

## SECTION I

### INTRODUCTION

Kenneth H. Cook

This introductory section includes (1) Previous Research Efforts, (2) Overview and Objectives of the Current Work, (3) Philosophy and Approach, (4) a major section on Overview of Work Accomplished: February 1971-January 1972, and (5) Summary. The overview of work accomplished does not intend to report results and conclusions but rather overviews the work accomplished and gives references to other sections which deal specifically with particular areas and topics.

#### 1. PREVIOUS RESEARCH EFFORTS

A previous research effort (July 1969-January 1971) by the SUPARS Research Group (Syracuse University Psychological Abstracts Retrieval Service), called SUPARS I, developed, implemented and evaluated a free-text reference retrieval system. The research group at the Syracuse University School of Library Science was contracted by the U.S. Air Force Systems Command, Rome Air Development Center to investigate a large-scale reference retrieval system in an on-line, interactive mode. One of the objectives of the work was to determine the reactions to such a system in a "real life" field environment, rather than a limited laboratory setting.

IBM's batch-mode Document Processing System (DPS) was extensively modified as the free-text processor to operate in an on-line interactive mode. A teleprocessing program was developed to allow the SUPARS/DPS package to interface with the university's 360/50 computer and 70 IBM 2741 communications terminals.

Machine readable versions of rented Psychological Abstracts (PA) tapes were translated, reformatted and then processed through DPS to form a data base of 35,000 document surrogates. The surrogates included bibliographic citations and abstracts for the January 1969-October 1970 issues of PA. A free service was thoroughly publicized and offered to the 17,000-member Syracuse University campus community for a three-month period in fall, 1970. The system operated four hours a day, Monday through Friday and provided users direct access to the data base, without any intermediaries, through use of the 2741 terminals. Users were provided with step-by-step examples of interaction in a User's Manual, a telephone aid service with experts available to discuss searching problems, and on-line demonstrations to classroom and other campus groups.

The system was evaluated through a multi-method approach, including the following techniques. (1) A cost-effectiveness study plotting the relationship between recall levels and computer processing cost of retrieved documents. Cost effectiveness curves were developed for various levels of search logic

complexity that were available in SUPARS/DPS. (2) A Semantic Differential attitude scale developed, implemented and evaluated at Syracuse for on-line retrieval systems, (3) In-depth interviews of the campus community, based on a stratified random selection of SUPARS users and nonusers, and (4) a descriptive analysis of the wide variety of users registered to use the system, including field of study, specific information requirements, and previous use of computerized systems.

The results of the SUPARS I project are available in a 6-part report of the work accomplished during July 1969-January 1971 under the title "Large-Scale Information Processing Systems" with each part sub-titled separately.

## 2. OBJECTIVES OF CURRENT WORK (SUPARS II)

The current effort, SUPARS II, is an extension and continuation of the work explained above and covers the period February 1, 1971, to January 31, 1972. The primary objective of the sponsoring agency was to "perform further experimentation with and evaluation of a large-scale on-line free-text retrieval system and to develop techniques for improving effectiveness." (1)

To meet the basic objectives, the requirements of the work to be performed called for:

- a. Developing recall-improving algorithms that could be added to the existing on-line program,
- b. Reducing the rate of vocabulary growth and increasing the vocabulary capacity of the Document Processing System, and
- c. Developing statistical data about system use to document the use, growth, and cost of the free-text retrieval system.

The "Overview of Work Accomplished" section beginning on page 4, is organized using these three major objectives as a framework for reporting the type of work done during the SUPARS II project.

The research work was accomplished by the SUPARS Research Group of the School of Library Science at Syracuse University. After an eight-month developmental period, the experimental free-text system was offered as a free bibliographic reference retrieval service to the students, staff and faculty of the 17,000-member campus. The period of experimental use began in November, 1971 and ended in mid-January, 1972, during which time it was available three to four hours a day, five days a week. During 1971, the SUPARS/DPS system ran under the control of the IBM 360/50 operating system and the 2314 disk pack facility. During January, SUPARS/DPS ran under a new IBM 370/155 operating system which replaced the 360. User access to the system was gained through the use of 100 IBM 2741 terminals using a standard APL typeface and keyboard and were located in most major buildings on campus.

The majority of programming involved in the modification of the basic Document Processing System was written in Basic Assembler Language.

