NASA-CR-195751

College of Science and Technology

(NASA-CR-195751) PROPOSAL FOR THE CONTINUED DEVELOPMENT AND MODIFICATION OF SOFTWARE BEING USED TO SUPPORT NASA'S TECHNOLOGY UTILIZATION PROGRAM (University of Southern Mississippi) 62 p

N94-71787

Unclas

Z9/61 0002873

Science and Technology

PROPOSAL FOR THE CONTINUED DEVELOPMENT

AND MODIFICATION OF SOFTWARE BEING USED TO

SUPPORT NASA'S TECHNOLOGY UTILIZATION PROGRAM

May, 1987

G. David Huffman

PROPOSAL FOR THE CONTINUED DEVELOPMENT AND MODIFICATION OF SOFTWARE BEING USED TO SUPPORT NASA'S TECHNOLOGY UTILIZATION PROGRAM

May, 1987

G. David Huffman
College of Science and Technology
University of Southern Mississippi
Hattiesburg, Mississippi 39406-5165

TABLE OF CONTENTS

Section Number	Section Title	Page Number
	TABLE OF CONTENTS	i
	LIST OF FIGURES	ii
	LIST OF TABLES	iii
1	INTRODUCTION	1
1.1	Bibliographic Database Searching and Technology Transfer	1
1.2	Survey of Commercial Software Systems	2
1.3	Software Framework for Technology Transfer	9
2	REVIEWING AND DETERMINING THE RELEVANCY OF CITATIONS	11
2.1	SORT-AID Post Processor	11
2.2	Program NABST	12
2.3	Program REVIEW	14
2.4	Program RANK	15
2.5	Subprogram PRINT	16
3	PROGRESS TO DATE	17
3.1	Benchmark Comparisons for C and Pascal	17
3.2	Benchmark Comparisons for Various Microcomputers	17
3.3	Preliminary Benchmark Results with SORT-AID (Version 4.0)	26
3.4	Summary of Benchmark Tests	28
3.5	User Evaluation of RANK	32
3.6	Analysis Framework for Weakly-Ordered Distributions	33
3.7	Test Results	36
4	PROPOSED WORK	45
4.1	Minor Modifications to SORT-AID	45
4.2	SORT-AID Enhancements	45
4.3	Statement of Work	49
4.4	Program Budget and Schedule	50
5	REFERENCES AND BIBLIOGRAPHY	52

LIST OF FIGURES

Figure Number	Figure Caption	Page Number
1	Industrial Applications Study Paradigm	4
2	Typical Commercial Software Packages for Use in Technology Transfer	5
3	System Flow of SORT-AID	13
4	Benchmark Comparisons for C and Pascal	19
5	Benchmark Comparisons for MS-DOS Computers	22
6	Benchmark Comparisons for UNIX Computers	23
7	Computational Time versus the Number of Citations	24
8	Computational Time versus the Number of Words Processed	25
9	Computational Times as a Function of Disk Access Times for MS-DOS Computers	30
10	Hypothetical Relevance Distributions. Actual Approximates Best Case	38
11	Hypothetical Relevance Distributions. Actual Approximates Worst Case	38
12	RANK Distributions. Typical cases, $\eta_a > 0.72$	40
13	RANK Distribution. $\eta_a = 0.66$	43
14	RANK Distribution. $\eta_a = 0.55$	43
15	RANK Distribution. $\eta_a = 0.47$	44
16	Flowchart of Enhanced SORT-AID	48

LIST OF TABLES

Table Number	Table Title	Page Number
1	Systems Relevant to Technology Transfer Tasks	6
2	Comparison of SORT-AID (Version 3.0) Operating Times on Selected Microcomputers	20
3	SORT-AID (Version 4.0) Operating Times on Selected Microcomputers--Preliminary Results	27
4	Benchmark Computer Metrics	29
5	Evaluation of Rank	39
6	Program Budget	51

1. INTRODUCTION

1.1 Bibliographic Database Searching and Technology Transfer

The NASA technology transfer process can be regarded as an information system. A primary output is known as the "industrial applications study". The actual processes involved in the industrial applications study are shown in flowchart format in Figure 1, see, e.g., Huffman (1987). The first step in the process is the formulation of the problem statement. Once the problem is formulated, a series of databases are selected. These may be either commercial, e.g., DIALOG, or public domain, e.g., NASA RECON. Using database thesauri, the search strategy is formulated and a series of keywords chosen. An on-line search is conducted and a number of abstracts retrieved. These are then reviewed for relevance and a number of documents ordered. The documents can then be analyzed and, if warranted, government and/or industrial contacts instituted. Information from experts in conjunction with the assessment of the published information is used to formulate the final report.

Another output of the NASA process is the "current awareness search". The steps carried out in a current awareness search are similar, but are normally terminated after the on-line search, i.e., the final report consists of the retrieved abstracts. Presumably, the recipient of the abstracts carries out the abstract review and analysis procedure and contacts the relevant experts.

The labor intensive steps in the above process are associated with the review of the abstracts for relevance, the analysis of the relevant publications and the preparation of the final report. Analysis and report writing labor can be reduced through the employment of computer information systems tools and techniques. This is discussed in the following sections.

1.2 Survey of Commercial Software Systems

Software development programs which are relevant to technology transfer are being carried out by government laboratories, businesses and universities. The various programs are diverse and difficult to categorize. They range from highly focused products for IBM PC's such as IN-SEARCH, e.g., Menlo (1984), to a broad-based program such as the Total Information System (TIS)--a DEC VAX 11/780-based program, e.g., Hampel, et. al. (1982). The various and sundry programs deal with pre-processing, i.e., a common command language, and post-processing, i.e., abstract sorting, word processing, etc. Both aspects are being considered in the present program, but with emphasis on the personal computer software. The reasons for this emphasis are:

i. The rapid increase in microcomputer capability, e.g., the new IBM AT supports 3 mb of main memory and 40 mg of hard disk memory--all in a desk top unit.

ii. The cost trends. Microcomputer costs have fallen while communications costs have increased. This factor favors distributed computing when compared to a central, i.e., gateway, computer.

iii. The availability of microcomputer software. There is no question that the number of technology transfer oriented programs has increased many-fold in the last year. This is a strong indication of the market trends.

Many of the tasks associated with conducting technology transfer studies are rather mundane--word processing, producing hard copy of abstracts, etc. While these tasks are fairly straight-forward, they are time consuming. A number of commercial software packages can be used effectively in these tasks.

As noted above, the carrying out of technology transfer studies is labor intensive and software systems can significantly improve productivity. A number of appropriate software packages are shown in Figure 2. These are divided into the general categories of pre-processors, gateways, citation analyzers, report writing and text processors. A short description of each program is given in Table 1. There are over-lapping functions among the software systems as well as the lack of any integration among modules.

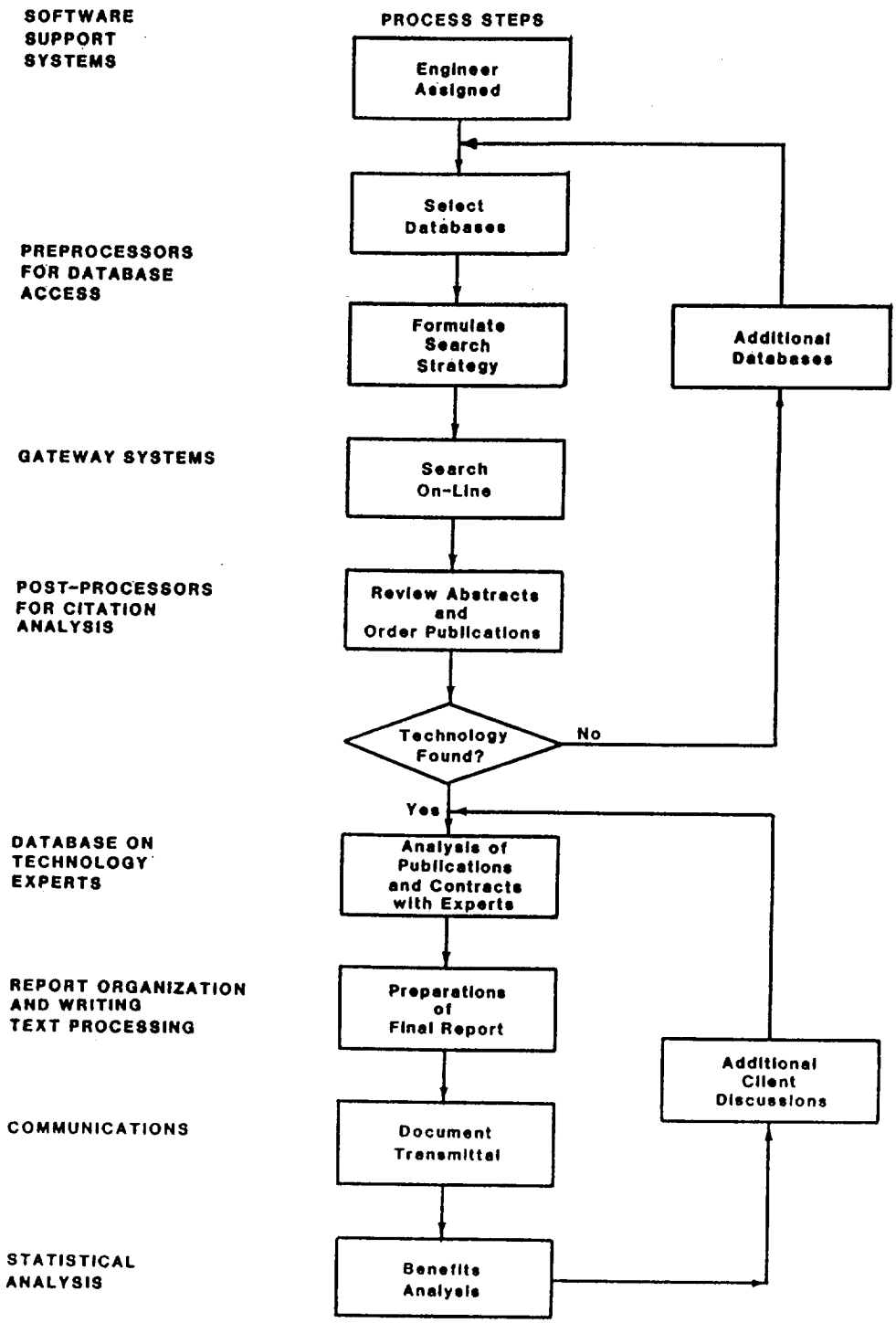


Figure 1. Industrial Applications Study Paradigm.

SOFTWARE SYSTEMS	FUNCTIONAL DATA
IN-SEARCH MIST +	Pre-processors for database access
Business Computer Network TOTAL INFORMATION SYSTEM	Gateway Systems
NUTSHELL	Post-processor for Citation Analysis
dBASE III	Database Program
OFFIX THINKTANK	Report Organization and Writing
MEMORY/SHIFT TOPVIEW	Integration
WORDSTAR SELECT WRITE	Text processing
CONNEXUS	Communications
STAT-PAK	Statistical Analysis

Figure 2. Typical Commercial Software Packages for Use in Technology Transfer

TABLE 1

SYSTEMS RELEVANT TO TECHNOLOGY TRANSFER TASKS

Program Title	Description	Environment[1]
IN-SEARCH	Front-end database searching program. Catalog of all DIALOG databases and resources. Search revision capabilities and keyword high-lighting within each reference.	IBM PC and compatibles. 192 K; two disk drives; smart or acoustic modem.
MIST +	Microcomputer communications program. Program contains a full programming language with specifications for telecommunicating, a database system and a text editor. The database can be turned into a full-fledged computer tele-conferencing system complete with electronic mail, conferences and on-line databanks.	IBM PC and compatibles. 256 K; two disk drives; hard disk recommended; smart or acoustic modem.
Business Computer Network	Logs on automatically to a number of on-line information systems. Program captures text on disk, writes messages off-line, sends them on-line and sends sequences to a printer.	IBM PC and compatibles. 128 K; two disk drives; Hayes compatible modem.

(1) Minimum equipment requirement. The software vendors are listed in Section 6. 128K denotes 128 kilobytes of main memory. IBM PC/XT or compatibles denotes a 16 bit microcomputer having a 10 Mb fixed disk drive. One disk drive denotes a 320-360Kb removable disk drive.

TABLE 1 (continued)
DESCRIPTIONS OF SOME SOFTWARE
SYSTEMS RELEVANT TO TECHNOLOGY
TRANSFER TASKS

Program Title	Description	Environment[1]
NUTSHELL	Citation analysis system which allows review, categorization, etc. Records can be indexed by title, author, keywords, etc.	IBM PC or compatibles. 128 K; one disk drive; smart or acoustic modem.
THINKTANK	Text processing program oriented toward report organization. Uses outlining techniques.	IBM PC/XT or compatibles 256 K; one disk drive.
OFFIX	Personal office system which mimics a file cabinet. Software can search a datafile for up to 10 fields simultaneously and then sort by one of the ten. Can also send information to a screen or printer.	IBM PC/XT or compatibles. 256 K.
MEMORY/SHIFT	This system allows you to run several programs simultaneously and to transfer data between them.	IBM PC/XT or compatibles. 128 K.
TOPVIEW	Provides multitasking and windowing.	IBM PC/XT or compatibles. 512K recommended; one fixed and one removable disk drive.
dBASE III	Database program which constructs and manipulates numeric and character data. Provides database manipulation directly from a keyboard. Provides capability for user-generated menus and application programs.	IBM PC or compatibles. 256K; two disk drives.

