alberta
T6G 2C2

ABSTRACT

A survey has been conducted to determine the state of the art of information science in Alberta, particularly with respect to computerized retrieval. The technical and educational skills necessary for effective online retrieval are analyzed. The level of end user satisfaction with online searching in the scientific community is high. A comparative analysis of host systems indicates that U.S. based retrieval systems are dominating the marketplace, due to superior service and a wider range of databases. An upward trend in usage of non-bibliographic databases is noted. The feasibility of two research projects is evaluated: the development of a command language and the design of a specialized terminal for standardized bibliographic information retrieval.

SYSTEME D'ENQUETE PAR ORDINATEUR EN ALBERTA
ETUDE DE SON APPLICATION

RESUME

Une étude a été éffectuée afin de déterminer la situation dans laquelle se trouve l'art de la Science de l'Information en Alberta, plus précisément les développements au niveau des relevés de document par ordinateur. Les abilités techniques et pédagogiques nécessaire pour le développement d'un tel système sont analysées. Les recherches actives présentement dans la communauté et la satisfaction des opérateurs sont élevées. Une analyse entreprise afin de comparer les circuits indique que les Etats-Unis domine le marché étant donné que leur système profite de service supérieur à ce qui est disponible au Canada. Nous avons remarqué une tendance envers l'utilisation de centres de données non-bibliographiques. Nous envisageons la possibilité de fusionner deux projets de recherche; le développement d'un vocabulaire principal et la piannification d'un centre spécialisé en système de relevé d'information bibliographique.

Alberta Research Council Contribution No. 959

ALBERTA COMPUTERIZED ENQUIRY SYSTEM

INTRODUCTION

The Alberta Oil Sands Supplementary Information Service (AOSSIS), as a part of the Research Council of Alberta, has a mandate to investigate methods for improving access to information, especially through the use of computers. Funding for the project is provided by the Alberta Oil Sands Technology and Research Authority (AOSTRA).

Computerized literature searching can be done through a number of information vendors, or 'hosts', that offer services on a variety of topics. Although each database covers a specific subject area, there exists a certain degree of overlap between databases. Even so, some subject areas may be unique to a particular database. More often than not, a search question spans a number of disciplines, creating the need to search several databases, possibly supplied by different vendors.

Three fundamental problems arise.

1. The type and number of data elements varies among databases, requiring the user to remember or have access to the database description.

2. Indexing and classification techniques are incompatible between databases.

3. Searching command languages change with each vendor.

These problems are largely semantic in nature, and have arisen due to a lack of standards for information retrieval systems. As information retrieval is a relatively new discipline, its growth has been diverse. Figure 1 illustrates the present situation.

The concept of a standard language for information retrieval has been proposed as a first step toward standardization. In such a configuration, the user needs to know only one retrieval language and one bibliographic data structure. A translating interface controls user and host communications. Figure 2 illustrates the concept of a common command language for accessing online databases.

To evaluate the need, from a user viewpoint, for a common enquiry language, a survey was conducted. The findings are presented in this paper.

SURVEY OBJECTIVES

Oil sands researchers and information specialists were surveyed to evaluate the usage and effectiveness of bibliographic information retrieval services as a tool in research, and to gauge reaction to some specific proposals.