### 3. PHILOSOPHY AND APPROACH

The philosophy of the current research is that the most promising area of improving free-text effectiveness is the human intelligence of the users and the by-products of free-text processing. This approach means giving users on-line access to the varied and complex search inquiries which have been constructed and submitted to the system. The approach also means recycling to the user the extensive vocabulary of terms generated by free-text processing, which would otherwise be stored internally and not be available to the user.

This basic philosophy was developed after consideration and recognition of alternative techniques and approaches to improving retrieval systems. These alternative approaches tend to emphasize intervention of the system designer through the use of statistical techniques for indexing, and lexographic decisions about the control of the terms allowed into the index.

The approach taken by the SUPARS II research group was that improvements to interactive, free-text searching should provide a user with various forms of cues, aid, and help during, before, and after the formulation of a search inquiry. This type of help recognized the advantage of, but did not emphasize, the use of on-line tutorials. Rather, the development of help stemmed from the notion that in an information center or library an individual tries to enlarge his strategy of information seeking through the help of the reference librarian. The reference librarian can generally suggest additional cues, sources, and alternative methods of searching -- aids which are rarely available in interactive searching of a retrieval system.

The idea behind the new algorithms developed in the SUPARS II study was to provide these cues in order to help prompt the on-line, interactive searcher during his negotiations with the computerized retrieval system. Because modern, interactive searching generally takes place at terminal locations remote from a reference librarian or any other human specialists in the user's interest area, one source of additional help was considered to be the human intelligence of all other users of the system. Because the intelligence of this user group is reflected in the search inquiries submitted to the system, techniques were developed to record, store, and reformat, the data for use by all other users. The section titled "Development of New Algorithms" on page 4 explains how this objective was reached.

Another underlying idea in the research reported here was that the development of search inquiries in free-text searching rests almost entirely on the primary searcher -- the individual with a specific information requirement statement -- and his ability to specify the inquiry search words and combinations, and the documents considered relevant to his requirement statement.

For a free-text retrieval system to remain adaptable and viable to a variety of user groups, the user must be given the maximum opportunity to develop his own unique and personal search inquiries that adequately reflect his requirements. The system built at Syracuse was developed for use by the

primary users -- the individual having his own specific information requirements -- and specifically did not provide intermediaries.

For future free-text retrieval systems of this type to be attractive to a wide range of users, the formulation of the inquiry will have to be easily accomplished by even the most unsophisticated, first-time user.

Whether an intermediary can effectively take the place of the primary searcher is still an empirical question. The question raised by the current research is how effectively the user can take advantage of new free-text, on-line algorithms to improve his recall level while maintaining some acceptable degree of precision.

To investigate this question, three main areas of work were performed, including: (1) development of new searching algorithms, (2) modification of free-text vocabulary storage, and (3) documentation, through a variety of techniques, of the use, growth, and cost of SUPARS/DPS. These three work areas parallel the basic objectives stated on page 2 and are used as an outline for a description of the work accomplished.

This overview provides a broad general review of the work accomplished, while the specific procedures, conclusions, and results are explained in more depth in other referenced sections following Section I.

#### 4. OVERVIEW OF WORK ACCOMPLISHED: FEBRUARY 1971-JANUARY 1972

This overview is organized into three parts, following the three main objectives: (1) Development of recall improving algorithms that could be added to the existing on-line program, (2) reducing the rate of vocabulary growth and increasing the vocabulary capacity of the Document Processing System, and (3) developing statistical data about system use to document the use, growth, and cost of the free-text retrieval system.

##### a. Development of New Searching Algorithms

The SUPARS/DPS searching capability developed during SUPARS I (July 1969-January 1971) allowed a user to interactively submit free-text words as input and retrieve document surrogates of bibliographic citations and abstracts.

In addition to the ability to search and retrieve documents from a document data base, two complementary algorithms were developed during the current SUPARS II in an effort to improve the user's ability to formulate free-text searches and to improve recall at an acceptable level of precision. The first idea was to store and make retrievable interactively the previous search inquiries of all users for on-line retrieval. The second idea was to make the free-text vocabulary terms available interactively along with the document postings. These two ideas are discussed below under (1) The Search Data Base, and (2) The Vocabulary Data Base.