TABLE 1 (continued)
DESCRIPTIONS OF SOME SOFTWARE
SYSTEMS RELEVANT TO TECHNOLOGY
TRANSFER TASKS

Program Title	Description	Environment[1]
WORDSTAR	Screen oriented word processing system featuring integrated printing. Displays both initial entry of text and alteration of previously entered text.	IBM PC or compatibles. 128 K; one disk drive.
CONNEXUS	Communications system providing electronic mail, bulletin boards, teleconferencing, etc. Includes password access system.	IBM PC/XT or compatibles. 256 K; two disk drives; smart modem.

1.3 Software Framework for Technology Transfer

Section 1.1 discussed the general steps carried out in technology transfer studies while Section 1.2 reviewed some relevant commercial software systems. This section describes both the framework and elements of a software support system for technology transfer professionals. This system is called a "searcher's workstation" and consists of a series of productivity tools and modules. Commercially available programs, i.e., modules, are being used where possible with unique programs developed only when no suitable systems exist. As discussed in Section 1.2, the software support system is being developed for use on a personal computer--not without precedent, e.g., Bertrand (1980) or Lefkovitz (1982).

Following the steps of Figure 1, the searcher needs a user facility which provides access to the online systems via telephone communications so that he or she can run searches, use the other features of the online system, and download citations. This need can be met by any of a large number of commercial and public-domain "terminal-emulation programs" which can run on the searcher's local computer.

In addition to this basic capability, it is desirable that the various online database systems be accessible through one common language for specifying searches, downloading citations, and manipulating the resulting files on his local computer. It is possible to obtain this common interface with a program running on the user's local computer. CONIT, Marcus (1982), is

the only available program which meets this need adequately, but CONIT does not seem to be transportable from its present, single (large computer) installation. IN-SEARCH is a commercial offering which works with a limited number of databases and is not user-extendible.

A large number of citations can result from a single search. It is sometimes a large task to classify these citations as to relevancy or sub-category, e.g., Marcus, (1978). The TIS GATEWAY, Hampel, et. al., (1982), offers some capability in this area in the form of tools for manipulating the abstracts and classifying them. In addition, the TIS GATEWAY also provides a limited analysis of citations by time of publication. However, no commercial automatic or semi-automatic programs are available.

Report organization and preparation is a major activity. In the area of writing aids, the offerings on the personal computers have out-paced anything before seen. Outlining, organizing, and word-processing tools abound. The searchers should certainly have this capability. There is the necessity of converting the files of citations as down-loaded from the on-line system into a format which can be included into the word-processor. Some word-processors have facilities for doing this; others do not, and it must be programmed. Ultimately, work in this area may lead to some type of automatic presentation of search results.

2. REVIEWING AND DETERMINING THE RELEVANCY OF CITATIONS

2.1 SORT-AID Post Processor

As noted in Sections 1.2 and 1.3, categorizing and evaluating citations is a labor intensive task. It is unusual for more than 30-40% of the citations retrieved in an on-line search to be directly relevant to the problem being investigated. The SORT-AID software system has been developed to assist the searcher in categorizing and evaluating citations. The SORT-AID system is designed to be useful in that interim stage in report preparation after the database querying and searching have been completed and before the actual writing process begins. Accordingly, SORT-AID provides facilities for classifying, manipulating, and organizing the abstract/citation files resulting from online database searching.

The SORT-AID system was originally developed for use on the DEC VAX 11-780 and was programmed in FORTRAN, i.e., SORT-AID--Version 1.0. The software system was subsequently redesigned for operation on an IBM PC or compatibles and was programmed in Turbo Pascal, i.e., SORT-AID--Version 2.0. While the latter version maintained all of the functionality of the earlier FORTRAN editions, computational times were judged to be excessive. Consequently, the software system was redesigned for a second time and programmed to run in both Aztec C and Unix C, i.e., SORT-AID--Version 3.0. Both of these versions are very effective in terms of computational times and various benchmark data is presented in Section 3. The functionality of the program has

been maintained and/or improved through all of the redesign phases.

The SORT-AID software system consists of four separate modules. Each of these is discussed in the following sections.

2.2 Program NABST

Figure 3 shows the system flow for the SORT-AID system. Input is the individual files resulting from multiple database searches. These files are combined by the NABST program. After the combined file is created by NABST, it can be accessed immediately by the user with the REVIEW program. Optionally, an automatic or semi-automatic relevance ordering can be created for the combined citation file with the RANK program. After the abstract/citations in a combined file have been classified by the user into report-specific categories, they can be printed according to those categories with PRINT. NABST must be run for each abstract/citation file which has been downloaded from a bibliographic database system.

In addition to the abstract/citation file, NABST requires for input a "DELIM.DAT" file. This file is created with the system editor and contains images of the delimiters which may appear in the input abstract/citation files to mark the beginnings of abstract/citations. Furthermore, in addition to the combined abstract/citation file, NABST creates a lexical statistics file which is used by RANK.

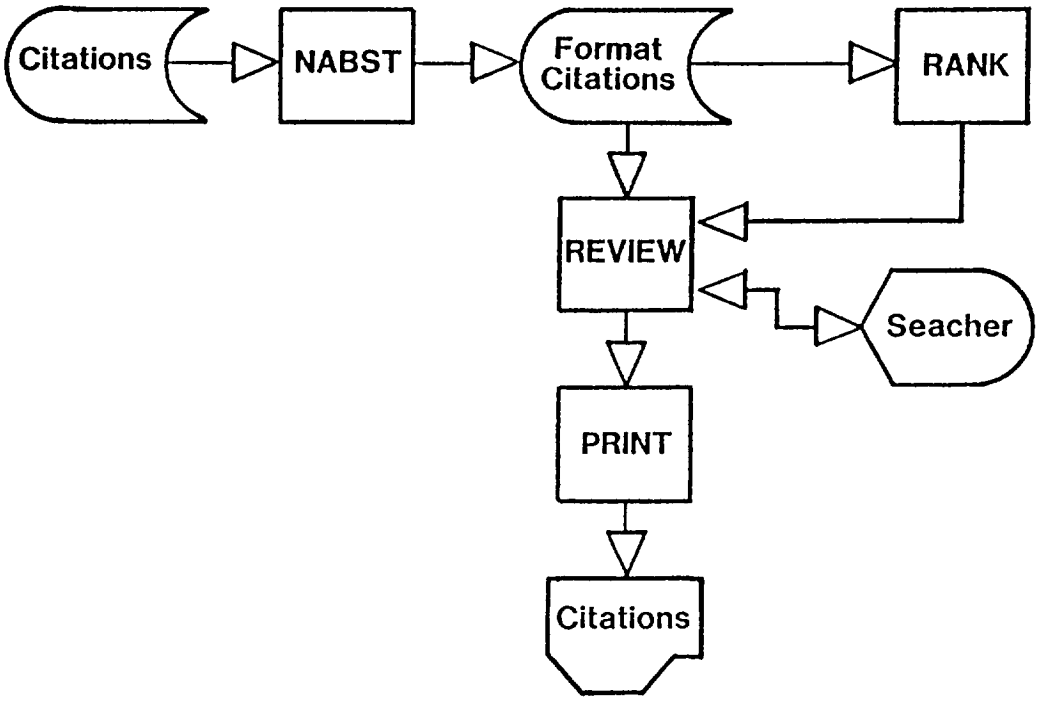


Figure 3. System Flow of SORT-AID.

2.3 Program REVIEW

REVIEW reads the combined abstract/citation file and allows the engineer/searcher to review the contents in its created order or according to a "relevance" order as determined by the RANK program. Optionally the abstracts are presented in the format they were received in (at the request of the users of the system).

REVIEW is command-driven with the following facilities offered:

i. Proceed to next screen, which will be more of a multi-screen entry, or to the next entry.

ii. Set abstract category, which allows the operator to assign a report-specific category or categories to be carried in the file with the citation. This category can be any character string without spaces or commas. Multiple categories are entered with commas in between and no spaces.

iii. Go directly to the next abstract.

iv. Go to the beginning of the current abstract.

v. Go directly to an abstract in the file addressed by its relative position in the file.

vi. Back up one abstract in the file.

vii. Search on the occurrence of an entered string. This is useful for finding entries containing a particular word, phrase, or author name, for example.

viii. Set search category; this allows the user to only see entries which have been previously set to a certain category or categories; "*" shows all categories, including un-categorized entries.

ix. Delete an entry from the file.

x. Enter notes into the file as a discrete entry; this entry can be categorized and processed like the others.

xi. Re-order an entry; this may be used to place a particularly interesting entry at the beginning of the file; or an uninteresting one at the end; this is done for the physical and the "relevance" orders.

xii. Load an internal memory register with the relative position of the current abstract.

xiii. Return to the abstract pointed to by the internal memory register.

xiv. Stop processing.

REVIEW is designed to allow the searcher to review and categorize abstracts until he has enough for his report, at which time he can stop processing. It also allows categorization to proceed iteratively via stepwise refinement; that is, the searcher can go through the file once applying gross categories, and then begin at the beginning looking at only a single gross category, applying finer categories.

2.4 Program RANK

RANK may be executed after the combined abstract file is completely created by NABST. RANK creates a file which contains a record for each abstract in the combined file. Each record contains the relative position of its associated abstract. These records are ordered in decreasing order by a "relevance" score.

To create the "relevance" score, RANK produces three sets of one hundred words each and displays them to the user. The user then specifies the relevance of each word and RANK calculates a "relevance" score.

In addition to the combined citation file, RANK requires a "stop list" file as input. This file contains common words which

cannot have report-specific meaning such as "a", "an", "and", and so forth. This file is created with the system editor.

2.5 Subprogram PRINT

PRINT is used to include abstracts of particular categories in a report and is a subprogram within REVIEW. The format of the abstracts may be changed, i.e., Chen (1985), and the page length and width may be specified. The abstracts are included in the "relevance" order or in physical order. PRINT is invoked by a command in the REVIEW program. Selection of abstracts for inclusion in the PRINT report follows the conventions set up in REVIEW.

3. PROGRESS TO DATE

3.1 Benchmark Comparisons for C and Pascal

As noted in Section 2.1, the SORT-AID software system has been redesigned twice. The first redesign converted the program from a minicomputer to a microcomputer environment (Version 2.0). The second redesign adapted the software architecture from a Pascal to a C environment (Version 3.0). The latter change had major performance impacts which are illustrated in Figure 4. Figure 4 compares the Pascal version of SORT-AID, i.e., denoted as P (Version 2.0), and the C version, denoted as C (Version 3.0), for the Kaypro PC and the Zenith 248, denoted as KP and Z248, microcomputers. The maximum improvement in computation times occurs for the 164 citation set with the Kaypro PC registering a 32% improvement and the Zenith 248 a 48% improvement. Not all of the improved performance can be traced to the different programming languages since software architecture was also optimized in the redesign and coding.

3.2 Benchmark Comparisons for Various Microcomputers

The performance of the SORT-AID software system--version 3.0 programmed in either Aztec or Unix C--has now been evaluated on a number of single user and multiuser microcomputers. The results of these tests are listed in Table 2. This table delineates the computer, the CPU processing speed, the disk access time, the number of citations, the total number of words analyzed in the citation set, the number of unique words in the citation set and

the computation time in minutes for the NABST and RANK software modules. Note that REVIEW operates in the interactive mode, e.g., see Section 2.3, and was thus excluded in the benchmark testing. The various data sets for the benchmark tests are shown in bar chart form in Figures 5 and 6.

Figure 7 shows the variation of computational times as a function of the number of citations for the Harris MCX microcomputer. As can be seen, the computational times vary non-linearly for small sets of citations and approach linearity for larger sets. Moreover,

$$t = f(n^2) \qquad n \leq 50$$

$$t = f(n) \qquad n > 50$$

where t denotes time in minutes and n the number of citations. The words processed by rank, e.g., see Table 2, vary in the same manner and the computational times are very nearly linear with regard to words processed. This is illustrated in Figure 8. Note that all of the computers tested demonstrate the same behavior.

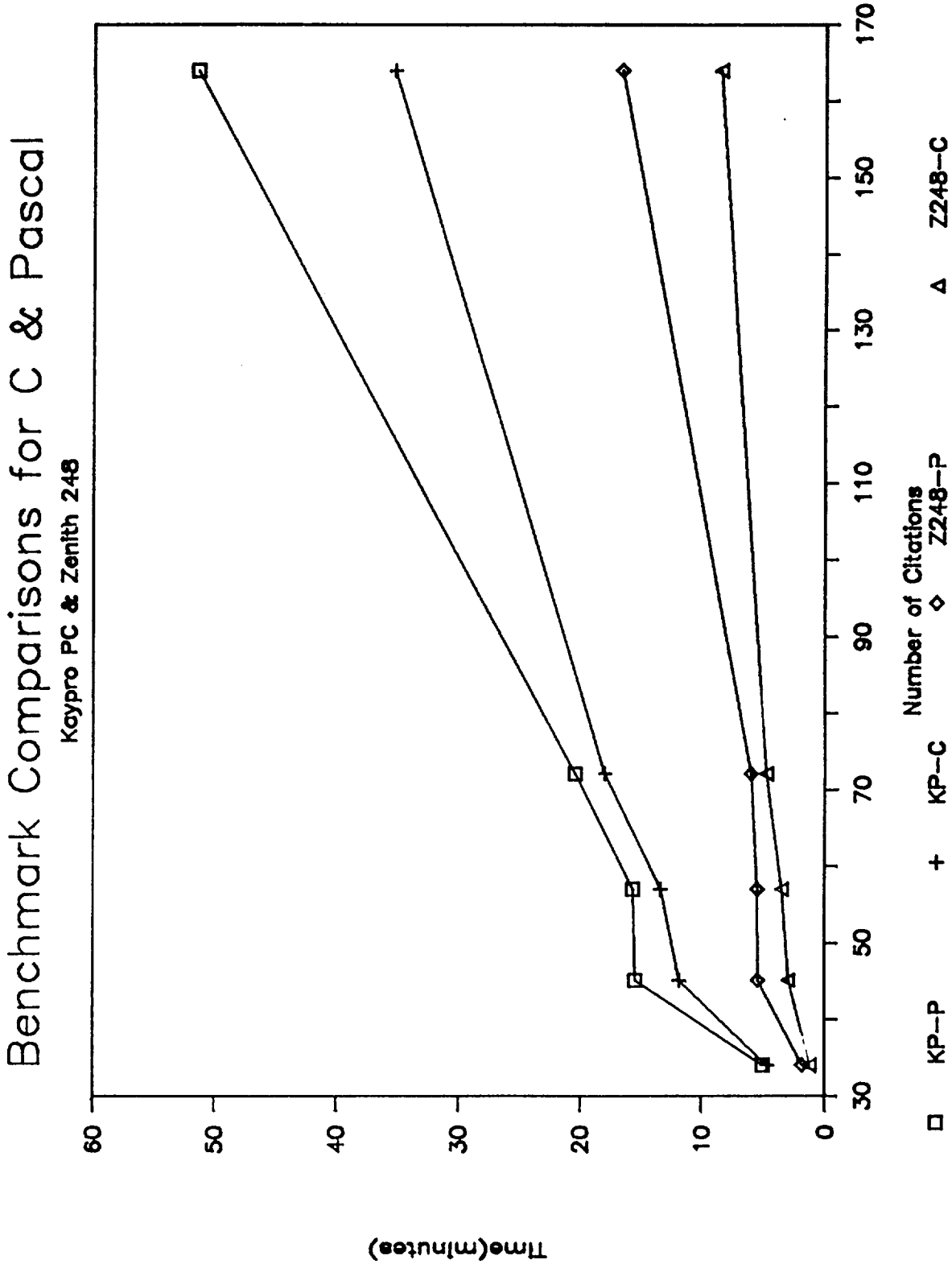


Figure 4. Benchmark Comparisons for C and Pascal.