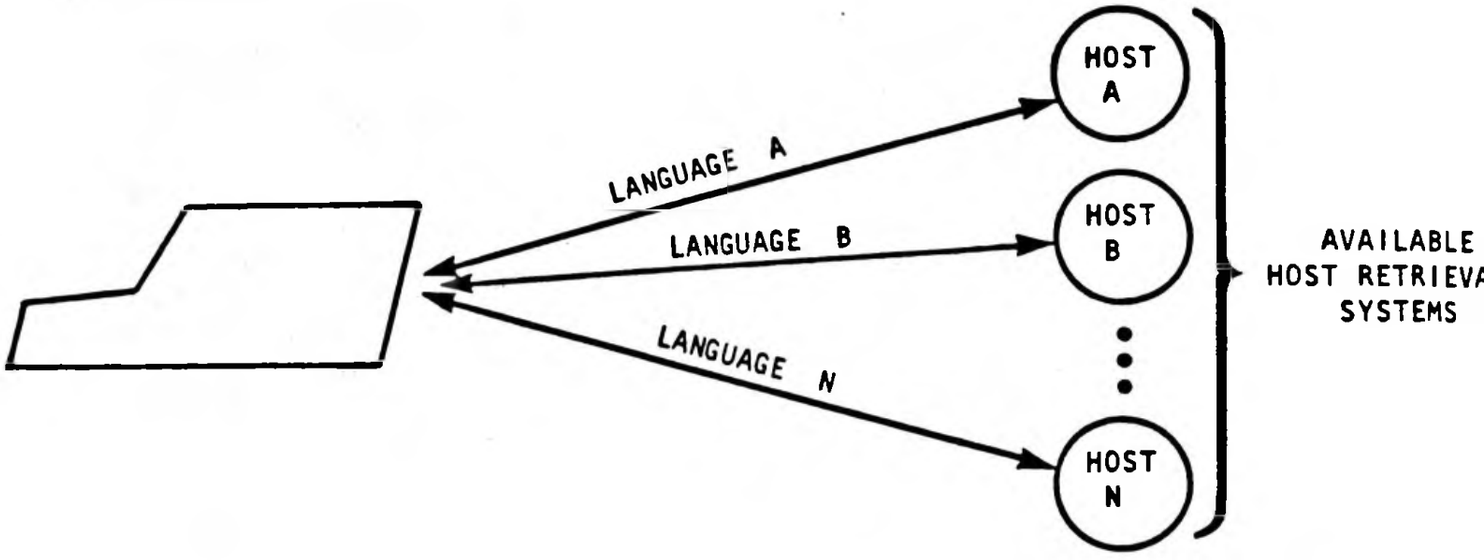

FIGURE 1: THE PROBLEM OF PROLIFERATING RETRIEVAL SYSTEMS

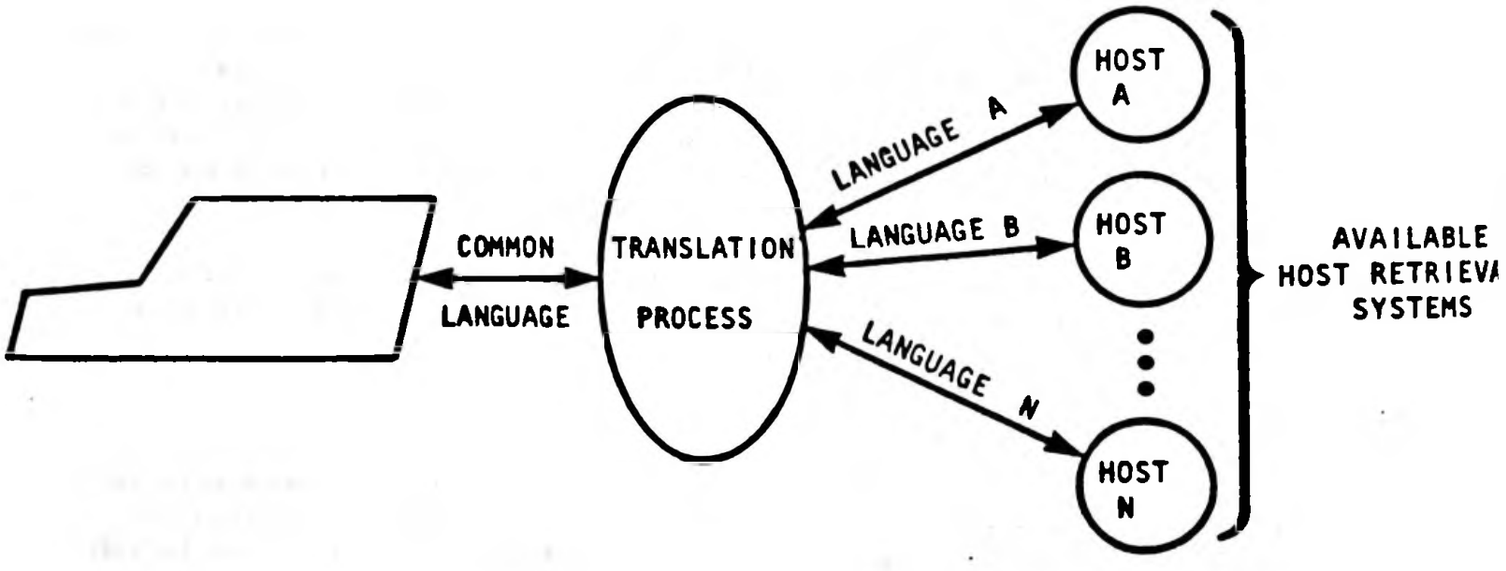

FIGURE 2: USER COMMUNICATION WITH COMMON COMMAND LANGUAGE

The objectives of the survey were to determine:

1. the range of current information-gathering techniques employed by researchers and information specialists; the awareness of online retrieval systems;

2. the amount of online searching being carried out, the confidence placed in it as a resource tool, and its effectiveness when compared to traditional information-gathering methods;

3. improvements necessary to overcome limitations of existing systems;

4. the biases for or against existing information retrieval systems and/or databases;

5. the role of information retrieval as it relates to the overall computing needs of the scientific community;

6. the need for and desirability of a standardized language for online information retrieval; the minimum requirements for such a language;

7. the demand for and marketability of a specially designed terminal for information retrieval; the desired capabilities of such a terminal.

The above formed the basis of the questionnaire used in the survey.

SURVEY GUIDELINES

During the past five years there has been a dramatic increase in the number of databases available online, the subject areas covered and the range of disciplines searchable. To limit the sample size of people to interview, regional and subject area constraints were defined.

Subject Area Constraints

Since project funding is provided through AOSTRA, the survey was aimed towards people involved in areas related to oil sands technology. Two classes were considered:

1. personnel involved in oil sands research, and

2. information specialists, primarily those who provide services for the first class of people.

Regional Constraints

The survey contains a regional bias as the primary consideration was with information retrieval within Alberta. A justification is the

law of diminishing returns. The time and money spent elsewhere would not justify the additional knowledge that would be gleaned. Regional biases diminish in importance with sophisticated usage of telecommunications network facilities.

SURVEY FINDINGS

General Remarks

Few 'end users' (the persons requesting the literature searches) did their own online searching. There appear to be two reasons.

