

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Syracuse University School of Library Science Syracuse, New York 13210		2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	
		2b. GROUP N/A	
3. REPORT TITLE FREE TEXT RETRIEVAL EVALUATION			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Final Report 1 February 1971 - 31 January 1972			
5. AUTHOR(S) (First name, middle initial, last name) Pauline Atherton Kenneth H. Cook Jeffrey Katzer			
6. REPORT DATE July 1972		7a. TOTAL NO. OF PAGES 196	7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. F30602-71-C-0185 Job Order No. 45940000		9a. ORIGINATOR'S REPORT NUMBER(S) None	
		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-72-159	
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited.			
11. SUPPLEMENTARY NOTES None		12. SPONSORING MILITARY ACTIVITY Rome Air Development Center (IRDT) Griffiss Air Force Base, New York 13440	
13. ABSTRACT The basic problems this research effort investigated were (1) the development, implementation, and evaluation of algorithms to improve recall levels in interactive, free-text retrieval using a modified version of IBM's Document Processing System (DPS), (2) the development of techniques for increasing the vocabulary capacity of DPS, and (3) unobtrusive statistical data gathering of system use, growth, and cost, through a previously developed computer program. A free-text document data base (DDB) of 46,828 bibliographic citations and abstracts from <u>Psychological Abstracts</u> was developed. Also, two interactively accessible data bases were developed and implemented to provide free-text vocabulary control and recall improvement directly to the user. No intermediaries were used in the retrieval process. These two algorithms were (1) a Vocabulary Data Base (VDB) containing the 106,702 unique free-text terms from the inverted file of processed documents in the DDB, and (2) a Search Data Base (SDB) containing previously submitted user search inquiries to the DDB. A two-month period of experimental use of the entire system with all three data bases by students, staff, and faculty of Syracuse University in fall 1971 provided the required "real-life" field environment. A total of 2399 search inquiries were submitted via the 2741 terminals. The system operated under both the 360/50 and the 370/155 operating systems. The capacity of the DPS vocabulary was increased by the development and successful implementation of computer programs that revised the DPS coding structure of the vocabulary file. The newly developed structure changing the 16-bit coding to 32-bit coding resulted in increasing the vocabulary capacity			

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Abstracts Indexes Information Retrieval						
13. Abstract (continued) from the former single-file limit of 65,534 to over 4 billion terms. An extensive user-oriented public relations/publicity, instruction, and education package was developed and implemented. This user emphasis resulted in a significantly greater number of registrants and actual users in the system, which did not use search intermediaries, than during a similar fall 1970 period. Results of a controlled cost-performance (recall) study indicated that use of the VDB or the SDB yielded better cost-performance levels, especially at higher recall percentages, than by use of the DDB alone. These results are the initial findings, and would require additional testing to substantiate their validity. Other evaluative techniques included a Semantic Differential attitude scale for interactive retrieval systems, a structured telephone interview of users, and a special number users could call for help in developing search inquiries. The STATPAC program for unobtrusively monitoring and retrieving data on system use, growth, and cost was modified and successfully used to provide evaluative data. General conclusions are that interactive, free-text cost-performance (recall) levels can be improved by direct user control of algorithms providing vocabulary control. Real-life application of these algorithms suggests that the recall improvement available through controlled vocabulary or indexing systems might be obtained readily by the free-text searcher who is provided with techniques such as those implemented in this research effort.						

N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM THE
BEST COPY FURNISHED US BY THE SPONSORING
AGENCY. ALTHOUGH IT IS RECOGNIZED THAT CER-
TAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RE-
LEASED IN THE INTEREST OF MAKING AVAILABLE
AS MUCH INFORMATION AS POSSIBLE.