TABLE 2

COMPARISON OF SORT-AID (VERSION 3.0) OPERATING TIMES ON SELECTED MICROCOMPUTERS

Computer	NABST (mins)	RANK (mins)	Total (mins)	No. Cit.	Unique Words	Total Words	CPU (MHz)	Disk Time (ms)
Harris MCX	0.05	0.22	0.27	34	459	1260	16.70	38.30
	0.18	0.68	0.87	45	1084	3700		
	0.18	0.73	0.92	57	1280	4023		
	0.23	0.98	1.22	72	1775	5090		
	0.73	1.98	2.72	164	2729	13266		
IBM RT	0.08	0.42	0.50	34	459	1260	6.00	40.00
	0.30	1.05	1.35	45	1084	3700		
	0.37	1.23	1.60	57	1280	4023		
	0.37	1.67	2.03	72	1775	5090		
	1.02	3.13	4.15	164	2729	13266		
Honeywell XPS-100	0.05	0.32	0.37	34	459	1260	16.70	38.30
	0.17	0.87	1.04	45	1084	3700		
	0.20	1.02	1.22	57	1280	4023		
	0.25	1.32	1.57	72	1775	5090		
	0.63	2.32	2.95	164	2729	13266		
IBM RT (new)	0.03	0.22	0.25	34	459	1260	10.00	40.00
	0.10	0.53	0.63	45	1084	3700		
	0.13	0.58	0.72	57	1290	4023		
	0.15	0.80	0.95	72	1775	5090		
	0.40	1.50	1.90	164	2729	13266		
Compaq 386	0.13	0.63	0.77	34	459	1260	16.00	30.00
	0.47	1.52	1.98	45	1084	3700		
	0.47	1.73	2.20	57	1280	4023		
	0.57	2.38	2.95	72	1775	5090		
	1.50	4.30	5.80	164	2729	13266		
Zenith 248	0.20	0.95	1.15	34	459	1260	8.00	35.00
	0.60	2.30	2.90	45	1084	3700		
	0.80	2.63	3.43	57	1280	4023		
	0.80	3.57	4.37	72	1775	5090		
	2.03	6.52	8.55	164	2729	13266		
IBM AT	0.33	1.58	1.92	34	459	1260	6.00	40.00
	1.07	3.88	4.95	45	1084	3700		
	1.10	4.43	5.53	57	1280	4023		
	1.33	5.93	7.27	72	1775	5090		
	3.57	11.02	14.58	164	2729	13266		

TABLE 2 (continued)

COMPARISON OF SORT-AID (VERSION 3.0) OPERATING TIMES ON SELECTED MICROCOMPUTERS

Computer	NABST (mins)	RANK (mins)	Total (mins)	No. Cit.	Unique Words	Total Words	CPU (MHz)	Disk Time (ms)
Zenith 148	0.57	2.58	3.15	34	459	1260	8.00	85.00
	1.77	6.12	7.88	45	1084	3700		
	1.97	6.97	8.93	57	1280	4023		
	2.27	9.43	11.70	72	1775	5090		
	5.93	17.60	23.53	164	2729	13266		
Kaypro PC	0.83	3.80	4.63	34	459	1260	4.77	65.00
	2.40	9.15	11.55	45	1084	3700		
	2.97	10.48	13.45	57	1280	4023		
	3.53	14.43	17.97	72	1775	5090		
	9.27	25.93	35.20	164	2729	13266		
Z148- 4.77 MHz	0.87	4.10	4.97	34	459	1260	4.77	85.00
	2.73	9.70	12.43	45	1084	3700		
	3.10	11.33	14.43	57	1280	4023		
	3.57	15.22	18.78	72	1775	5090		
	9.43	28.12	37.55	164	2729	13266		

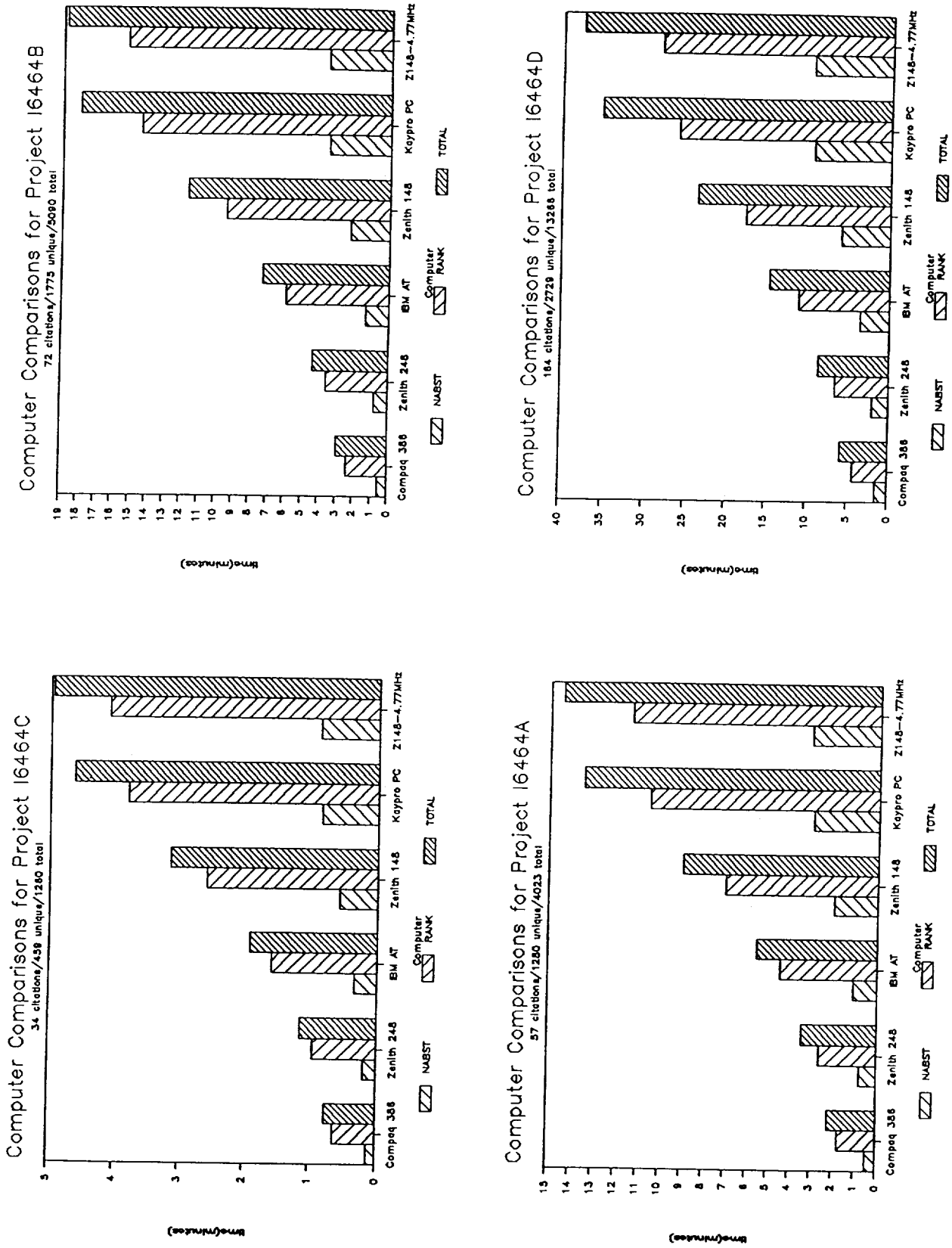


Figure 5. Benchmark Comparisons for MS-DOS Computers.

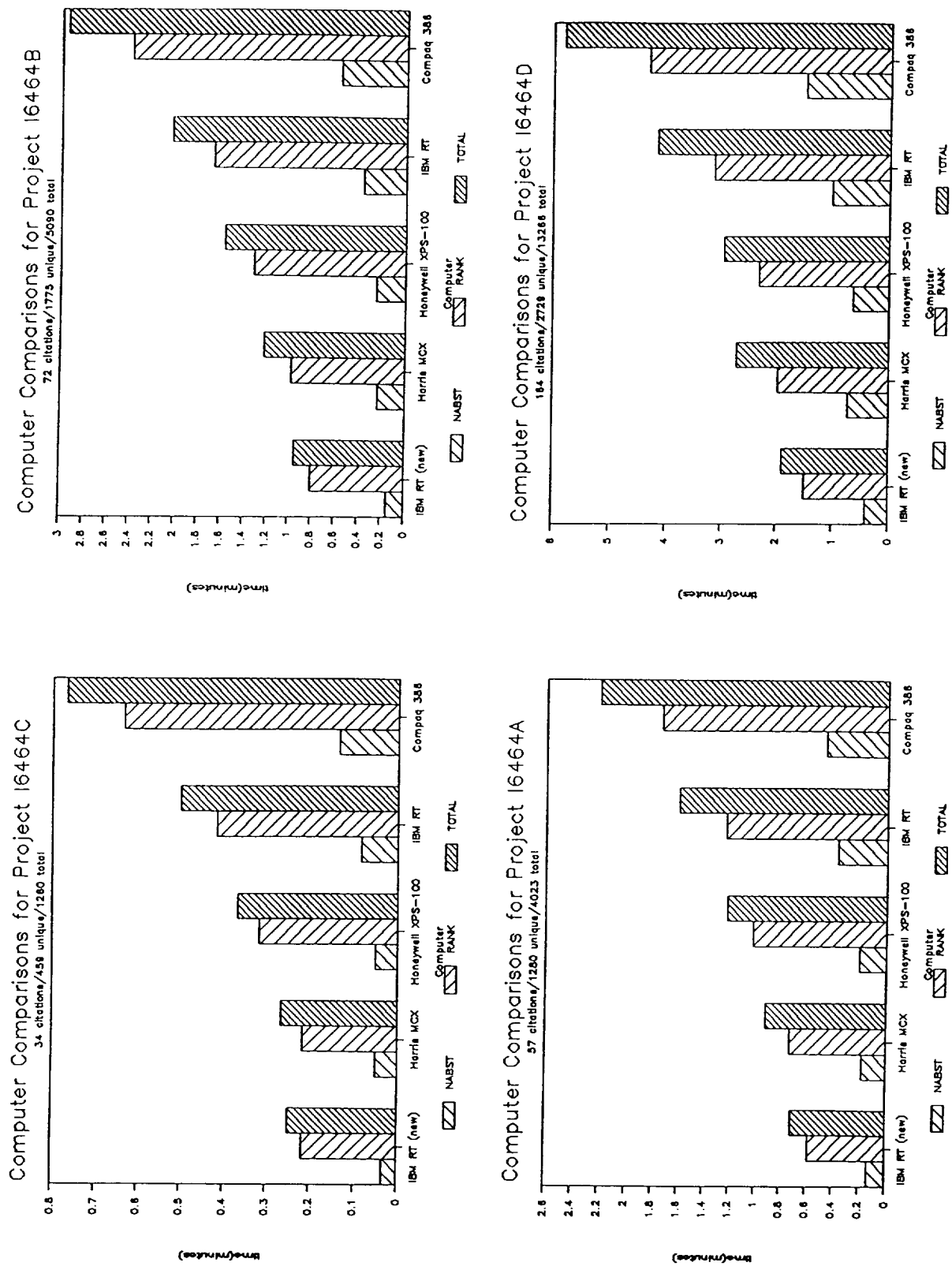


Figure 6. Benchmark Comparisons for UNIX Computers.