1. Researchers have relatively infrequent needs for such a service. Most are unwilling to commit the amount of time required to gain and maintain proficiency in online searching, especially when there are qualified personnel available to do the searching for them.

2. The second reason is less obvious: the professionals surveyed indicated a general antipathy for online searching. Bibliographic information retrieval is regarded as a technical support task.

Researchers exhibited a markedly greater interest in the interactive use of numerical databases, such as physical properties, experimental data, field test data.

AOSTRA funding has stimulated the majority of the oil sands research done in Alberta by the university and government sectors. Although the original intent of the survey was to interview only personnel involved in oil sands research, most oil companies conduct their research in their laboratories in the United States. Oil company representation in the survey consisted mostly of oil sands or heavy oil production-oriented personnel.

Many of the information specialists interviewed search all disciplines of science and technology. The most valuable comments were obtained from this group as they have a greater appreciation of the bibliographic retrieval process than do the end users.

User Qualifications

Subject Knowledge. The single most important asset for effective searching is knowledge of the subject area, gained either through formal education and/or work experience or searching experience. In any case, one must become familiar with the subject indexing strategies employed in each database. Online index and thesaurus facilities are inadequate for effectively guiding the user through an unfamiliar subject area.

Technical Skills. Subject knowledge, although important to effective searching, is not enough: knowledge of the technical skills of online searching is necessary. Presently, the amount of technical knowledge required prohibits people with casual interests or infrequent

needs from carrying out their own searches, regardless of their mastery of the subject terminology. Keeping current with system changes and continual usage is vital to efficient searching. A training session and/or a set of comprehensive manuals are essential.

