

## SECTION V

### CONCLUSIONS AND RECOMMENDATIONS

KENNETH H. COOK

This section presents short, summary statements about the major conclusions of the work accomplished, followed by a discussion of recommendations. To provide the reader with information regarding a specific conclusion of particular interest to him, a reference to the applicable section(s) of this report is given following the conclusion.

The basic objectives for this research effort have been spelled out in Section I; it also includes a complete overview of the work accomplished. The different sections that follow Section I present detailed descriptions of some of the more important aspects of the project.

#### 1. MAJOR CONCLUSIONS

The major conclusions are:

##### a. Recall Improving Algorithms

Augmentation of free-text searching with algorithms used interactively, such as the vocabulary data base and the search data base, lead to improved performance (recall) levels, at lower computer cost. These improvements give the user the flexibility of retaining the advantages of completed free-text searching while still being able to elect the additional use of searching aids which improve the efficiency and effectiveness of the system. (Sections I and IV)

1. Use of the Vocabulary Data Base (VDB) to retrieve documents from the Document Data Base (DDB) yields better cost-performance levels than use of the DDB alone. (Section IV)

2. Use of the Search Data Base (SDB) to retrieve documents from the DDB yields better cost-performance levels than use of the VDB or use of the DDB alone. However, care must be taken in interpreting this finding as the data supporting this evidence was collected under controlled conditions which were less than optimal. (Section IV)

3. Efforts to improve the user's recall ability with an interactive retrieval system can possibly make use of the cumulative human intelligence of users of that system. Techniques such as these which are continually adaptive to the behavior of users present an alternative to a fixed, "one-shot" controlled vocabulary or indexing system.

4. One unanticipated finding was that significant differences in searching styles existed among individuals equally well trained in free-text retrieval. These differences appear in sizeably different cost-performance curves for individuals. Studies of these human differences may contribute

significantly to improved design of interactive languages and algorithms to improve recall. (Section IV)

b. Improving SUPARS/DPS Vocabulary Capacity

The capacity of the SUPARS/DPS vocabulary file can be made approximately 65,000 times as large as the original capacity through new coding techniques. (Section II)

1. The SUPARS/DPS vocabulary file was increased to a size of 4,294,967, 296 unique free-text terms from the previous limit of 65,534 by implementing programs using 32-bit coding for words, rather than the previously used 16-bit coding (Section II)

2. The increased vocabulary capacity allowed the construction of one data base holding 106,702 terms and 46,828 documents rather than having to build many data bases, each limited by the previous DPS single file vocabulary limit of 65,534 words per data base. (Section II)

c. Computer Programs Recording System Use, Cost, and Growth

Computer programs that unobtrusively monitor, collect, and store the user's searching interaction and important parameters of the system (cost, computer time, data, terminal used, etc.) have been successfully developed. These STATPAC programs have proven to be useful means of recording the use, growth, and cost of an interactive retrieval system. (Section II)

d. Publicity, Instruction and Education

The use of a strong publicity/public relations, instruction, and education package has resulted in gaining high user awareness, a continued positive attitude of users, and increased use of the free-text retrieval system without employing search intermediaries. (Section III)

e. Systematic Evaluation of User Attitudes and Reactions

Three techniques were found to be effective means of systematically evaluating user attitudes and reactions: (a) a Semantic Differential Attitude Scale, (b) a structured telephone interview and (c) records of user dissatisfaction, information inquiries, and reactions through a special telephone aid service for users.

f. Growth of the Document Data Base

The growth rate of the free-text document data base indicates that for the first 8,859 documents processed through SUPARS/DPS, an average of 3.86 new words per document were added to the vocabulary of unique free-text terms. By the time 46,828 documents from Psychological Abstracts were processed, the average number of unique words being added to the data base for the last 13,035 documents was reduced to 1.34. (Section II)

g. Cost Comparisons Between SUPARS I and SUPARS II

1. Using total document retrieval as a measure of cost, the best version of SUPARS II (e.g. the VDB as a searching algorithm) had a better cost-performance relation than all of the searching techniques used in SUPARS I to improve document retrieval effectiveness. (Section IV)

2. Using average dollar cost, a completed search inquiry in SUPARS II was approximately 14% less expensive than a comparable search inquiry in SUPARS I. This cost improvement is mainly due to the increased performance of the IBM 370/155 which was used to record cost figures for SUPARS II.