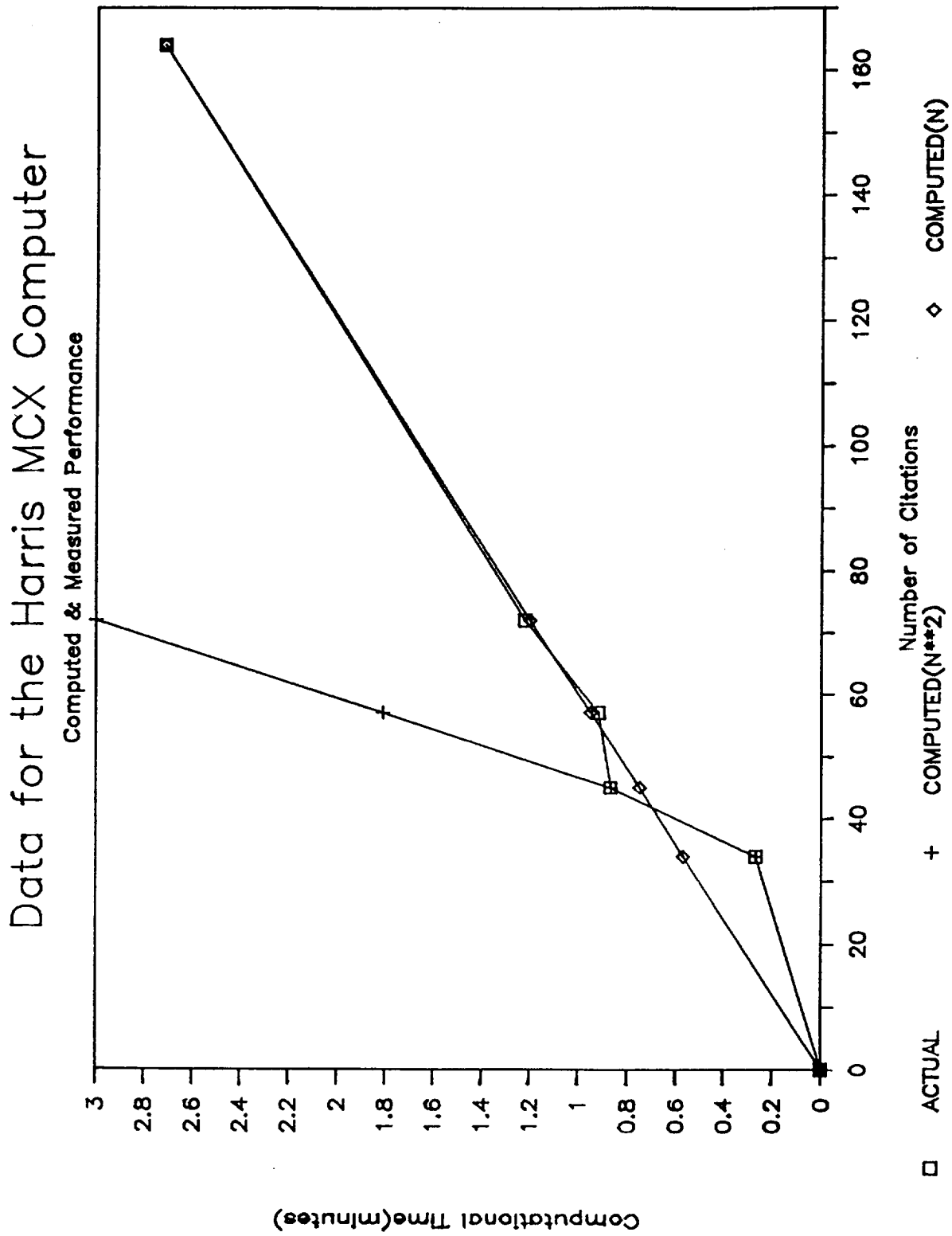


Figure 7. Computational Time versus the Number of Citations.

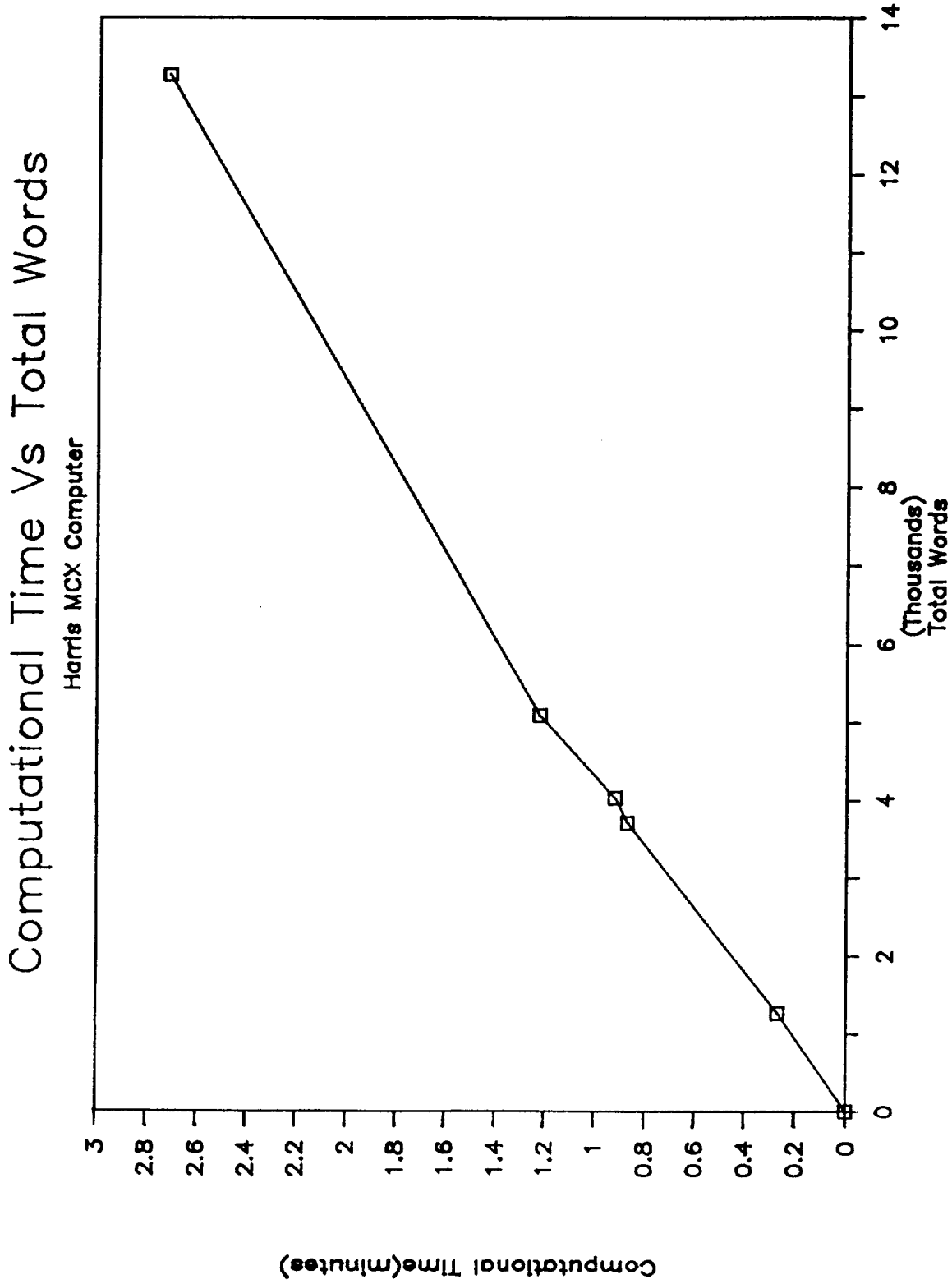


Figure 8. Computational Time versus the Number of Words Processed.

3.3 Preliminary Benchmark Results with SORT-AID (Version 4.0)

As a result of some University-supported research activities, the disk data storage algorithms of SORT-AID (Version 3.0) were redesigned. While these changes have not been fully evaluated, preliminary performance results are impressive. A limited comparison of execution times is shown in Table 3.

Note that the execution time per citation varies with the size of the citation set being analyzed. The values tend to decline as the number of citations increases reaching a fairly constant value for sets with 150 or more citations. Table 3 shows a 48% reduction in sec/citation for the Zenith 248 for Version 4.0 as compared to Version 3.0 with the Honeywell XPS-100 demonstrating a 34% reduction in operating times.

TABLE 3

SORT-AID (VERSION 4.0) OPERATING TIMES
ON SELECTED MICROCOMPUTERS
(PRELIMINARY RESULTS)

	Version 3.0				Version 4.0				
Computer	NABST (mins)	RANK (mins)	Total (mins)	Sec/ Cit	NABST (mins)	RANK (mins)	Total (mins)	Sec/ cit	No cit
Zenith 248	7.50	19.88	27.38	3.5	3.63	10.62	14.25	1.8	475(1)
Honeywell XPS-100	2.42	6.82	9.24	1.2	1.27	4.80	6.07	0.8	475(1)
Zenith 248(2)	2.03	6.52	8.55	3.1	1.06	3.39	4.45	1.6	164
IBM RT new(2)	0.40	1.50	1.90	0.7	0.26	0.97	1.23	0.4	164
Compaq 386(3)	1.50	4.30	5.80	2.1	0.73	2.30	3.03	1.1	164

(1) 7149 unique words and 45446 total words.

(2) Zenith 248 and IBM RT (new) Version 4.0 data scaled from Zenith 248 and Honeywell XPS-100 Version 3.0 and Version 4.0 data using total words.

(3) Compaq 386 Version 4.0 data scaled from Zenith 248 Version 4.0 data.

3.4 Summary of Benchmark Tests

The computer systems used in the benchmark tests vary over the entire spectrum of microcomputers in both cost and performance. Some typical performance metrics for the computers are given in Table 4. For a given operating system, i.e., MS-DOS, the computational times correlate fairly well with disk access time. This is illustrated in Figure 9. The operating system has a substantial impact on the computational times. The Harris MCX, Honeywell XPS-100 and the IBM RT employ a multiuser Unix-based operating system. As can be seen from a comparison of Tables 2 and 4, the computational times are far less than would be predicted by either CPU cycle speed or disk access times.

Computer prices vary with the retail outlet, peripherals and options and discounts. As a result, specific prices for the benchmark computer are not presented. Price ranges can however be stated. The Kaypro PC and the Zenith 148 are in the $1,000-$2,000 range; the IBM AT, the Zenith 248 and the Compaq 386 are in the $3,000-$5,000 range; the IBM RT in the $13,000-$16,000 range; and the Harris MCX and Honeywell XPS-100 in the $20,000-$25,000 range. Note that the IBM RT, the Honeywell XPS-100 and the Harris MCX will support multiple users. The number of users varies from approximately 4 to 8 depending on the specific configuration.

TABLE 4

BENCHMARK COMPUTER METRICS

Computer	CPU	CPU Cycle Rate	Average Disk Access Time
Kaypro PC	8088	4.77 MHz	65 milli-sec.
Zenith 148	8088	4.77/8 MHz	85 milli-sec.
IBM AT(1)	80286	6 MHz	40 milli-sec.
Zenith 248	80286	8 MHz (no wait state)	35 milli-sec.
Compaq 386	80386	16 MHz	30 milli-sec.
IBM RT(2)	not avail.	6 MHz	40 milli-sec.
Harris MCX(3)	68020	16.7 MHz	38.3 milli-sec.
Honeywell XPS-100	68020	16.7 MHz	28 milli-sec.
IBM RT (new)	not avail.	10 MHz	40 milli-sec.

(1) IBM AT:
hard disk drive capacity 20/30 meg.

(2) IBM RT:
transfer rate 5 million bits/sec.
hard disk drive capacity 40 meg.

(3) Harris MCX 70 Mb disk:
head positioning (track to track) 8 milli-sec.
maximum positioning time 60 milli-sec.
latency time 8.3 milli-sec.

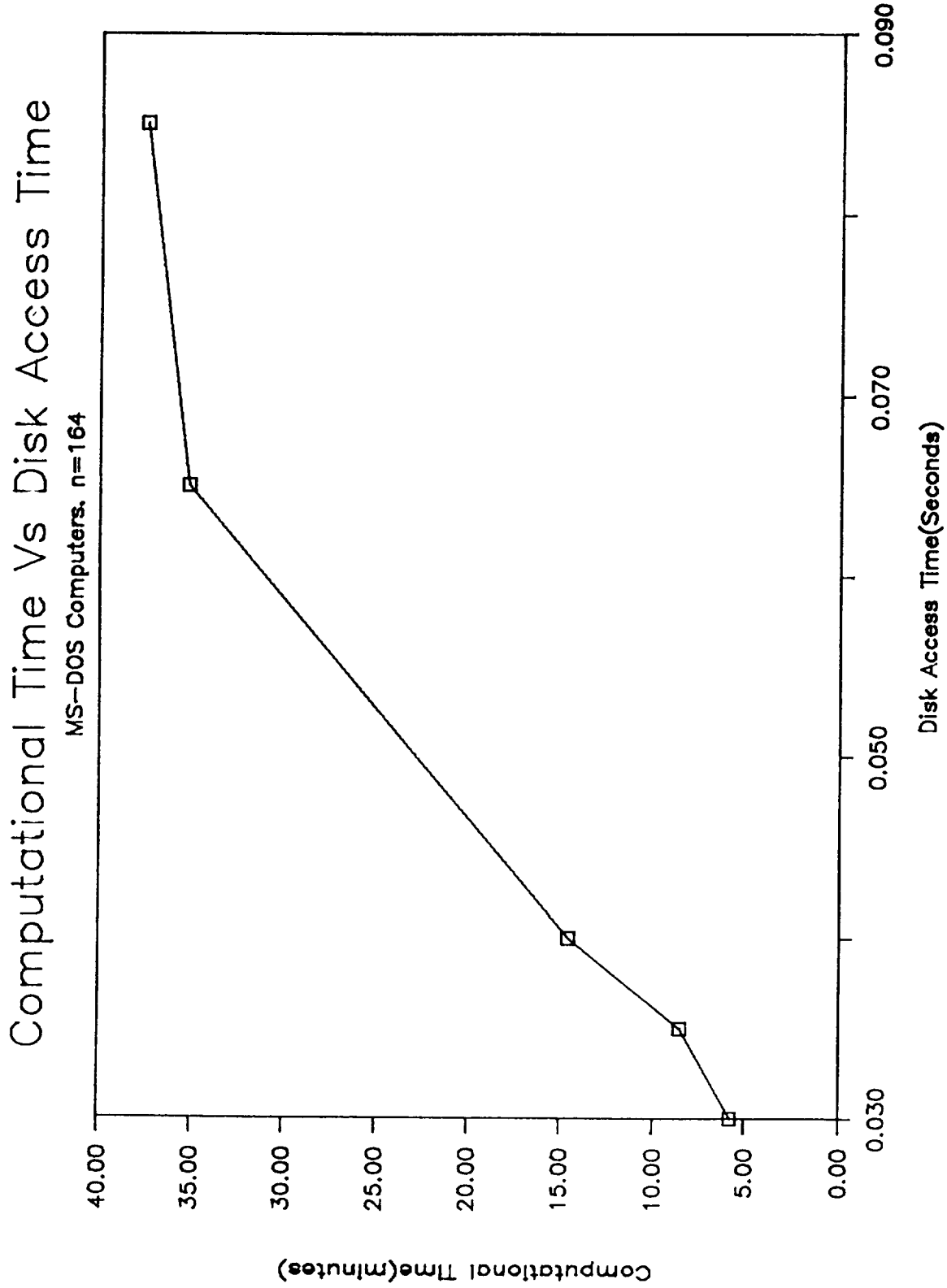


Figure 9. Computation Times as a Function of Disk Access Times for MS-DOS Computers. 164 Citations.