One of the difficulties with developing and constructing free-text



searches is the demand placed on the user to know and use a variety of synonyms related to his specific interest area in an inquiry. Ideally, a user would be able to talk with someone familiar with his interest area and be provided with a variety of words and other cues. The user could then develop or formulate a search inquiry to the system that would have the specificity or exhaustiveness needed to maximize retrieval. Although some specialists are able to come up with most potential search words, most users don't have the opportunity for an expert to be immediately available for consultation.

To be responsive to the needs of an on-line user with no intermediary, a free-text system has to augment the intellect of each user and provide synonym control and other cues to help him develop effective searches. The concept of computer augmentation of the human intellect is not new, but has not yet been fully applied to the improvement of on-line reference retrieval systems.

The new algorithms developed during the current research effort are attempts to extend the search formulation capabilities of the free-text searcher. The two new algorithms discussed below are separately searchable data bases of (1) the previous search inquiries of all users -- the search data base -- and (2) the entire range of free-text vocabulary terms -- the vocabulary data base. These two new ideas were provided to the user in addition to the primary data base of bibliographic citations and abstracts -- the document data base.

#### (1) The Search Data Base

The search data base was developed by taking as input the previous search inquiries that users had submitted to the system when searching for documents. After reformatting this data, the searches were processed through the SUPARS/DPS programs and an inverted file of search words and documents were formed. However, the "documents" were not bibliographic citations and abstracts, but the "text" of the previously used searches. A detailed description of the development of the search data is contained in Section II.

Users were allowed to submit search inquiries to the search data base in the same form as the document data base -- in fact, all search inquiries use basically the same three-part structure: (1) specification of data base to be searched, using one or two keyboard characters, (2) search words and Boolean operators, and (3) specifications of type of output desired.

An example of how the search would be entered is given on the following page in Figure 1, where the indented lines are the computer, and the lines to the far left are the user. Each user-typed line is transmitted to the computer by depressing a carriage return key on the terminal keyboard.

The search inquiry shown on page 6 in Figure 1, uses the symbol AS to specify the search data base as the one to be queried. Each line containing search words can be substituted later as a unique individual term. For example, under L3, the user is asking for searches that contain L1 (either CONCEPT or MOTIVATION) and also L2 (ESTEEM). This final line, L3 specifies the words that the user wants to appear in a stored search before it can be

AS

SEARCH NO. 00012543	-----	Specifies data base of searches
L1	-----	Labels and search number automatically generated
CONCEPT OR MOTIVATION		
L2		
ESTEEM		
L3		
L1 AND L2;	-----	Use of labels as search words in final search re- quirement semi-colon stops generation of labels
LIST SEARCHES	-----	Output form desired requests listing of entire search
END	-----	END statement is signal for processing to begin

Figure No. 1 Search Inquiry Using Search Data Base

retrieved. The statement LIST SEARCHES is the form of output the user wants listed for retrieved items. END statement, followed by depressing the carriage return key, is the signal for the system to begin processing the search inquiry.

Two types of output can be requested when using the search data base:  
(1) LIST SEARCHES, which requests a printing of the entire search as originally submitted to the system, along with data on the number of documents which were retrieved and printed for the search, and (2) LIST WORDS, which requests a printing of all unique words found in the retrieved searches, along with the number of searches in the data base containing these unique words.

An example of LIST SEARCHES output is given in Figure 2; a LIST WORDS output is shown in Figure 3.

As Figure 2 shows at the top, the search inquiry is ended with the END statement, and the system reply states that SEARCHING is taking place, followed by the maximum number of items (in this case searches) possible, which is 12. The command )GO automatically prints 10 searches at a time like the one shown, and then reports that 2 of the 12 total searches retrieved may be available.

END

SEARCHING

MAXIMUM ITEMS POSSIBLE: 0000012

)GO

```
00017
L1 MOTIVATION OR ESTEEM OR ACHIEVEMENT;
LIST WORDS
520  10  007234
```

Example of one retrieved  
search from the search  
data base

[9 more searches printed out here]

MAY BE 2 MORE ITEMS

Figure No. 2 Output Example of "LIST SEARCHES" in Search Data Base

The message NO MORE ITEMS AVAILABLE would have been printed if only 10 documents were retrieved and printed. To print the remaining two searches, the user types )GO.

In the boxed example of a retrieved search, the first number (00017) is the accession number assigned to this search by SUPARS/DPS input processing. The search lines and words as originally submitted by a user are printed next, along with the form of output requested, which in this case was LIST WORDS. The three numbers below the LIST statement represent (1) the number of documents originally retrieved with this search, (520); (2) the number of documents actually printed (10), and the sequential number originally assigned to this search as identification, i.e., this was the 7234th search processed (SEARCH NO. 007234).