The ability to search is an acquired skill. Perusal of vendors' newsletters and other literature is required to maintain a high level of ability. Naive users are forced to make use of an intermediary resource person, which does not speak well for the tutorial or instructive capabilities of present systems.

Library Skills. Most of the information specialists were connected with library services of some kind. They felt that formal library school education could not be equated with online searching performance. Such training, however, does provide an appreciation for the value of other reference tools, such as conventional printed indexes.

User Satisfaction

End users expressed a high level of satisfaction with online search results. Most subject areas are well covered by the available databases (several exceptions are outlined later). The speed of computerized retrieval seems satisfactory.

Online searching enables information specialists to provide a more comprehensive service than would be otherwise possible. Computerized information retrieval has assumed a productive role in the realm of information services and is here to stay.

Host Systems Evaluation

Bibliographic Information Services. The choice of host system is largely determined by the user's area of interest, which in this survey was primarily petroleum production and related disciplines. The two host systems used most frequently are ORBIT IV from System Development Corporation (SDC) and DIALOG from Lockheed Information Systems (LIS). Both systems offer a large number of databases, high system reliability, and sophisticated search logic. The Petroleum Abstracts and American Petroleum Institute databases are available from SDC.

Usage of Canadian hosts systems has declined relative to U.S.-based systems (FitzPatrick and Townsend, 1978). CAN/OLE, the National Research Council's online system, was criticized for its database coverage, poor system reliability and minimal search logic. QL Systems Ltd. of Kingston, Ontario, was attacked along different lines -- its search logic is awkward and redundant. The SPIRES database management system, offered at the University of Alberta, was recognized as a potentially useful system. However, it needs more vigorous advertising to make the user community aware of its capabilities.

The usage of other bibliographic retrieval systems is low enough to warrant little or no comment.

Non-Bibliographic Information Services. A sharp trend has developed towards usage of retrieval systems for non-bibliographic data such as numerical data and newspaper items. This includes the Dow Jones database, the Statistics Canada (CANSIM) databases, and the New York Times and Toronto Globe and Mail databases. Usage of these databases is often by management, where decisions depend on the most current information available.

Information Deficiencies

Subjects Poorly Covered. The greatest information deficiency is in the area of government publications at all levels. Although many information specialists are aware of the GAP database, very few are using it, the main reason being that the database has not been up to date. At the time of writing this paper, the database is complete.

There are several specific subject areas which were mentioned as having poor online coverage. They are listed below:

1. corrosion -- the printed index for Corrosion Abstracts should be online.

2. metallurgy

3. consumerism

4. textiles -- the Textile Technology Digest should be online.

5. plastics -- the British Rubber and Plastics Research Association printed index should be online.

6. Canadian arts

Many of the complaints listed may stem not from a lack of online coverage, but rather from a lack of knowledge about where to look for the desired information. The COIN (Computerized Information) Directory, an online inventory of databases, indicates that as many as 13 databases cover one of the above topics.

Non-Comprehensive Databases. As most search questions are interdisciplinary in nature, several databases and/or hosts may be searched. The 'save search' feature of SDC and Lockheed can be used to transfer a search to a different database, saving the user from retyping the search logic. There is still a problem, however, when

1. the databases have different thesauri. Search terms must be tailored to the individual database semantic interpretations.

2. the databases exist on more than one host system. At present there is no means of transferring a search from one system to another.

Numerical Databases. Many researchers indicated that they would like to see the development of more numerical databases pertaining to oil sands. At present, there is no comprehensive body of information concerning physical properties of oil sands that is publicly available.

Reference Duplication. If search results are obtained from several sources, there may be duplicate references. Identification of duplicate references for merging search results is difficult because there is no standard bibliographic data structure.

Desirability of a Common Command Language

The usefulness of a common search language has been expressed elsewhere (Atherton, 1978). The feasibility of implementation through an interpreter approach has been established in separate studies (Marcus, 1975, Marcus and Reintjes, 1977, Rosenthal, 1978).