It is important to note that the same version of SORT-AID, i.e., Version 3.0 or 4.0, was tested on both MS-DOS and UNIX computers. The IBM RT, the Honeywell XPS-100 and the Harris MCX support substantial amounts of main memory. SORT-AID is configured for operation on microcomputers having 524 K of main memory. Large numbers of citations mandate a large number of data transfers between disk and main memory. While this could be largely eliminated on the IBM RT, Honeywell XPS-100 or Harris MCX, no software architecture changes were implemented to utilize the full memory capacity of these computers. Alterations of this type would undoubtedly improve the already impressive performance of these computers.

The data presented is sufficient to predict the performance of SORT-AID on different classes of computers and for different numbers of citations. Computational times tend to vary linearly with the number of citations and relationships of this type are suitable for predicting computational times for an arbitrary number of citations.

In summary, the computational times for NABST and RANK have been substantially reduced in SORT-AID (Version 3.0 and 4.0). Computational times for the IBM RT (new) are approximately 0.4 sec/citation, 1.1 sec/citation for the Compaq 386 and 1.5 sec/citation for the Zenith 248. The values represent a 10-to-50 fold reduction over the execution times for SORT-AID (Version 2.0) operating on an IBM XT or compatible, i.e., 18.3 sec/citation. These reductions substantially enhance the utility

of the software--a searcher can process 200 citations in 1.5 to 5.5 minutes dependent upon operating environment.

3.5 User Evaluation of RANK

As described in Section 2.4, the RANK module provides a thesaurus of terms from the citations which can aid the user in categorizing and ordering the citations. In an effort to quantify the effectiveness of RANK, the University of Southern Mississippi's School of Library Science is now completing a RANK evaluation program. Over one hundred projects consisting of multiple database searches have now been conducted. The searches were conducted by faculty members of the School of Library Science with technical experts drawn from the faculty of the University. All project searches were conducted on-line using either a Kaypro PC or Zenith 248--donated by Zenith, SORT-AID (Version 2.0) and SMART-COM. The search procedures followed the outline presented by Huffman (1987). The preliminary results are described in the following paragraphs.

To date, 103 projects have now been completed through the search phase. Of these, 77 have been completed through the user-supplied relevance phase and analyzed. As noted above, the searches were conducted on-line using the NASA RECON, Dialog, BRS, and Orbit databases. Both the searcher and technical expert were present during the on-line search with the technical expert supplying the relevance factors for the characteristic words generated by RANK, see, e.g., Huffman (1987). The citations resulting from the search were down-loaded to the microcomputer

and printed in the as-received order. The technical expert then evaluated all citations for relevance using a 1-to-10 scale, i.e, 1 not relevant and 10 relevant. The RANKed citation order was stored on removable disks and was not available to either the searcher or technical expert. The citation ordering was then analyzed using the RANKed order, the user-supplied relevance and the theoretical framework described in the following section.

3.6 Analysis Framework for Weakly-Ordered Distributions

As noted above, the analysis of RANK is built on the user-supplied relevance values. The determination of relevance is, itself, subject to some uncertainty, see, e.g., Saracevic (1975) and Robertson (1977). In the present study, the technical experts were instructed to base the relevance determination on the fraction of the citation that appeared to be relevant, i.e., 1 for no relevance, 10 for complete relevance and 2 through 9 for intermediate fractions of relevance.

RANK is a relevance ordering system and changes neither the precision nor recall of the original search. The intent is to concentrate the relevant documents at the beginning of the collection, i.e., weakly-order the citations from most relevant to least relevant, e.g., Copper (1968). This ordering reduces the citation review process and can be viewed as a partial-automation of the task.

The ordering is carried out by dividing the collection of citations into groups, e.g., 70 citations can be divided into 7 groups with 10 citations per group. The user-supplied relevance

is converted into a fraction relevant Z_j via

$$Z_j = \frac{1}{G_j} \sum_{g=1}^{G_j} \left(\frac{r_g}{10}\right) \quad j=1,J \qquad [1]$$

where r_g is the relevance value, 1 to 10, for the gth document within the group, G_j is the number of documents per group, j is the group number and the factor 10 normalizes the values to 1. j=1 represents the first group in the collection and j=J represents the last group in the collection with the weak-ordering determined by RANK.

The weak-ordering efficiency can be determined by computing the deviations between the best, worst, random, and RANK (actual) ordering. The user-supplied relevance values can be used to determine the precision for the complete citation collection, i.e.,

$$R = \sum_{i=1}^{I} \left(\frac{r_i}{10}\right) \quad i=1,I \qquad [2]$$

where I denotes the total number of citations in the collection and R the equivalent number of relevant documents. The R value can be used to construct the best, worst and random distributions by placing the R documents first in the collection, last in the collection and distributed uniformly throughout the collection.

Consider the following example. A citation set consists of 90 documents divided into 9 groups of 10 citations each. The user-supplied r_i values amount to 31, i.e., R = 3.1. The "best"

distribution has Z values of 1 for groups 1 and 2, 0.5 for group 3 and 0.1 for groups 4 through 9. The "worst" distribution has values 0.1 for groups 1 through 6, 0.5 for group 7 and 1 for groups 8 and 9. The "random" distribution has a Z value of 0.35 for all groups. These distributions are shown graphically in Figures 10 and 11.

The "actual" distribution must be bounded by the best and worst cases. The ordering efficiency can be defined as the deviation between the best, worst, and actual cases, i.e.,

$$\sigma^{(b-w)} = \left\{ \frac{1}{J} \sum_{j=1}^{J} \left(Z_j^{(b)} - Z_j^{(w)} \right) 2 \right\}^{\frac{1}{2}} \qquad [3]$$

$$\sigma^{(a-w)} = \left\{ \frac{1}{J} \sum_{j=1}^{J} \left(Z_j^{(a)} - Z_j^{(w)} \right) \right\}^{\frac{1}{2}} \qquad [4]$$

where σ denotes the mean-square deviation and the superscripts b, w and a denote the best, worst and actual distributions. A similar equation can be written for the random case and

$$\sigma^{(r-w)} = \left\{ \frac{1}{J} \sum_{j=1}^{J} \left(Z_j^{(r)} - Z_j^{(w)} \right) \right\}^{\frac{1}{2}} \qquad [5]$$

where the superscript r denotes random.

Since all distributions must have a σ value less than or equal to $\sigma^{(b-w)}$, an "efficiency" can be defined as

$$\eta_a = \frac{\sigma(a-w)}{\sigma(b-w)} \qquad [6]$$

$$\eta_r = \frac{\sigma(r-w)}{\sigma(b-w)} \qquad [7]$$

where $0 \leq \eta_a \leq 1$ and $0 \leq \eta_r \leq 1$.

Figures 10 and 11 show two hypothetical distributions, i.e., Z ACTUAL. The distribution of Figure 10 obviously approaches the best case and has an η value of 0.94. The distribution of Figure 11 approaches the worst case and has an η of 0.15. The random distribution has an η value of 0.59.

Equations [1] through [7] were used to analyze the test data which is described in the following section.

3.7 Test Results

The results of the evaluation of RANK are summarized in Table 5. Data is presented for both the frequency, f, and signal-to-noise, s/n, ordering, see Huffman (1987). The average ordering efficiency for the complete set of projects is 0.68 which represents a 23% improvement over a random distribution. The RANK distribution approaches the best case, i.e., $\eta_a > 0.8$, in 21% of the projects and is poorer than the random distribution in only 4% of the cases.

A few typical distributions are shown in Figures 12, 13, 14, and 15. Figure 12 depicts distributions with $\eta_a > 0.72$. Figure 13 shows a distribution with an η_a value of 0.66 which is

approximately the average value for the entire set of citations. Figure 14 depicts a distribution with an η value which approximates a random distribution, i.e., $\eta_a = 0.55$ and $\eta_r = 0.52$. Figure 15 shows one of the few cases where the RANKed order was poorer than random, i.e., $\eta_a = 0.47$ and $\eta_r = 0.57$.

In summary, the test results to date clearly indicate that RANK--with characteristic word evaluation by the technical expert--significantly improves the distribution of citations within the collection.

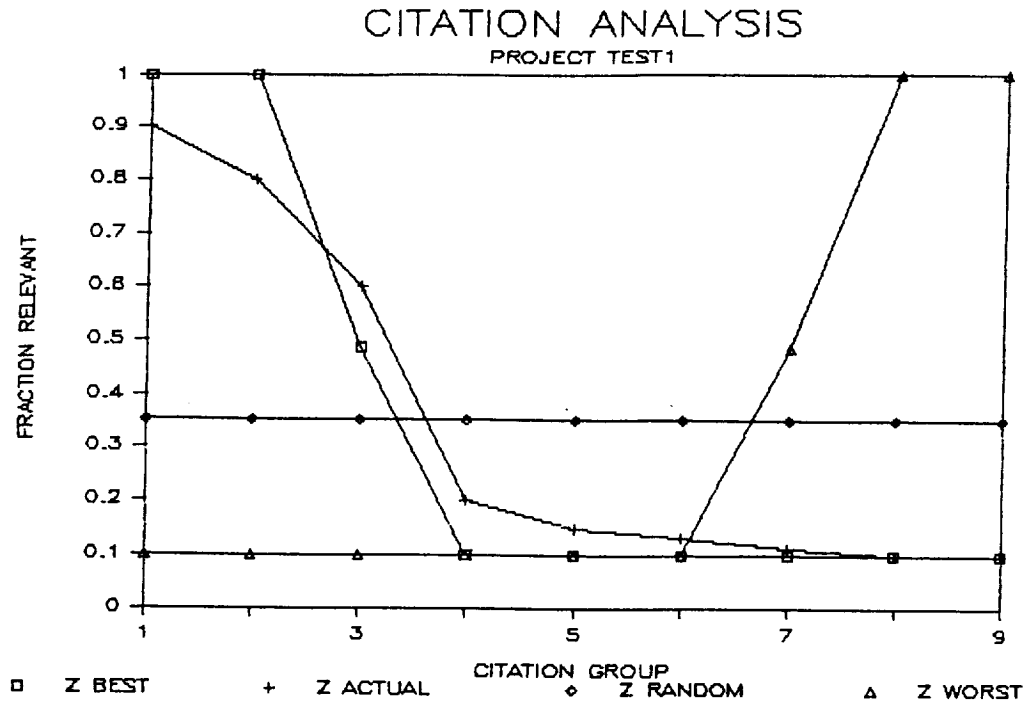


Figure 10. Hypothetical Relevance Distribution. Actual Approximates Best Case.