As Figure 3 shows, the other search type of output for the search data base is the LIST WORDS option. LIST WORDS prints just the unique word found in all searches which were retrieved from the search data base. The user is given the message "MAXIMUM ITEMS POSSIBLE: 0000011" which indicates that there are 11 other unique words found in searches that contain L1 (either concept or motivation) AND also L2 (esteem) which was the search inquiry of Figure 1. Up to 100 words at a time will be printed when the user has selected the LIST WORDS option. After electing to print the 11 words retrieved by typing the command )GO, the system responds with the message "NO. IN ( ) IS NO. OF SEARCHES" reminding the user that the postings following each word

END

SEARCHING

MAXIMUM ITEMS POSSIBLE: 000011

GO

NO. IN ( ) IS NO. OF SEARCHES

CONCEPTS	(000139)	CONCEPTUAL	(000120)	CONCEPTUALIZATION	(000102)
MOTIVE	(000090)	MOTIVES	(000040)	MOTIVATES	(000034)
ACHIEVE	(000032)	THOUGHT	(000030)	THINK	(000021)
IMAGE	(000014)	MOTIVATED	(000002)		

NO MORE ITEMS AVAILABLE

Figure No. 3 Output Example of "LIST WORDS" in Search Data Base

ΔV

SEARCH NO. 000367

L1

ESTEEM;

LIST WORD(S)

END

Figure No. 4 Vocabulary Data Base Input: Determining Status of Single Word

in the LIST WORDS output represent the number of searches containing these words. When all eleven words have been printed the message "NO MORE ITEMS AVAILABLE" indicates that all items retrieved have been printed. Had the MAXIMUM ITEMS POSSIBLE total been more than 100, the user would simply have to repeat the )GO command to have the next 100 (or less) words printed.

## (2) The Vocabulary Data Base

A second on-line option available to the user was the vocabulary data base. The vocabulary data base was different from the search data base because no reprocessing of data had to be made through the SUPARS/DPS loading process. The vocabulary containing all unique free-text terms and the number of documents containing each word was simply the portion of the inverted file developed from the original processing of documents.

Because the vocabulary was not reprocessed through DPS, as the stored searches were, and could not be accessed through use of the basic DPS search module programs, new search module programs had to be constructed. In the version implemented during SUPARS II, users were limited to entering as input only one word (or character string) per search because of programming limitations. Later versions of the vocabulary search module could extend this capability to more than one word.

To allow the user to "browse" through the alphabetical listing of words in the vocabulary data base, three search options were considered and two were implemented.

The first search option allowed a user to determine if his potential search word was in the vocabulary of free-text words. The second option allowed the user to enter any character string prefix of from 1 to 25 characters and have all words containing that prefix retrieved. The third option would have allowed a user to enter a word and specify the "n" words before and/or after it from the alphabetical list to be printed as output. Because of programming difficulties (explained in Section II), this option was not made available.

As Figure 4 indicates, the form for entering a search of the vocabulary data base is basically the same as the search data base shown in Figure 1. In the Figure 4 example, by typing the characters ΔV, the user specifies the vocabulary data base as the one he wants to search through. To determine, for example, if the word "Esteem" is contained in the free-text vocabulary of acceptable search words, the word is entered under label number 1, L1. The output form is specified in the search by typing LIST WORD, or LIST WORDS. The output that would be received by a user if the word "Esteem" were in the vocabulary data base is shown in Figure 5. The reply that the "MAXIMUM ITEMS POSSIBLE is 00001" indicates that the single input word has been retrieved. The user can either elect to type )GO and have the word and document postings number printed, as shown, or begin a search of any one of the two other data bases by typing the appropriate delta Δ symbol, i.e. Δ for document data base, or ΔV for vocabulary data base.

The message "NO. IN ( ) IS NO. OF DOCUMENTS" in Figure 5 reminds the user that the number in parentheses after the retrieved word represents the



END

SEARCHING

MAXIMUM ITEMS POSSIBLE: 000001

)GO

NO. IN ( ) IS NO. OF DOCUMENTS

ESTEEM (000123)

NO MORE ITEMS AVAILABLE

Figure 5. Vocabulary Data Base Output: Determining Status of Single Word

number of documents in which the word is found. This message was used to avoid confusion with the LIST WORDS option in the search data base that posts the number of searches in which each word was found.