2. RECOMMENDATIONS MADE

a. Continue Development of User Control of Vocabulary and Synonyms

A need continues to exist in interactive free-text searching for some form of user vocabulary or synonym control. Two approaches can be taken. The first is to generate or obtain synonyms and incorporate them into the DPS Synonym List option that is available but has not been used. One way of obtaining synonym-like terms would be to generate strings from terms related in the Search Data Base. This technique would be adaptive to the changing body of terms employed by the user population and could grow as usage and users changed. The second approach would be to continue development of the Search Data Base and the Vocabulary Data Base in an interactive mode. The cost-effectiveness of the Search Data Base as a recall-improving technique could be judged more adequately if a greater number of search inquiries were included to cover a wider range of topics within the Document Data Base. The Vocabulary Data Base provides the best cost-effectiveness relationship for recall improvement and could be enhanced as a searching aid by continuing the development of options allowing users to view or print chosen segments of the list of terms.

b. Maintain a Non-Reentrant Search Module

Regarding the non-reentrant capacity of the current search module portion of the SUPARS/DPS retrieval system, recent experience with the 370/155 operating system indicates that the faster speed of this system would be about the same as a system with reentrant capabilities. In addition, a re-entrant search module with computer time sharing by many users might result in the same time delays as a non-reentrant module, with more core allocated.

c. Improve User Access to Hardware

Because of user dissatisfaction with having too many users vying for available processing space and not being able to gain access to the data base, either (1) increase the amount of computer core available to allow more than the current limit of eight users at a time searching, or (2) control the proportion of potential users to the available number of users who could search at one time. One possibility might be to establish an

input queue of users that would collect a search inquiry and channel it to the next available processing space. A queue of this type was begun during the current research but was discontinued because of programming difficulties.

d. Investigate Searching Styles and Techniques of Free-Text Users

Section IV, "Cost-Performance Study" has clearly indicated that wide differences exist in an individual's ability to efficiently and effectively search and retrieve documents. An important consideration in the improvement of interactive, free-text searching would be to begin research to determine these user characteristics. We suspect that knowledge of these characteristics may be as important in improving a retrieval system as the dollars expended in software/hardware improvements of existing facilities. This information about differences could be incorporated into existing procedures which were part of the interactive searching language and the user educational/training function.

e. Develop Interactive Algorithms Based on User Styles and Techniques

One way to make a system maximally adaptive to the differences among users is for system developers to incorporate in the processing and search sub-system some characteristics of these differences. For example, some possible ways of accomplishing this objective would be to (1) develop personalized synonym lists for high priority or high frequency users, (2) identify parameters of documents considered relevant by a user and use these parameters as a screening device for documents retrieved by him, or (3) display to the user the discriminant ability of search inquiry components through a term/document matrix available, as an option, after the processing of each inquiry.

## REFERENCES

1. Rome Air Development Center, Griffiss Air Force Base, New York.  
"Statement of Work for Free-Text Retrieval Evaluation." PR No.  
I-1-4683, 10 September 1970, p.3.
2. Large-Scale Information Processing Systems, Report of the SUPARS  
Project, Syracuse University School of Library Science, Syracuse,  
N.Y., July 1971. [5 Sections]
3. Large-Scale Information Processing Systems, Report of the SUPARS  
Project, Syracuse University School of Library Science, Syracuse,  
N.Y., July 1971, Section IV-B "User Component of the System", p.2.
4. King, Donald W., Neel, Peggy W., and Wood, Barbara L. Comparative  
Evaluation of the Retrieval Effectiveness of Descriptor and Free-Text  
Search Systems Using CIRCOL. Final Report of Westat Research Inc.  
to Foreign Technology Division, Wright-Patterson AFB, Ohio. 119 pages.  
Contract No. F30602-70-C-0205.
5. Paisley, William J., "Information Needs and Uses" In: Cuadra, Carlos  
(Ed.) Annual Review of Information Science and Technology. Chicago:  
Encyclopedia Britannica, Inc., 1968, p.4.
6. IBM System/360 Document Processing System (360A-CX-12X), Program  
description and operations manual. H20-0477-1.