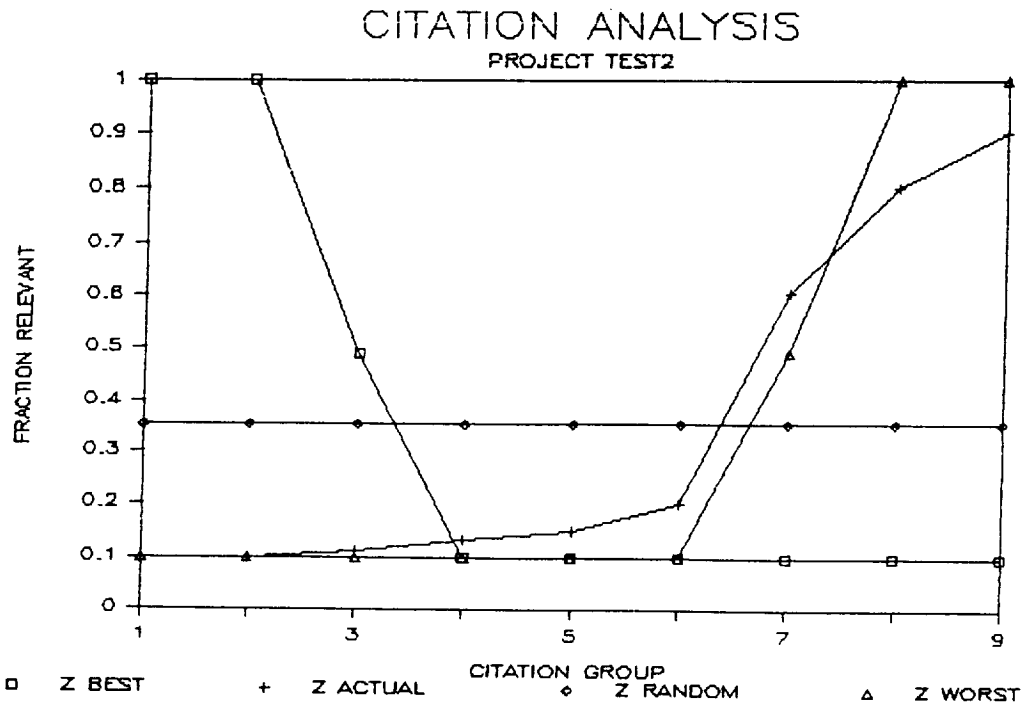
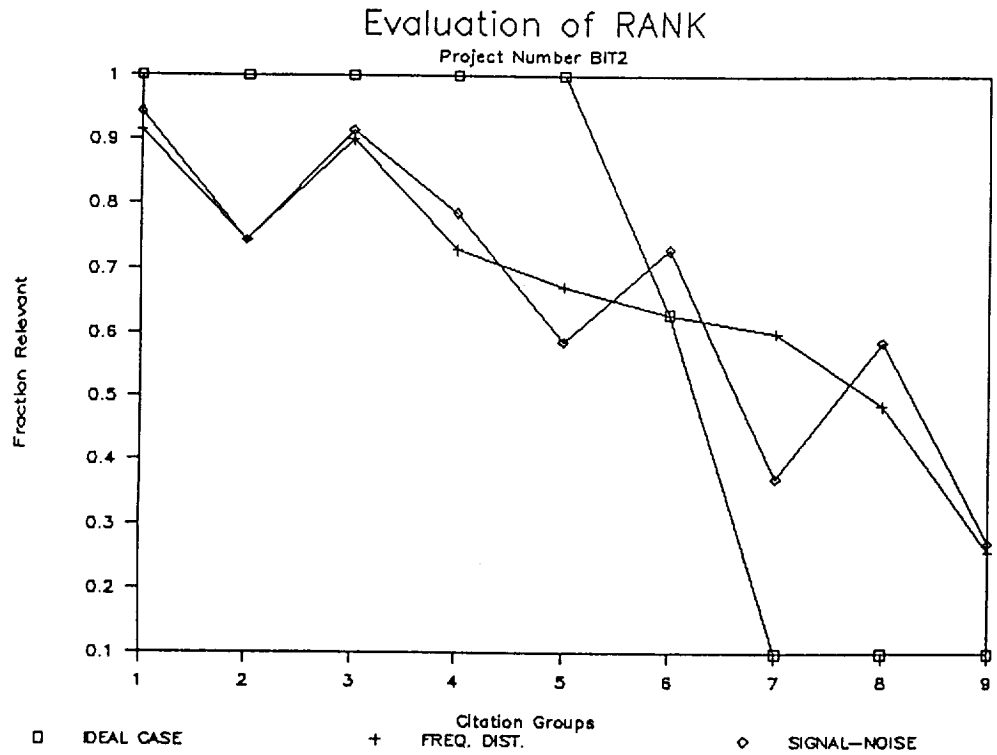


Figure 11. Hypothetical Relevance Distribution. Actual Approximates Worst Case.

TABLE 5
EVALUATION OF RANK

PROJECT	I	R	ORDERING EFFICIENCY			PROJECT	I	R	ORDERING EFFICIENCY		
			f	s/n	Random				f	s/n	Random
AGED2	119	65.6	0.66	0.71	0.50	JOR2	51	49.4	0.87	0.77	0.63
ALL2	82	28.8	0.68	0.64	0.59	JUL2	70	27.9	0.47	0.51	0.57
ANDE2	33	17.0	0.71	0.80	0.51	KENT2	101	62.8	0.69	0.64	0.52
AST2	21	15.8	0.58	0.59	0.58	KIL2	45	22.3	0.67	0.64	0.51
BESS2	25	14.7	0.61	0.62	0.50	LAN2	35	20.2	0.89	0.87	0.50
BET2	12	5.3	0.58	0.58	0.54	LAU2	7	1.9	0.99	0.88	0.62
BIT2	63	41.5	0.76	0.79	0.54	LEBO2	93	88.9	0.68	0.68	0.67
BLA2	85	43.5	0.55	0.52	0.52	LEO2	15	6.1	0.92	0.83	0.55
BRID2	41	22.5	0.72	0.84	0.50	LEW2	68	44.6	0.81	0.81	0.54
CAI2	40	37.5	0.70	0.70	0.66	LOT2	188	80.9	0.69	0.64	0.55
CART2	54	27.4	0.55	0.59	0.54	MCC2	34	20.0	0.61	0.56	0.50
CHA2	162	70.8	0.66	0.62	0.55	MEAS2	63	34.0	0.62	0.59	0.50
CHU2	99	58.8	0.60	0.64	0.51	MEL2	21	11.9	0.53	0.51	0.50
CORT2	5	3.1	0.68	0.34	0.52	MUS2	19	3.6	0.97	0.85	0.66
DAN2	24	10.0	0.85	0.85	0.56	NEIS2	22	19.8	0.89	0.89	0.65
DEAR2	200	144.4	0.68	0.67	0.58	NO2	69	11.1	0.73	0.72	0.66
DIP2	38	34.7	0.85	0.83	0.66	OKB2	23	5.9	0.87	0.85	0.65
DIX2	29	20.4	0.73	0.80	0.55	PAL2	13	6.2	0.63	0.55	0.55
DO2	85	39.4	0.89	0.80	0.54	PEE2	11	5.6	0.74	0.56	0.55
EMO2	6	2.8	0.96	1.00	0.52	PROS2	45	23.3	0.52	0.62	0.51
EPP2	30	8.6	0.82	0.82	0.61	RBC2	47	18.4	0.61	0.63	0.58
EPS2	19	12.0	0.54	0.55	0.51	READ2	40	5.3	0.66	0.75	0.66
ESS2	55	25.0	0.48	0.51	0.55	REOR2	32	11.2	0.66	0.65	0.59
EVER2	60	46.7	0.63	0.64	0.59	ROBE2	26	19.9	0.73	0.73	0.58
FIGG2	26	11.3	0.62	0.51	0.57	RON2	74	30.4	0.68	0.67	0.57
FULL2	40	21.0	0.60	0.57	0.50	SAM2	58	34.2	0.38	0.38	0.50
GAR2	75	36.6	0.53	0.53	0.53	SBOY2	22	12.2	0.58	0.53	0.50
GOOD2	35	23.6	0.54	0.52	0.54	SCH2	11	6.0	0.52	0.40	0.52
GRAY2	10	5.5	0.50	0.37	0.50	SLO2	33	21.4	0.59	0.59	0.51
GRU2	21	11.1	0.54	0.57	0.50	SMI2	69	24.9	0.64	0.63	0.58
HAM2	44	35.3	0.65	0.65	0.61	SOU2	48	31.8	0.63	0.61	0.54
HAS2	105	49.4	0.56	0.58	0.55	STO2	30	11.3	0.75	0.68	0.57
HIL2	31	20.5	0.53	0.53	0.53	TAST2	55	33.9	0.63	0.52	0.52
HORS2	13	12.1	0.55	0.59	0.64	TEST2	85	42.9	0.61	0.57	0.52
INT2	6	4.4	1.00	1.00	0.58	VAN2	42	24.3	0.58	0.54	0.51
JEAN2	14	8.4	0.73	0.60	0.50	VIRT2	10	5.6	0.78	0.42	0.50
JES2	13	9.1	0.83	0.83	0.55	WHIT2	40	22.0	0.52	0.56	0.50
JOAN2	88	61.0	0.60	0.59	0.56	WOOT2	30	15.2	0.72	0.77	0.51
JOHN2	12	4.5	0.88	0.92	0.57						
SUBTOT	1920	1110	26.12	25.81	21.46	SUBTOT	1745	917.3	26.06	24.60	21.15
TOTAL	3665	2027.40	52.18	50.41	42.61						
AVERAGE			0.68	0.65	0.55						
% IMPROVE-MENT OVER RANDOM			22.5	18.3							
PRECISION,%		55.3									



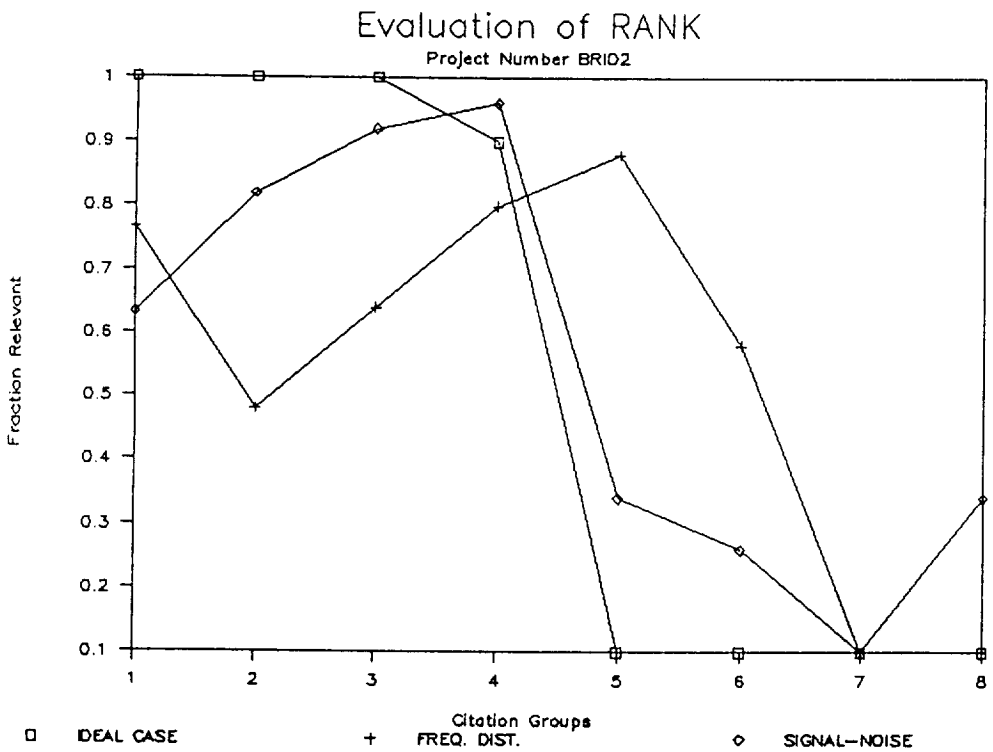
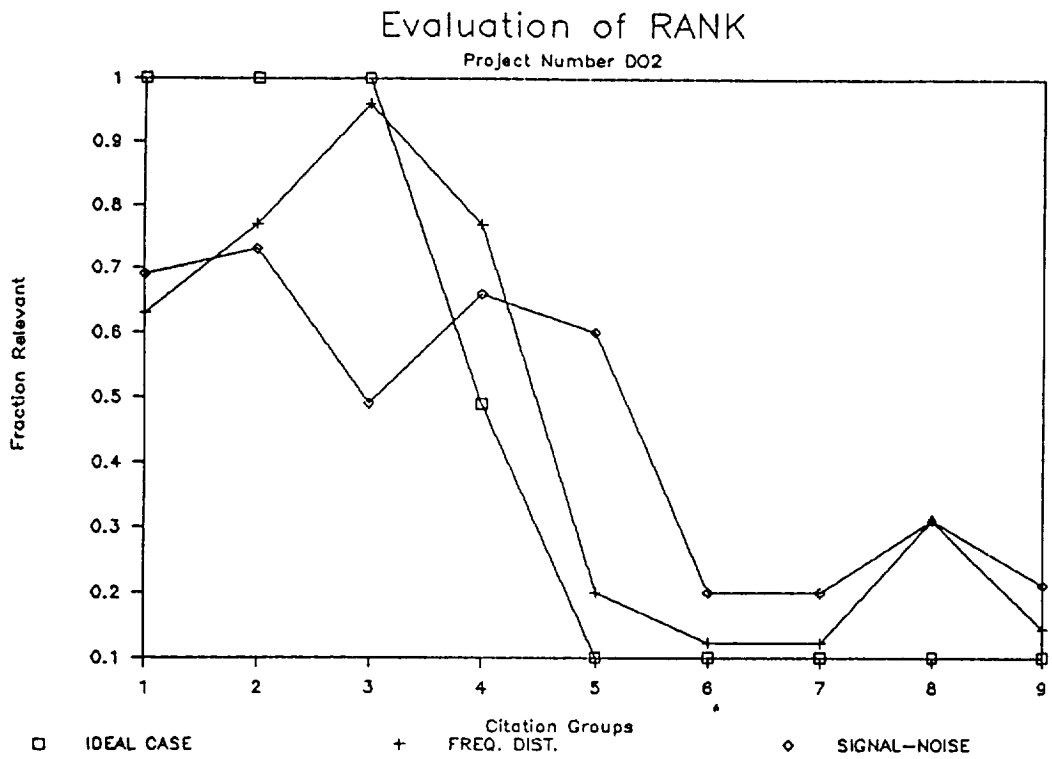


Figure 12. RANK Distributions. Typical Cases, $\eta_a > 0.72$.