The second search option available to the user searching the vocabulary data base is the truncation expander. The terms "truncation" and "expander" mean that a user can type in any prefix truncation of a word followed by a special symbol [Motivate(?)] and the system will "expand" that root to retrieve from the vocabulary any word with the prefix "M-O-T-I-V-A-T-E." The output for this option will print up to 100 words and document postings with a single )GO command, and would be basically the same as the example in Figure 5 with up to 25 rows of four words a row printed.

### (3) Integration of All Three Data Bases in Searching

By integrating the use of all three data bases, a user could continually revise his searching strategy. A preliminary statement of his information requirements could lead to the selection of certain "key" words which would be candidates for the actual search inquiry.

Before a search inquiry was submitted to the document data base for the retrieval of documents, a user could revise his overall strategy through the use of the vocabulary and/or the search data base.

For example, a user could query the vocabulary data base to retrieve all free-text terms with a specified prefix. Terms having no relation to the area of interest could be eliminated (e.g. "RATCLIFF" and "RATIFY" would contain the prefix "RAT" but be of little help in a study of small rodents) and those with some connection could be retained for further use. Also, the number of document postings for each useful word could be used to further sort out terms that retrieved in the user's estimation too many or too few documents.

A user who felt that he had a sufficient number of terms in order to proceed could submit an inquiry directly to the document data base and begin the retrieval of documents. A user who needed more terms related to his informational requirements could directly begin querying the search data base to find additional terms.

Input could consist of words and word combinations a user extracted from his information requirement statement, or any other source. Output would be all the previously submitted search inquiries containing those terms specified in the input format. The user could print out the entire search inquiry as output or merely a listing of all unique terms found in the retrieved inquiries.

New terms found through queries of the vocabulary and search data bases could then be used to develop and submit new inquiries to the basic data base of documents. In turn, new terms found in some of the relevant documents retrieved could be used to modify the inquiries submitted to the document data base or could be used to find additional terms from the search or vocabulary data base.

This iterative process of searching through first one data base and then another in any order or at any individual pacing rate could help the user to develop search inquiries that were more effective for his specific requirements.

The limitations of these two new searching algorithms are fairly evident. First, the usefulness of a search data base depends on its variety and depth. Before it can become maximally useful to most users, the data base has to "build up" a stockpile of search inquiries in a wide range of sub-areas. Depending on the number of potential users and the subject matter, this building up process could take a substantial time. The most popular and frequently submitted inquiries could represent a wide range of users or any small specialized sub-group.

#### (4) Alternatives Considered but Not Implemented

One alternative which was not a direct part of the objectives and was not implemented because of time and programming limitations was the establishment of a user queue that would control and report to users their order of access to the data bases. The queuing system was originally devised to overcome the previous user dissatisfaction with not being able to enter an inquiry because the maximum amount of computer core necessary was already in use. The queuing system was developed to establish two separate queues:

(a) the primary queue, which would consist of those users who could have search inquiries submitted and processed with the available computer core available, and (b) the secondary queue, which would allow additional users to enter their search inquiry, have it held in temporary storage, but not have it processed. Each user in the secondary queue would be told how many people were ahead of him in the queue. When a space in the primary queue was vacated, each member of the secondary queue would move up one space and be given a printed message that "There are x users ahead of you." When "x" was zero, the user would be first in line in the secondary queue.

The queuing system had the advantages of (a) giving the user a chance to enter a search inquiry without being "bounced" from the system, (b) transmitting feedback to the user from the system about his search inquiry status, (c) allowing more users than just those having their search processed to begin interaction with SUPARS/DPS.

However, the complexities in the development of the programming task to maintain control over the status and locations of 15 to 20 different users, to recognize and compensate for inquiries that were interrupted, and to account for user inquiries that were abruptly terminated accounted for the discontinuation of the queuing system development.

## 5. MODIFICATION OF THE FREE-TEXT VOCABULARY FILE

A second objective of the sponsoring agency dealt with a way to handle the large free-text vocabulary that was developed and stored for use in the inverted file of documents. For SUPARS II, the SUPARS/DPS vocabulary file contained 106,702 unique words for the 46,828 documents stored in the document data base.

Because free-text processing generates very large vocabularies such as these, some research choice has to be made between (a) developing techniques to control the vocabulary and reduce its growth while still providing representation of the data base, or (b) providing for the expansion and continued development of the vocabulary.