FREE TEXT RETRIEVAL EVALUATION

**Pauline Atherton
Kenneth H. Cook
Jeffrey Katzer**

Syracuse University

**Approved for public release;
distribution unlimited.**

ic

FOREWORD

This report was done by the Syracuse University Psychological Abstracts Retrieval Service (SUPARS) Research Group at the Syracuse University School of Library Science, under contract F30602-71-C-0185, Job Order Number 45940000, for Rome Air Development Center, Griffiss Air Force Base, New York. Mr. Nicholas M. DiFondi (IRDT) was the RADC Project Engineer.

This report represents a continuation of work conducted under contracts F30602-69-C-0013 and F30602-70-C-0190 during the period 1 July 1969 and 31 January 1971.

The current research contract covers work accomplished during 1 February 1971 to 31 January 1972, and deals with the development, implementation, and evaluation of new algorithms to improve recall in an interactive, on-line, free-text retrieval system.

Individual authors of different sections include: Section I, Kenneth H. Cook; Section II, Lynn Trump and Mr. Cook; Section III, Sandra Browning, June Brower, Jeffrey Katzer, Patricia Moell, and Peggy Mucia; and Section IV, Mr. Katzer.

Acknowledgement is given to the following individuals: Dean Roger C. Greer and Dr. Allan F. Hershfield of the School of Library Science who provided continued support for the SUPARS project; Mr. William J. Jones, Director, Syracuse University Computing Center, for his cooperation and assistance in the implementation of the on-line system; and to Mrs. Brenda Lefebvre, whose overall editorial and typing made this report possible.

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS).

Nicholas M. DiFondi

Approved: NICHOLAS M. DIFONDI
Technical Evaluator

Franz H. Dettmer

Approved: FRANZ H. DETTMER
Colonel, USAF
Chief, Intel & Recon Division

FOR THE COMMANDER:

Fred I. Diamond

FRED I. DIAMOND
Acting Chief, Plans Office

ABSTRACT

The basic problems the current research effort (February 1, 1971-January 31, 1972) investigated were (1) the development, implementation, and evaluation of algorithms to improve recall levels in interactive, free-text retrieval using a modified version of IBM's Document Processing System (DPS), (2) the development of techniques for increasing the vocabulary capacity of DPS, and (3) unobtrusive statistical data gathering of system use, growth and cost through a previously developed computer program. A free-text document data base (DDB) of 46,828 bibliographic citations and abstracts from Psychological Abstracts was developed. Also, two interactively accessible data bases were developed and implemented to provide free-text vocabulary control and recall improvement directly to the user. No intermediaries were used in the retrieval process. These two algorithms were (1) a Vocabulary Data Base (VDB) containing the 106,702 unique free-text terms from the inverted file of processed documents in the DDB, and (2) a Search Data Base (SDB) containing previously submitted user search inquiries to the DDB. A two-month period of experimental use of the entire system with all three data bases by students, staff, and faculty of Syracuse University in fall 1971 provided the required "real-life" field environment. A total of 2399 search inquiries were submitted via the 2741 terminals. The system operated under both the 360/50 and the 370/155 operating systems. The capacity of the DPS vocabulary was increased by the development and successful implementation of computer programs that revised the DPS coding structure of the Vocabulary file. The newly developed structure changing the 16-bit coding to 32-bit coding resulted in increasing the vocabulary capacity from the former single-file limit of 65,534 to over 4 billion terms. An extensive user-oriented public relations/publicity, instruction and education package was developed and implemented. This user emphasis resulted in a significantly greater number of registrants and actual users in the system, which did not use search intermediaries, than during a similar fall 1970 period. Results of a controlled cost-performance (recall) study indicated that use of the VDB or the SDB yielded better cost-performance levels, especially at higher recall percentages, than by use of the DDB alone. These results are the initial findings, and would require additional testing to substantiate their validity. Other evaluative techniques included a Semantic Differential attitude scale for interactive retrieval systems, a structured telephone interview of users, and a special number users could call for help in developing search inquiries. The STATPAC program for unobtrusively monitoring and retrieving data on system use, growth, and cost was modified and successfully used to provide evaluative data. General conclusions are that interactive, free-text cost-performance (recall) levels can be improved by direct user control of algorithms providing vocabulary control. Real-life application of these algorithms suggests that the recall improvement available through controlled vocabulary or indexing systems might be obtained as readily by the free-text searcher who is provided with techniques such as those implemented in this research effort.