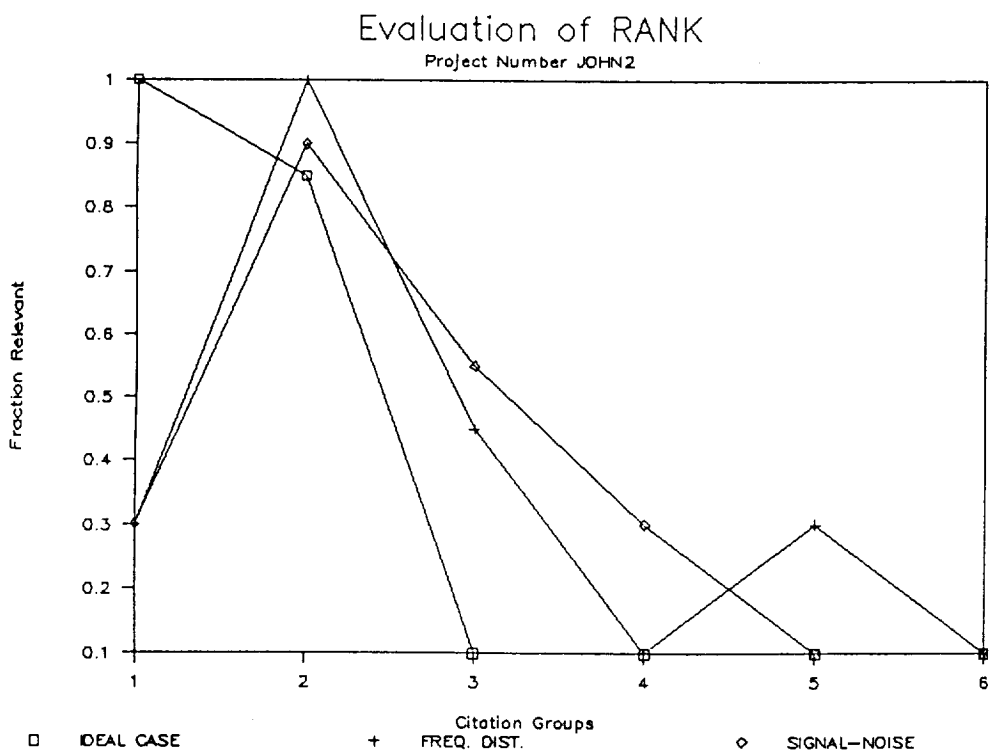
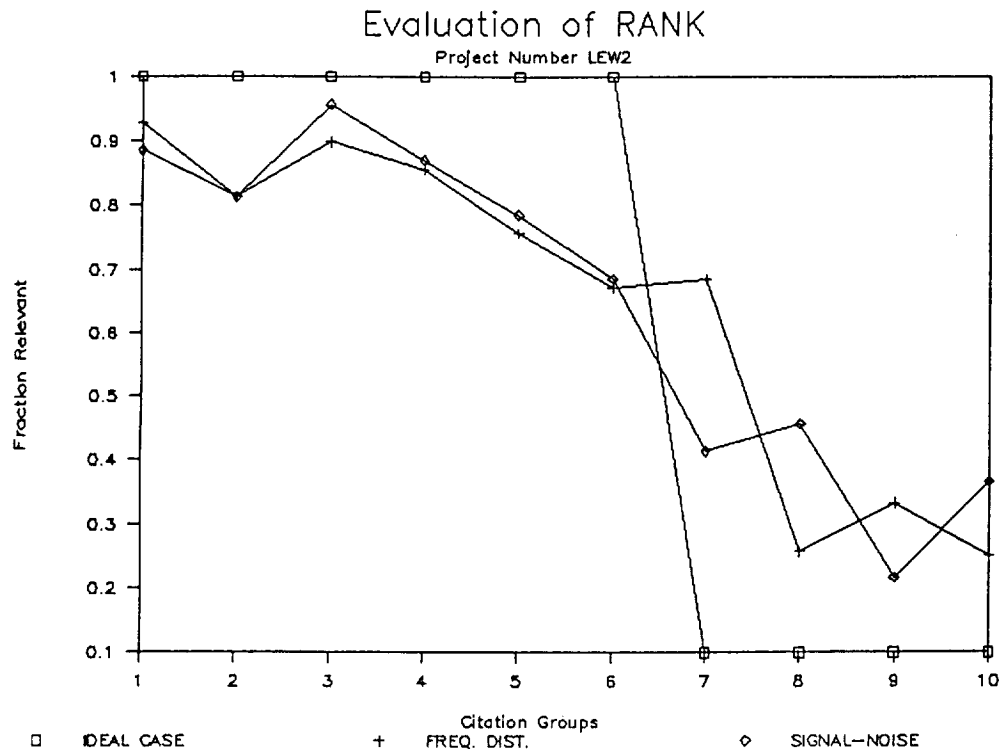


Figure 12 (continued).



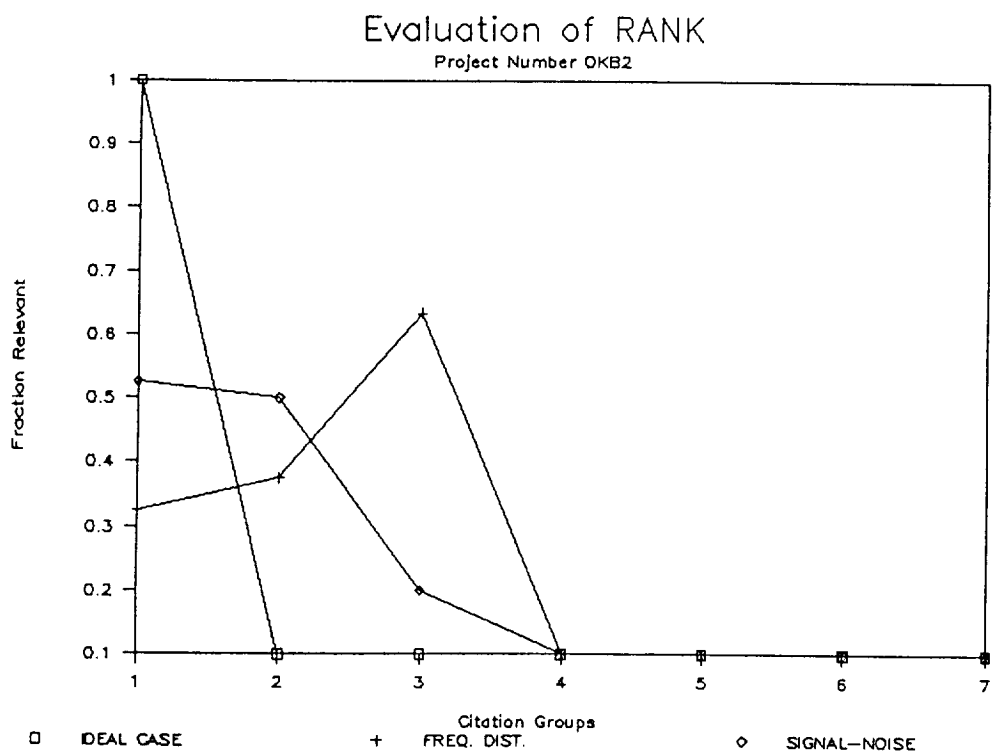


Figure 12 (continued).

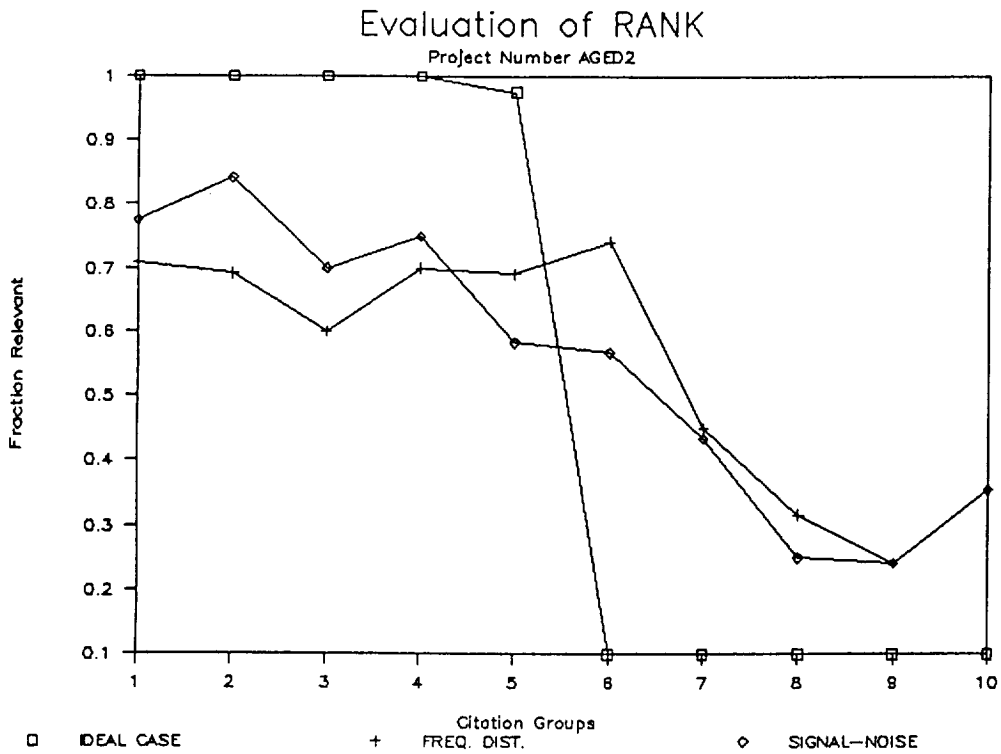


Figure 13. RANK Distribution. $\eta_a = 0.66$.

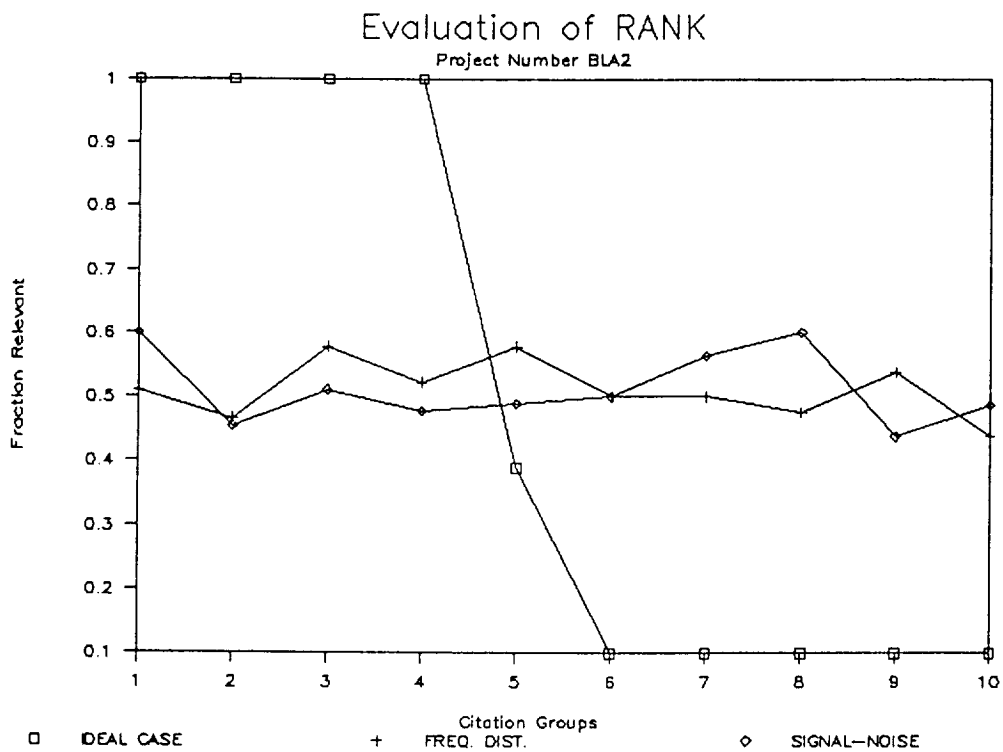


Figure 14. RANK Distribution. $\eta_a = 0.55$.

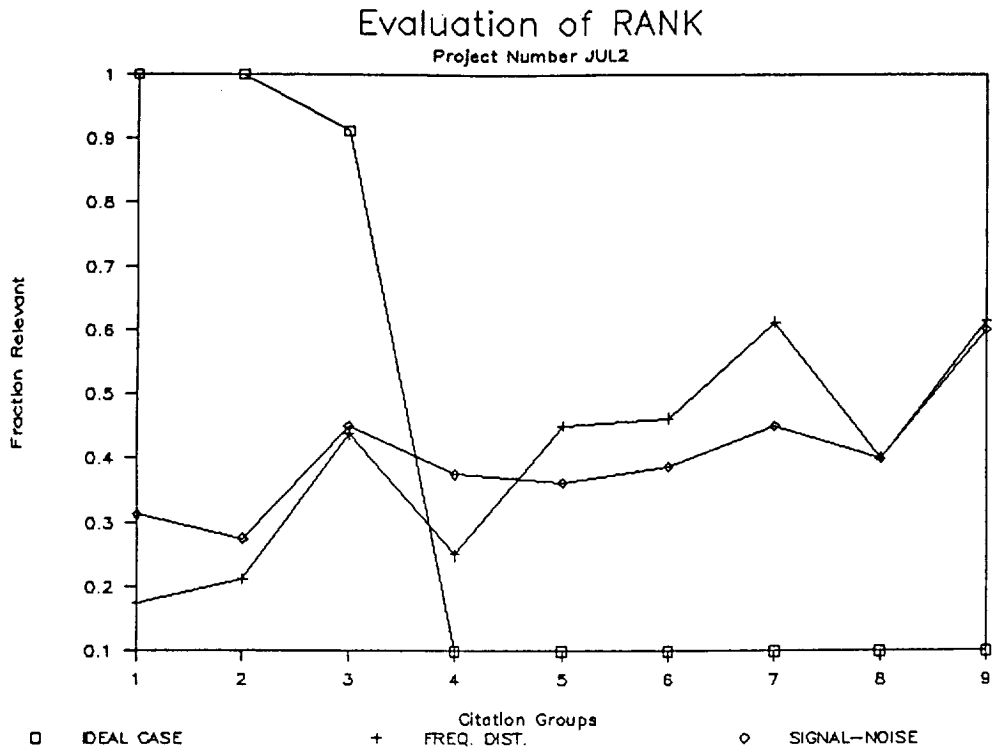


Figure 15. RANK Distribution. $\eta_a = 0.47$

4. PROPOSED WORK

4.1 Minor Modifications to SORT-AID

As we discussed in Section 3.4, preliminary benchmark results with SORT-AID (Version 4.0) are encouraging. These modifications will be fully-verified and the new version of the program evaluated for consistency of results with earlier versions.