The research direction taken during this effort was directed at developing techniques for storing and making available all free-text words derived from document process, with the exclusion of common words, such as "and", "or", "but", etc. The choice was made on the basis of three considerations: (a) The maintenance of the research posture of a free-text investigation called for the retention of a non-restricted choice of words and terms in the vocabulary. Restrictions on the vocabulary would involve, in some measure, man-made decisions that would result in one form or another of a controlled, rather than a free vocabulary, (b) restrictions on search word choice eliminate the exhaustivity and specificity essential for effective free-text search inquiries, especially when the primary user has the responsibility for formulating his own search, and (c) a relatively easy but time consuming and major programming technique was available for effecting the expansion of the vocabulary.

In addition, empirical evidence from the SUPARS I research effort indicated that after 35,874 documents were processed, each new document contributed only one new unique word to the free-text vocabulary. At the beginning of processing an average of 8 new words were added for each new document. A complete description of the data base growth is given in Section II.

To allow for an expansion of the vocabulary, a modification was made to the coding structure of the original DPS program. The coding in the original program is 16-bit structure, i.e.  $2^{16}$ . This meant that the vocabulary for any one file could be no larger than  $2^{16}$ , or 65,534 words. Under this arrangement, the only way to have a larger vocabulary would be to define and construct a

completely new file. This second file would have to be searched separately and would be one which "started from the beginning" storing words already found in the first file. Among other limitations, this multi-file structure meant repetitive and inefficient searching and retrieval of documents.

By reprogramming the original DPS coding with a 32-bit structure, i.e.  $2^{32}$ , a greatly increased capacity could be developed. This 32-bit coding provides for over 4 billion unique words to be stored. The programming work involved changing each segment of the old search module programs from 16-bit to 32-bit coding, and then debugging and testing the new programs during test loadings of documents. See Section II for documentation of this programming work.

This new coding structure allowed a single SUPARS/DPS file to grow to a total of 196,702 unique words, after the free-text processing of 46,828 documents.

#### 6. DOCUMENTING COST, USE, AND GROWTH OF SUPARS/DPS II

The final major objective of the sponsoring agency was the reporting of the cost, use, and growth of SUPARS/DPS II. These three areas were documented in the following ways and are explained in detail below:

- a. Cost -- a cost-performance study of the operating efficiency and effectiveness of the SUPARS/DPS system
- b. Use -- a structured phone interview with a representative sample of users and nonusers; an attitude measure of users from a semantic differential scale; and an analysis of user initiated calls for help and information to a special telephone line, and
- c. Growth -- analysis of the growth of the three major data bases; analysis of user demographic data and rate of user registration with the system; and analysis of the unobtrusively recorded on-line user interaction during the 2-month period of system operation.

##### a. Documenting Cost of SUPARS/DPS II

A cost-performance study was conducted during January 1972 in order to take advantage of results from a newly installed IBM 370/155 which was used to run the SUPARS/DPS II system from January on.

The procedures used and the evaluation of the data are given in detail in Section IV.

The objective of this study was to develop cost-performance curves for the three data bases that could be employed by the user during the on-line searching process. These three data bases included: (1) documents, (2) free-text vocabulary, and (3) previously submitted search inquiries.

The measure of cost was the number of documents retrieved; a secondary cost measure was the central processing unit time of computer processing of a SUPARS/DPS search inquiry as measured by the IBM 370/155 operating system. The measure of system performance was the recall ratio, i.e. the ratio of the relevant documents retrieved to the total number of relevant documents in the data base.

By determining separate cost-performance relationship curves for each of the three data bases, empirical data was obtained to judge the relative efficiency and effectiveness of these three algorithms in an interactive, free-text system.

b. Documenting Use of SUPARS/DPS II

(1) Structured Phone Interview

One technique used to assess the use and nonuse of SUPARS/DPS was a structured interview administered by telephone during December, 1971. The purpose of the interviews, based on random samples of users and nonusers, was to obtain reliable estimates of the reactions to SUPARS of user and non-user groups.

The pre-tested interview guide was based on extensive personal interview schedules conducted during Fall, 1970.