EVALUATION

The objective of this study was to develop, implement and evaluate methods for increasing vocabulary file space and improving the retrieval effectiveness of a free-text indexed on-line document retrieval system at Syracuse University. The system first operated on an IBM 360/50 computer and recently on a 370/155 computer. The document data base consisted of 46,828 bibliographic citations and/or abstracts from Psychological Abstracts. Vocabulary storage space was increased by developing computer programs to convert the half-word (16-bit) coding scheme as defined by the IBM/DPS program to full word (32 bit) coding. Methods of improving retrieval effectiveness include a vocabulary data base and a search data base as on-line searching aids. Results are reported in terms of nine levels of Recall (the portion of relevant documents retrieved), total retrieval (the number of documents retrieved to achieve a specific Recall level), and cost-performance (the cost incurred to achieve a specific Recall level).

There are several significant conclusions derived from the results of this effort:

1. The conversion from 16 bit to 32 bit coding has increased the limit on a single vocabulary file size from approximately 65,000 words to over 4 billion words. Without this capability, upon reaching the 65,000 word limit a new vocabulary would have to be defined and then created by the free-text indexing program resulting in inefficient use of core due to redundant information between files and longer search cycles.
2. The use of the vocabulary data base as a searching aid results in better cost performance than using the other data bases. Since the vocabulary data base is a portion of the inverted file developed from the original processing of documents, it is relatively inexpensive to add to the system and reduces cost performance by decreasing the need to search the more expensive document data base.
3. Statistically significant differences in total retrieval at all levels except the 10% level of Recall reflected variations in efficient use of the system by search experts. Each expert was equally knowledgeable in the subject area, equally trained in the use of free-text retrieval, and used the same information requirement statements to formulate his searches.

However, each chose a different search strategy in the hopes of minimizing total retrieval. This finding indicates that it is difficult for system experts to find or establish efficient search methods. Since most users may be knowledgeable in their fields but not necessarily experts in the use of on-line retrieval, attempts by novice users to formulate efficient search strategies may be difficult to achieve.

4. Total document retrieval is very high at all levels of Recall. Although use of the vocabulary data base or the search data base does reduce total retrieval considerably from that achieved by using the document data base, too much time and effort would be required of the user to scan for relevant documents.

As a result of this study future research can be directed toward establishing better search methods to reduce total retrieval, continue work on the Recall improving algorithms to insure total retrieval improvement does not negate their effect on system effectiveness, and identify differences in the experts methods of searching to determine the impact on the general population of users and make adjustments accordingly.

Nicholas M. Di FonDI
NICHOLAS M. DIFONDI

TABLE OF CONTENTS

SECTION	TITLE	PAGE
I.	INTRODUCTION	
	1. Previous Research Efforts	1
	2. Objectives of Current Work (SUPARS II)	2
	3. Philosophy and Approach	3
	4. Overview of Work Accomplished: February 1971 - January 1972	4
	a. Development of New Searching Algorithms	
	(1) The Search Data Base	5
	(2) The Vocabulary Data Base	9
	(3) Integration of All Three Data Bases in Searching	10
	(4) Alternatives Considered but not Implemented	11
	5. Modification of the Free-text Vocabulary File	12
	6. Documenting Cost, Use, and Growth of SUPARS/DPS II	
	a. Documenting Cost of SUPARS /DPS II	13
	b. Documenting Use of SUPARS/DPS II	14
	(1) Structured Phone Interview	14
	(2) Attitude Measures	14
	(3) User Initiated Calls	15
	c. Documenting Growth of SUPARS/DPS II	
	(1) Growth of the Three Major Data Bases	15
	(2) Profiles of User Registrants	15
	(3) STATPAC	16
	7. Summary	16