The idea of a common command language found general, although not universal, approval. The exceptions were:

1. researchers who felt it was unlikely they would ever do their own online searching.

2. information specialists who either concentrate on using only one or two retrieval systems, or who are extremely knowledgable about all the systems they use. In fact, they fear a degradation of performance if the command language cannot adequately support all the nuances of existing systems.

A common command language is most beneficial to searchers with infrequent needs, or those who must keep up to date with many host systems. This group comprises the majority of the online users, and includes an even larger 'silent majority' of potential users who are intimidated by computer usage in general.

Desirability of Specialized Terminals

The terminals presently used by most searchers are hard copy (paper printing) units operating at 300 baud (30 characters per second). Users are satisfied with their terminal equipment, but would upgrade to higher transmission speeds. Cathode ray tube (CRT) display units would be considered only if there was an attached printer.

A specially designed terminal for information retrieval must be multipurpose, including functions such as APL programming language, automated library routines, etc. The marketability of the hardware depends on cost and time savings benefits. User reaction to terminal

concepts may change with declining hardware costs and increased user awareness.

Communications Networks

All online users surveyed make use of at least one data communications network. Most use the Datapac packet switching network; some use Tymnet and Telenet as well. Relatively few problems were cited with communications networks, and users were unanimous in endorsing the cost benefits of networking. Few users were aware, however, of the potentials of network facilities, other than for information retrieval.

The Role of the Alberta Research Council

The Alberta Research Council has maintained a high profile in information retrieval since the formation in 1967 of the Alberta Information Retrieval Association (AIRA). The Council has promoted and developed online services by providing a search service, and by cooperatively constructing databases such as Government of Alberta Publications (GAP), Solar and Wind Energy Research Program (SWERP) and Alberta Oil Sands Index (AOSI). The Research Council has also played an important role in developing networking facilities in the province through an AOSTRA funded project. The most tangible contribution was the purchase of a PDP 11/40 minicomputer to interface to Datapac. Software development and maintenance is provided by the University of Alberta.

As a result of the efforts of the Research Council, information specialists in the province expect the Council to provide leadership in the area of online searching.

CONCLUSIONS

Common Command Language - A Viable Project

From the interviews and related research publications, further development of a common command language for information retrieval is a viable project. The goal of reducing the amount of searching skills required for effective online searching is a realistic one. It may be possible to bring the use of retrieval systems within the grasp of casual end users.

A set of minimum requirements can be outlined for such a language.

1. It must be simple to use, and must possess a high degree of stability.

2. Differences in communications access and host commands should be transparent to the user.

3. Boolean operators are required; the ability to do combinatorial logic in a single command is desirable.

4. Bibliographic record formats should be standardized as much as possible. There should also be a range of output formats available.

5. Instructive or tutorial features should allow naive users to produce useful results while simultaneously improving their online skills.

6. Users should be given the option to enter commands either in the common language or the specific host language. Certain host language features may be too specific for inclusion in the common command language.

7. Search logic should be transferable between databases and/or host systems.

8. Search sets should be numbered and search set numbers should be usable entities in forming further queries.

9. A 'mail' facility should be provided for sending messages between users or between user and supplier.

10. Local offline printing of search results should be possible.

The Command Language in Perspective

The language developed by AOSSIS will standardize access to bibliographic information retrieval systems. A broader class is generalized database management systems, which offer a greater range of capabilities, such as numerical databases and an applications program interface. The common command language design should be flexible enough to allow extension to extraction and manipulation of numerical data.

Development Strategy

Development of the common command language will take place in two stages. First, the software will be developed on the University of Alberta's Amdahl computer as shown in Figure 3. The advantages of initial development on a timesharing system are lower capital costs, greater flexibility and broader access.

As a second stage, all or part of the technology will be transferred to a prototype terminal, as illustrated in Figure 4. The command language could then be used directly by specialized terminals, or by any terminal via the timesharing system.