The survey was aimed at obtaining data on the following six questions:

- a. Can we identify the variable(s) which differentiate the population of registrants into users and nonusers?
- b. What mechanisms are effective in the education of SUPARS users?
- c. How often and to what extent is the SUPARS system used?
- d. What are the major problems encountered by users?
- e. How do users judge the quality of their search output?
- f. How do users and nonusers characterize their reactions toward the system?

The techniques used for selecting the sample, the procedures used in interviewing, a summary and discussion of the responses, and conclusions of the telephone interview are given in Section III of this report.

(2) Attitude Measures

A second technique used to obtain data on the usage of the SUPARS system was through the application of the Semantic Differential Attitude Scale. The Semantic Differential was used because it could identify independent



dimensions of users' attitudes toward the system. The sample of 102 people given the attitude scale represented users of the system from such academic fields as psychology, education, journalism, social work, and library science.

The study conducted during 1971 was a replication of the 1970 study (reported in the 1971 SUPARS Final Report, Section IV-B-5) and indicated similarity of response that supported the reliability of the testing instrument for assessing user reaction to an interactive bibliographic retrieval system.

A complete description of the study which was conducted is given in Section III.

### (3) User Initiated Calls

A third means of examining use of the SUPARS system was through an analysis of data from user initiated telephone calls to a special HELP! line manned by experts in the use of the system. The basis of making the line available was to allow users to know that there was "someone out there who is interested and will help." This type of personalized help is a major factor in overcoming the stereotype image of the cold, impersonal computer over which the user has no control. Especially with users new to the computerized retrieval system, the chance to talk to a human and be given an explanation of why the system might not be working, or what type of remedial action could be taken when a user error was made, was extremely important in maintaining a satisfied and reasonably interested group of users.

Various categories of inquiries made by users are grouped, totaled, and examined under such areas as (a) requests for general information, (b) user problems in interacting with the system, (c) computer related problems, and (d) problems related to too many users in the queue.

The analysis of the various inquiries made through the HELP! line is presented in detail in Section III.

## c. Documenting Growth of SUPARS/DPS II

### (1) Growth of the Three Major Data Bases

The documentation of the growth of the SUPARS/DPS system began with the three major data bases that were accessible on-line and interactively, by users. A complete description of the growth of these data bases is given in Section II of this report. Included are summaries of the number of documents processed by SUPARS/DPS for various periods of time, and the growth rate of the free-text vocabulary. Also documented are the development of a data base for previously submitted search inquiries, and another data base of the free-text vocabulary terms.

### (2) Profiles of User Registrants

A second technique used to document the growth of the SUPARS/DPS system was through a tabulation and thorough analysis of the demographic data

supplied by each individual who registered to use SUPARS. This analysis includes the cumulative growth of the registrant population, and a descriptive examination of the registrant's status, department, time spent in teaching, research, etc., user of Psychological Abstracts and other computer related activities, and previous experience with computers in general.

This profile of registrants is given in more detail in Section III.

### (3) STATPAC

Another means used to document the growth of the SUPARS system was by means of data obtained from the STATPAC program, which unobtrusively recorded and collected the entire user interaction and other system parameters. The type of data recorded for later batch-mode retrieval included: (a) user identification number, (b) date of interaction, (c) terminal use, (d) clock time used during searching, (e) the actual line-by-line interaction, including the user's search inquiry and printed output in a condensed version.

Other data included the computer CPU time to process an inquiry, measures of computer channel use and input-output sequences, and a cost figure for the processing and output.

The STATPAC programs were written as a general retrieval system which included a variety of identifiable fields (such as date, time, terminal, output, cost, etc.) and Boolean and other logical operators that could be used to combine fields in various combinations for retrieval.

Two standard summaries were printed at two-week intervals to monitor the various types of use of the system. Other summaries could be tailored individually according to the field or field combinations desired for a more specific analysis of data that was stored.

The analysis of the STATPAC program, and examples of data that reflect the growth of the SUPARS interactive retrieval system are shown in Section III.

## 7. SUMMARY

This introductory section has presented the major objectives of a study of free-text retrieval evaluation conducted by the SUPARS Research Group at the School of Library Science, Syracuse University. The major portion of the section was devoted to a description of the work accomplished during the period February 1, 1970 - January 31, 1972 in relation to the three major objectives:

- a. Developing recall-improving algorithms that could be added to the existing on-line program.
- b. Reducing the rate of vocabulary growth and increasing the vocabulary capacity of the Document Processing System.
- c. Developing statistical data about system use to document the use, growth, and cost of the free-text retrieval system.