SECTION	TITLE	PAGE
II.	DATA BASE	
	1. Document Data Base	17
	2. Vocabulary Data Base	
	a. Special Programs Developed for a Vocabulary Data Base	22
	b. Expanding the Capacity of the Vocabulary Data Base	24
	3. The Search Data Base	
	a. Definition of Data Base Description	27
	b. Modifying Search Module and Reformat Programs	27
	c. Specification of Output Forms	29
	4. Summary	29
	5. Implications and Projections	29
	APPENDICES	
	Appendix I System Overview	32
	Appendix II Translate Psychological Abstracts	34
	Appendix III Search Reformat	49
III.	THE USER	
	1. Profile of Registrants	57
	a. Cumulative Growth of the Registrant Population	58
	b. Demographic Data	60
	(1) University Status	60
	(2) Departmental Status	60
	(3) Time Spent on Teaching, Research and Other Activities	60

SECTION	TITLE	PAGE
	(4) Previous SUPARS Registration and Use	64
c.	Use of <u>Psychological Abstracts</u> by Registrants	64
	(1) Recent Use of <u>Psychological Abstracts</u>	64
	(2) Future Use of <u>Psychological Abstracts</u>	68
	(3) Recent and Future Use of <u>Psychological Abstracts</u> for Preparation of a Term Paper, Thesis, or Speech	71
	(4) Specific Need for <u>Psychological Abstracts</u>	71
d.	Computer Experience of Registrants	76
	(1) Previous Experience with Computer Terminals	76
	(2) Previous Experience with Computer-Based Retrieval Systems	77
e.	Summary	77
2.	Publicity	78
3.	4220 LOGS	78
	a. Description of Log Summary B	79
	b. Summary	80
4.	STATPAC	
	a. Summary 1	85
	b. Summary 2	88
	c. Summary	90
5.	Telephone Interviews of SUPARS Registrants	91
	a. The Sample	93
	b. Completed Versus Non-completed Interviews	93

SECTION	TITLE	PAGE
	c. Users: Description	94
	d. Users: General Usage Patterns	95
	e. Users: Reactions	99
	f. Nonusers: Description	101
	g. Summary	104
6.	Semantic Differential	106
	a. Description of Procedure	106
	b. Data Organization	112
	c. Results	113
	d. Comparison with the Results of Last Year's Study	113
	e. Conclusions	118
	f. Discussion	118
7.	Implications and Projections	119
	a. User Orientation	119
	b. Publicity and Instruction Functions	119
	c. User Control of Interaction	120
	d. Obtaining User Response	121
	e. Conclusion	121
APPENDICES		
Appendix IV	Program Description of SUPARS STATPAC	122
Appendix V	SUPARS Telephone Survey Question and Registration Form	134
Appendix VI	Introduction to Semantic Differential Questionnaire	147

SECTION	TITLE	PAGE
IV	COST-PERFORMANCE ANALYSIS	150
	1. Definitions	150
	2. Method	152
	a. Experimental Design	153
	b. IRSs	154
	c. Identification of Relevant Documents	156
	d. Search Experts	156
	e. Procedure	157
	f. Analysis	157
	3. Results	158
	a. Use of Optional Data Bases	158
	b. Computer Costs	161
	c. Cost-Performance	163
	d. Comparing the Cost-Performance of Two Years of SUPARS Operation	172
	4. Discussion	177
	a. The VDB	177
	b. The SDB	179
	c. Differences Among IRSs and Experts	180
	5. Summary of Findings	181
	APPENDICES	
	Appendix VII Instructions to Subjects	183
	Appendix VIII Analysis of Variance Summary	185