This second phase of the project will provide a testbed for new hardware and software ideas. It will be useful for demonstration purposes, thus increasing public awareness. By the time the terminal is developed, component costs are bound to have decreased, improving the affordability of such a device. Even so, commercial production of the

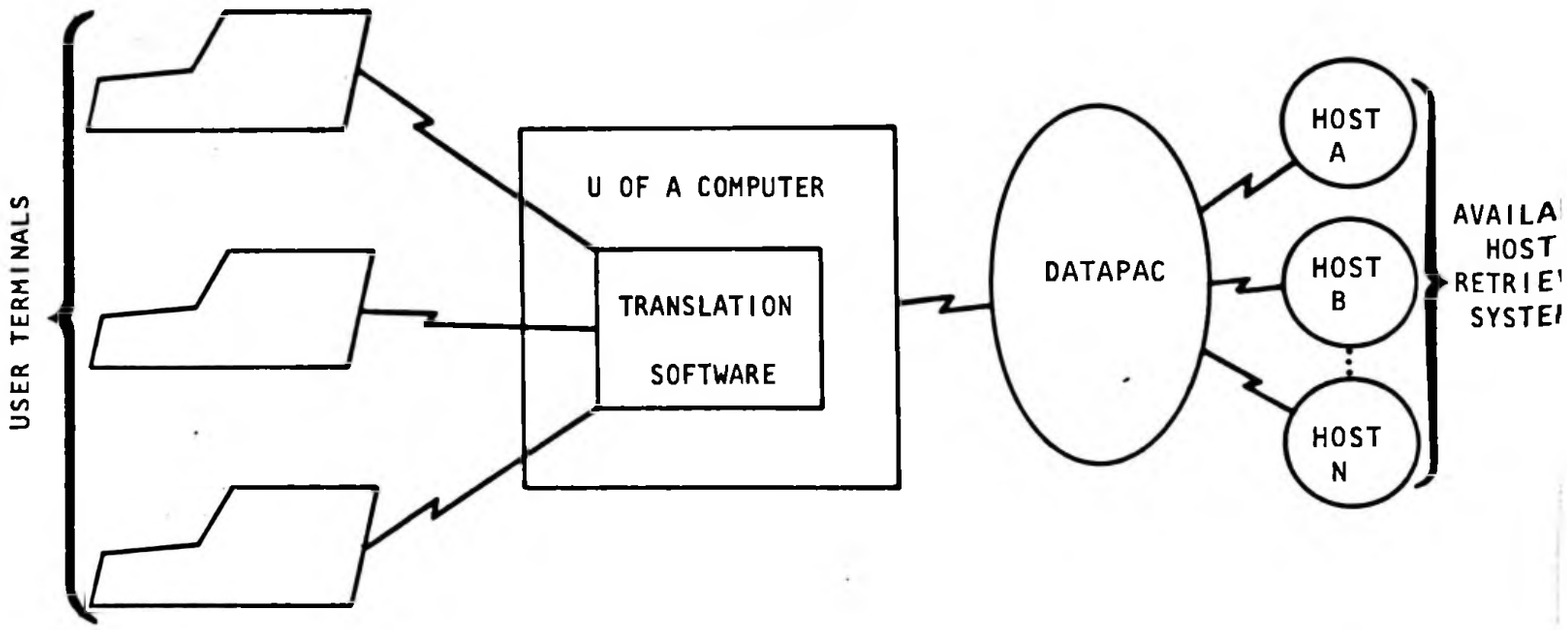

FIGURE 3: COMMAND LANGUAGE SOFTWARE ON U OF A COMPUTER

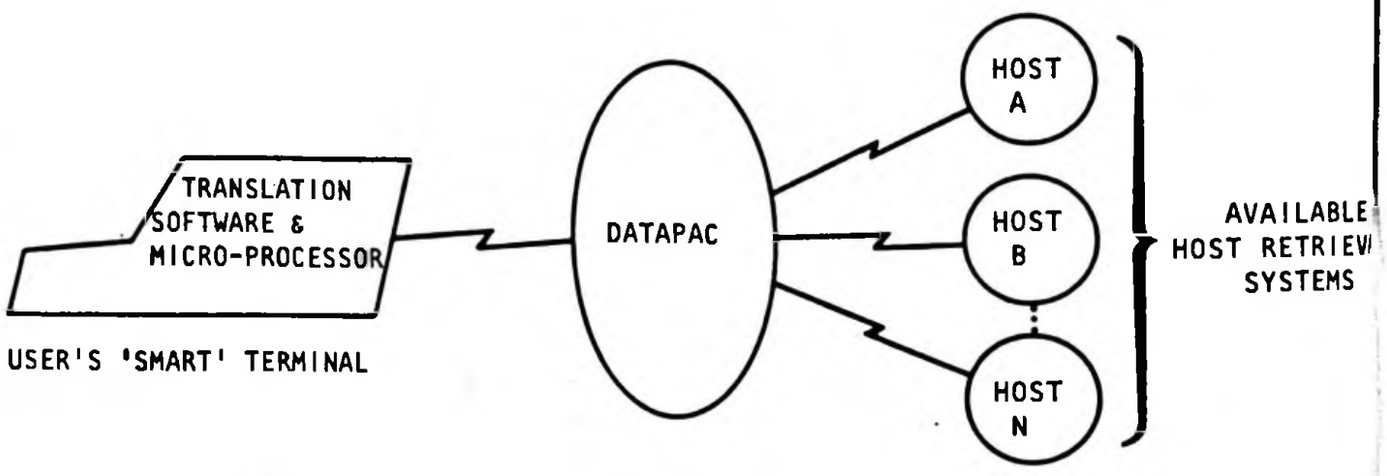

FIGURE 4: COMMAND LANGUAGE SOFTWARE IN A 'SMART' TERMINAL

terminal may not be economically feasible.

SUMMARY

The usage and effectiveness of bibliographic information retrieval services has been evaluated. Problems in online searching have been identified and suggested improvements presented.

User reaction to two specific project proposals has been noted:

1. a common command language for bibliographic information retrieval, and

2. a specially designed terminal for online searching.

Strategy for the development of the proposed projects has been formulated.

ACKNOWLEDGEMENT

The AOSSIS project is the result of a research grant from the Alberta Oil Sands Technology and Research Authority, whose support in the development of online information retrieval is gratefully acknowledged.

REFERENCES

Atherton, P., "Standards for a User System Interface Language in Online Retrieval Systems", Online Review, Volume 2, Number 1, March 1978.