SECTION	TITLE	PAGE
V.	CONCLUSIONS AND RECOMMENDATIONS	192
1.	Major Conclusions	192
a.	Recall-Improving Algorithms	192
b.	Improving SUPARS/DPS Vocabulary Capacity	193
c.	Computer Programs Recording System Use, Cost, and Growth	193
d.	Publicity, Instruction and Education	193
e.	Systematic Evaluation of User Attitudes and Reactions	193
f.	Growth of the Document Data Base	193
g.	Cost Comparisons Between SUPARS I and SUPARS II	194
2.	Recommendations Made	194
a.	Continue Development of User Control of Vocabulary and Synonyms	194
b.	Maintain a Non-Reentrant Search Module	194
c.	Improve User Access to Hardware	194
d.	Investigate Searching Styles and Techniques of Free-Text Users	195
e.	Develop Interactive Algorithms Based on User Styles and Techniques	195
	REFERENCES	196

LIST OF FIGURES AND TABLES

FIGURE		PAGE
1	Search Inquiry Using Search Data Base	6
2	Output Example of "LIST SEARCHES" in Search Data Base	7
3	Output Example of "LIST WORDS" in Search Data Base	8
4	Vocabulary Data Base Input: Determining Status of Single Word	8
5	Vocabulary Data Base Output: Determining Status of Single Word	10
6	Growth Rate of SUPARS/DPS II Document Data Base	20
7	Cumulative Growth of the SUPARS/DPS Vocabulary Data Base	23
8	Modification of Characters in Search Data Base	28
9	Cumulative Growth of SUPARS II Registrant Population	59
10	Computer Cost per Search to Achieve Nine Levels of Recall: A Comparison Among Three Treatment Conditions	168
11	Two Methods of Computing Dollar Cost of Searching on DDB Using LIST BRIEF Output Format	170
12	Number of Documents Retrieved (Total Retrieval) to Achieve Nine Levels of Recall: A Comparison Among Three Treatment Conditions	169
13	Number of Documents Retrieved (Total Retrieval) to Achieve Nine Levels of Recall: A Comparison Between SUPARS I and SUPARS II Searches to the DDB	175
14	Number of Documents Retrieved (Total Retrieved) to Achieve Nine Levels of Recall: A Comparison Among the SUPARS II Conditions and the Most Efficient Use of SUPARS I	178

TABLE		PAGE
I	Loading History of SUPARS/DPS II Document Data Base	19
II	Track Allocation and Usage of Document Data Base	21
III	Track Allocation and Usage of Search Data Base	26
IV	Growth of Search Data Base	25
V	Number of SUPARS Registrants by University Status	61
VI	Number of SUPARS Registrants by Major Departmental Categories	62
VII	Number of SUPARS Registrants by Departmental Categories	63
VIII	Percentage of Time Engaged in Teaching and/or Learning	65
IX	Percentage of Time Engaged in Research	66
X	Percentage of Time Engaged in Other Activities	67
XI	Registration for SUPARS Last Year	68
XII	Usage of SUPARS by Last Year's SUPARS Registrants	69
XIII	Average Usage of <u>Psychological Abstracts</u> in the 2-3 Month Period Preceding Registration	70
XIV	Anticipated Usage of <u>Psychological Abstracts</u> in the 2-3 Month Period Following Registration	72
XV	Past Use of <u>Psychological Abstracts</u> for Paper, Thesis, and Speech Preparation	73
XVI	Anticipated Future Use of <u>Psychological Abstracts</u> for Paper, Thesis, and Speech Preparation	74
XVII	Anticipated Types of Use of <u>Psychological Abstracts</u>	75
XVIII	Previous Experience with Computer Terminals	76
XIX	Previous Experience with Computer-Based Retrieval Systems	77

TABLE		PAGE
XX	Log Summary A	81
XXI	Log Summary B	84
XXII	STATPAC Summary I	87
XXIII	Mean CPU Time and Cost Over the Three Data Bases	89
XXIV	Frequency of Completed Interviews	92
XXV	Academic Status of Users	94
XXVI	Frequency of Attempted Use of SUPARS	96
XXVII	Average Time Spent Using SUPARS in one Session at the Terminal	96
XXVIII	Frequency and Percentage of Estimated Number of Searches Made	97
XXIX	Problems Experienced by Users Which Hampered Attempts to Sign On	98
XXX	Problems experienced by SUPARS Users After Signing On	99
XXXI	Frequency and Percentage of Responses: "Have you located more relevant information with SUPARS?"	100
XXXII	Academic Status of Nonusers	101
XXXIII	Attitudes Toward and Contact with SUPARS	103
XXXIV	Concepts Used in the SUPARS Semantic Differential Packet	107
XXXV	Example of a Semantic Differential Used in the Present Study	108
XXXVI	Classification of Respondents by Completion of Semantic Differential	109
XXXVII	Classification of Respondents by Number of Searches Made	110
XXXVIII	Number of Days Students Used SUPARS	112

TABLE		PAGE
XXXIX	Concept Means	114
XL	Standard Deviations for Concept Means	116
XLI	Experimental Design	154
XLII	Characteristics of Subjects and IRSs	155
XLIII	Sources of Variation and Degrees of Freedom in Cross-over Design	158
XLIV	Use of Three Data Bases to Achieve the 90% Recall Level	159
XLV	Estimates of the Computer Cost of Searching the Three Data Bases	162
XLVI	Estimates of the Mean Cost of Searching Under the 'D' Experimental Condition at Nine Levels of Performance (Recall)	164
XLVII	Estimates of the Mean Cost of Searching Under the 'S' Experimental Condition at Nine Levels of Performance (Recall)	165
XLVIII	Estimates of the Mean Cost of Searching Under the 'V' Experimental Condition at Nine Levels of Performance (Recall)	166
XLIX	Differences Between SUPARS I and SUPARS II	173
L	Estimates of the Computer Cost of Searching to DDB Counting All Entered Searches (N=246)	174

LIST OF FIGURES FOR APPENDICES

FIGURE		PAGE
APPENDIX I		
1	System Overview	33
APPENDIX II		
1	Input Record Description	35
2	Translate Psychological Abstracts Logic Diagram	36
3	Translate Psychological Abstracts Flow Chart	37
4	Reformatted Data	41
5	Reformat Psychological Abstracts Logic Diagram	42
6	Reformat Psychological Abstracts Flow Chart	44
APPENDIX III		
1	Data Base Description	51
2	Input Record Description	52
3	Output Record Description	53
4	Reformat Searches Module 1 - BIBFLDS	54
5	Reformat Searches Module 2 - FORMSRCH	56
APPENDIX IV		
1	Data Flow of SUPARS Log	124
2	Program Flow for Producing STAT Output	125
3	Program for Conversion of Log to STAT Usable Form	126
4	MACRO - General Outline of STAT Programs	129
5	PROCESS - Main Working Section of STAT Programs	130

APPENDIX V

1	SUPARS Registration Form	146
---	--------------------------	-----

APPENDIX VIII

1	Differences Among Three Treatment Conditions at the 10% Recall Level	186
2	Differences Among Three Treatment Conditions at the 20% Recall Level	187
3	Differences Among Three Treatment Conditions at the 30% Recall Level	188
4	Differences Among Three Treatment Conditions at the 40% Recall Level	188
5	Differences Among Three Treatment Conditions at the 50% Recall Level	189
6	Differences Among Three Treatment Conditions at the 60% Recall Level	189
7	Differences Among Three Treatment Conditions at the 70% Recall Level	190
8	Differences Among Three Treatment Conditions at the 80% Recall Level	190
9	Differences Among Three Treatment Conditions at the 90% Recall Level	191

USAGE OF TERMS

Because the definitions of terms used in the information technology field are not completely standardized nor consistent, the terms in this report are explained below. An effort has been made, where possible, to follow the most consistently used and reasonable meaning to convey a concept. Where specialized or more specific usages of these terms are employed in specific sections of this report, an explanation will be given by the section author.