Fitzgerald, M.L., "Common Command Language for File Manipulation and Network Job Execution: An Example", National Bureau of Standards Special Publication 500-37, August 1978.

FitzPatrick, A.B., and Townsend, V.S., "The Economics of Information Retrieval Services", European Computing Congress 1978, London, England, May 1978, pp. 689-705, Online Conferences Ltd.

Kimbleton, S.R., Wood, H.M., and Fitzgerald, M.L., "Network operating systems -- An implementation approach", AFIPS Conference Proceedings, Volume 47, 1978.

Marcus, R.S., "Networking Information Retrieval Systems Using Computer Interfaces", Proceedings of the 38th ASIS Annual Meeting, Volume 12, October 1975.

Marcus, R.S., and Reintjes, F.J., "Computer Interfaces for User Access to Heterogenous Information Retrieval Systems", NTIS Report PB 269 126, April 1977.

Negus, A.E., "EURONET Guideline: standard commands for retrieval systems", Final report on a study carried out for the Commission of the European Communities, DG XIII, December 1977.

Rosenthal, R., "A Review of Network Access Techniques with a Case Study: The Network Access Machine", National Bureau of Standards Technical Note 917, July 1976.

Rosenthal, R., and Lucas, B.D., "The Design and Implementation of the National Bureau of Standards' Network Access Machine (NAM)", National Bureau of Standards Special Publication 500-35, June 1978.

Townsend, V.S., "Design of ACE - A Computerized Enquiry: Preliminary Command Language Considerations", Internal document, April 1977.