- (a) Delta: The character (upshift "H") on the 2741 keyboard that is used in SUPARS/DPS user interaction to access the document data base and initiate a search inquiry.
- (b) Delta S: The character "delta" and the letter "S" (S) which are used in SUPARS/DPS user interaction to access the search data base and initiate a search inquiry.
- (c) Delta V: The character "delta" and the letter "V" (V) which are used in SUPARS/DPS user interaction to access the vocabulary data base and initiate a search inquiry.
- (d) Dictionary: The internally stored list of unique free-text terms processed by DPS and the document frequency count for each work. The DPS dictionary forms one part of the inverted file.
- (e) Document: In this study, the term "document" stands for the bibliographic citation and abstract that are used as a surrogate of the original journal, article, proceeding, book, etc.
- (f) Document Data Base (DDB): Consists of SUPARS/DPS processed documents of Psychological Abstracts which are interactively accessible on-line by users. The DDB is one of three data bases available to the SUPARS/DPS user (others are vocabulary data base and search data base).
- (g) Document Processing System (DPS): The IBM free-text, batch mode programs that convert machine readable textual data into searchable and retrievable data sets organized in inverted file structure.
- (h) Free-Text: Specifically refers in this study to the Document Processing System. The general reference is to an organized system allowing the indexing and retrieval of documents or their surrogates by any of the terms used in a defined text, rather than terms derived by a controlled set of terms.
- (i) Information Requirement Statement (IRS): The verbal or written statement of an individual's interest area as generally related to documents or their surrogates. The IRS is the publically verifiable indication of the internally held construct, "information

- (j) Label (labelled line): The portion of a search inquiry, such as I1, I2, etc. that identifies and stands for the search words and operators used to act on those words; a label can be used itself in an inquiry as a search word.
- (k) Operator: The user language accepts as input one or more keywords which represent the IRS of the searcher. Keywords may be combined with Boolean operators (AND, OR, NOT) or grammatical operators (those which specify the desired proximity of keywords within a sentence or those which specify the root of a word).
- (l) Search: The search inquiry, the user/computer interaction, and the printed output, if any. The beginning of a new search inquiry marks the end of a search.
- (m) Search Data Base (SDB): Consists of the previously stored and processed search inquiries made to SUPARS/DPS. The SDB is one of three data bases available to the user and was newly developed during the current research. (Others are the document data base and the vocabulary data base.)
- (n) Search Inquiry: The user arrangement of words, word combinations and logical operators in a form acceptable as input for machine processing. A SUPARS/DPS search inquiry would consist of the free-text terms combined with Boolean and other logical operators, the request for output, and an "end" statement. Examples of search inquiries are given in Section II.
- (o) Search Word: Free-text word(s) or term(s) used as part of a search inquiry. A KEYWORD is a synonym for a search word.
- (p) STATPAC: The Statistical Package used in conjunction with SUPARS/DPS to unobtrusively collect, store and retrieve the elements of user interaction and other system parameters such as time, terminal number, cost, etc. STATPAC includes a highly flexible retrieval system in itself which allows the operator to specify and retrieve various combinations of data reflecting user interaction or system performance. In addition to standard summaries printed periodically, the operator could request, for example, a listing of the computer time used for all searches of the document data base by one-time users after a certain calendar date.
- (q) SUPARS/DPS: Syracuse University Psychological Abstracts Retrieval Service/Document Processing System. The modified DPS program developed at Syracuse University which allows on-line, interactive searching of free-text data. SUPARS/DPS I refers to the research work conducted from July 1969 to January 1971. SUPARS/DPS II refers to the work conducted from February 1971-January 1972.
- (r) Vocabulary: The on-line, interactively accessible dictionary that is stored by DPS. The term "vocabulary" rather than "dictionary" is used to connote the words and terms accessible to the user

that can be used as free-text index terms.

- (s) Vocabulary Data Base (VDB): Consists of the on-line, interactively accessible DPS dictionary of free-text terms. The VDB is one of three data bases available to the user and was newly developed during the current research. (Others are document data base and search data base).