The current version of RANK produces distributions with a greater concentration of relevant citations at the beginning of the collection. The frequency method is marginally superior to the signal-to-noise technique. Both methods now function as independent entities with user selecting one approach or the other. The original, minicomputer version of the program combined the two methods, but this approach was dropped in microcomputer versions because of execution time problems. The reduction in execution times realized in Version 3.0 and 4.0 makes possible the joint use of both methods. This will be explored in the coming year and if initial testing shows a substantial increase in ordering efficiency, a complete evaluation will be carried out.

4.2 SORT-AID Enhancements

As a result of recent discussions with the information service managers of the NASA Industrial Application Centers, see, e.g., Jack (1987), a series of SORT-AID enhancements have been suggested. The bulk of the enhancements are directed at

standardizing the myriad citation formats which result from multiple database searches. In particular, the enhancements are:

i. Translate citations via automatic editing from the host database format to a standard format. The translator module will be designed to be user extendable, i.e., the user will have the capability to specify citation delimiters and descriptors. The standard format will be specified by personnel at the NASA Scientific and Technical Information Facility (NASA-STIF). While the translator will be designed for the broadest applicability--NASA-STIF personnel will supply examples of typical and atypical citations--no one can guarantee that the algorithms will function for all possible citation formats.

ii. Merge the translated citations into a single file for additional post-processing. Once item (i) is accomplished, the existing versions of SORT-AID are capable of this function.

iii. Eliminate duplicate citations. With all citations in the same format, this task is relatively straight-forward.

iv. Create indexes and/or concordances for the merged citation sets by field, i.e., author, subject terms, title, abstract, etc. The RANK module of SORT-AID currently builds an index for all unique terms in the citation set. With all subfields specified, RANK can build indexes for subfields or the complete citation.

v. Provide general citation analysis and review tools. The REVIEW module of SORT-AID currently supports approximately 20 analysis and review functions. These can be readily applied to the standard citations.

vi. Production of print-ready bibliographies. The PRINT submodule of SORT-AID currently provides user-specified print formats. These capabilities can be enhanced. The use of a standard citation format will provide considerable latitude in output format.

The enhanced version of SORT-AID is shown in flowchart form in Figure 16. The major change in the existing program is the addition of the TRANSLATOR module. The bulk of the other enhancements will consist of fairly minor modifications to RANK and REVIEW. Note that both TRANSLATOR and REVIEW are user extendable.

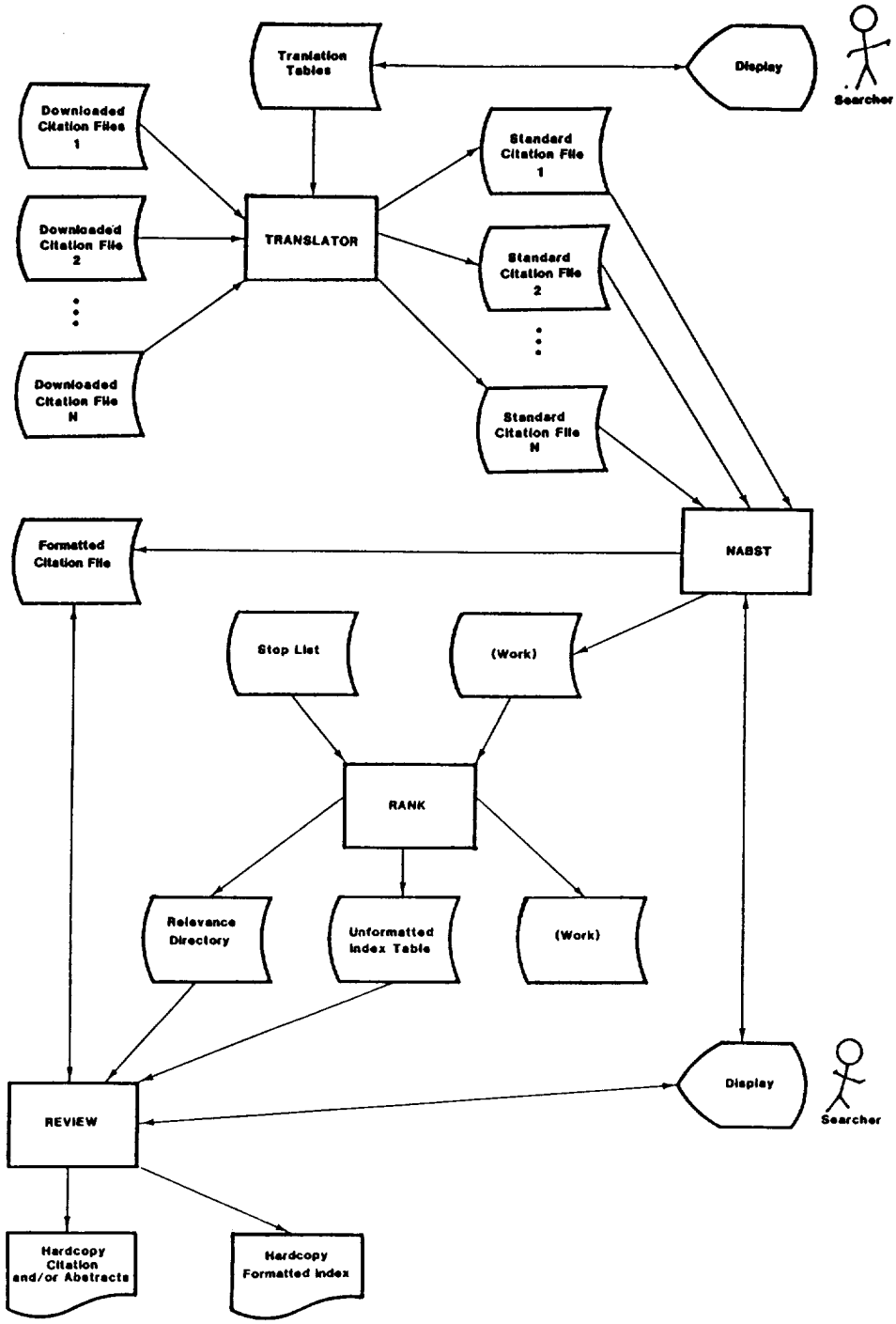


Figure 16. Flowchart of Enhanced SORT-AID.

4.3 Statement of Work

Task 1

The University shall design, develop, test and document an augmentation to SORT-AID, hereafter referred to as the TRANSLATOR module, that performs the following functions:

i. receives citation files obtained from selected bibliographic data bases and reformats them into a standard schema; and

ii. supports user selection of citation delimiters and descriptors.

Task 2

The University shall design, develop, test, and document enhancements to the currently operable SORT-AID package (Version 3.0) that improve the computational efficiency, disk utilization, and ordering efficiency.

Task 3

The University shall investigate enhancements to the RANK module of SORT-AID that perform the following functions:

i. produce an index table;

ii. allow generation of field specific, index tables according to user selection; and

iii. allow generation of field specific, relevance directories according to user selection.

Task 4

The University shall investigate enhancements to the REVIEW module that perform the following functions:

i. print a formatted index according to the field specific tables produced by the enhanced RANK module; and

ii. print a formatted listing of citations and/or abstracts.

Task 5

The University shall produce periodic reports of activities and accomplishments and a final report encompassing activities associated with Tasks 1 through 4.

Deliverables

The following deliverables will be provided as a result of this statement of work:

i. a final report on project activities;

ii. executable load modules and associated support files for the MS-DOS operating environment; and

iii. a user's manual for SORT-AID as enhanced in Tasks 1 through 4.

4.4 Program Budget and Schedule

The program budget is contained in Table 6 with a total program cost of $101,073--$90,963 to NASA and $10,110 to USM. The program is of 12 months duration with all deliverables to be provided at the completion of the program period.

TABLE 6
PROGRAM BUDGET

	NASA-FUNDED				USM-FUNDED			
PERSONNEL								
Name/Position	Level of Effort MM	Salaries & Wages	Fringe Benefits	Total	Level of Effort MM	Salaries & Wages	Fringe Benefits	Total
G David Huffman, Principal Investigator	2.00	11,000	1,870	12,870	1.00	5,500	935	6,435
Dennis Vital, System Analyst	10.50	21,000	4,410	25,410				
TBN, System Analyst	7.50	15,000	3,150	18,150				
Deborah Theisen, Technical Typist	2.50	3,499	910	4,409				
Subtotal	22.50	50,499	10,340	60,839	1.00	5,500	935	6,435
INDIRECT COSTS(45% of Salaries & Wages)								
Subtotal				22,725				2,475
TRAVEL								
Item/Description								
8 Person days @ $125/day				1,000				
3 Airfares @ $500/fare				1,500				
Subtotal				2,500				
SUPPLIES, COMMODITIES, & COMMUNICATIONS								
Office supplies, media, etc				900				
Reproduction & graphics				1,500				
Online database searches				2,500				
Local & toll telephone charges								1,200
Subtotal				4,900				1,200
TOTAL				90,963				10,110
GRAND TOTAL				101,073				

5. REFERENCES AND BIBLIOGRAPHY

ANSI/X3/SPARC Database System Study Group, "Reference Models for DBMs Standardization," private communication, October, 1984.

Bertrand, D., and Bader, C. R., "Storage and Retrieval of Bibliographic References Using a Microprocessor System," Int. J. Bio-Medical Computing, Vol. 11, 1980.

Bieber, H., "Technology Transfer in Practice," IEEE Trans on Engr. Management, Vol. EM-26, 1969, pgs. 144-147.

Bramwell, Bob, "BROWSE: An On-line Manual and System Without an Acronym," Asterisk, December, 1983.

Burlem, W. S., "Motivations for Technology Transfer," J. Tech Trans, Vol. 1, 1977, pgs. 53-61.

Bush, V., "As We May Think," Atlantic Monthly, 176, July, 1945.

Carroll, John M., and Carrithers, Caroline, "Training Wheels in a User Interface," Comm. of the ACM, August, 1984.

Chapman, R., "NASA Partnership with Industry: Enhancing Technology Transfer," Denver Research Institute Report, 1983.

Chen, Violet, M.S. Project Report, "An Algorithm for Recognizing and Reformatting Substrings," University of Southern Mississippi, Dept. of Computer Science, 1985.

CODASYL, "A Status Report on the Activities of the CODASYL End User Facilities Committee," SIGMOD Record, Vol. 10, Nos. 2 and 3, 1978.

Cooper, W. S., "On Selecting a Measure of Retrieval Effectiveness-Part I", J. Am. Soc. Inf. Sci., March-April, 1973, pgs. 87-100.

Cooper, W. S., "On Selecting a Measure of Retrieval Effectiveness-Part II", J. Am. Soc. Inf. Sci., November-December, 1973, pgs. 413-424.

Cooper, W. S., "Expected Search Length: A Single Measure of Retrieval Effectiveness Based on the Weak Ordering Action of Retrieval Systems", American Documentation, January, 1968, pgs. 30-41.

Digman, L. A., "Organizational Factors Related to Technology Transfer and Innovation," J. Tech Trans, Vol. 2, 1977, pgs. 61-68.

Elsasser, T. and Schneider, R., "Two Conflicting Approaches to Adaptation of Technology Transfer, Part I," J. Tech Trans, Vol. 2, 1977, pgs. 69-75.

George, P., Jolly, J. A. and Creighton, J. W., "The Linker's Contribution to Technology Transfer," J. Tech. Tran., Vol. 3, 1978, pgs. 51-61.

Goehring, M. F., "Automating the Search Process-Phase I SORT-AID Implementation and Characterization," Aerospace Research Applications Center, Indianapolis, IN, February 1986.

Grosch, A. N. "Configuring a Professional Microsystem for Information Processing," Microcomputers for Information Management, Vol. 1, 1984, pgs. 15-29.

Hampel, V. E., et. al., "Down Loading and Post-Processing of Bibliographic Information with the TIS Intelligent Gateway Computer," On-Line '82 Conference, November 1-3, 1982.

Heimbigner, Dennis, "Towards an Integrated Environment for Accessing External Databases," Proceedings of SECIBD ACM-SIGOA Conference on Office Information Systems, June 25-27, 1984, Toronto, Canada, Vol. 5, No. 1-2.

Herzog, A. J., "The 'Gatekeeper' Hypothesis and the International Transfer of Scientific Knowledge," J. Tech. Trans, Vol. 6, 1981, pgs. 57-72.

Howard, Steven, M. S. Project Report, "A Tool for Experimenting with Using Hypertext for Determining the Relevance of Citations," University of Southern Mississippi, Dept. of Computer Science, 1985.

Huffman, G. David, et. al., "Aerospace Research Applications Center. Volume I. Technical Proposal," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, June 1983.

Huffman, G. David, et. al., "Development of a Technology Transfer Workstation," Hawaii International Conference on System Science, January, 1986.

Huffman, G. David, "Continuation of the Development of Software to be Used in Support of the NASA Technology Utilization Program--Interim Report", University of Southern Mississippi, July, 1986.

Huffman, G. David, "Technology Transfer--First Quarterly Report," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, December 1983.

Huffman, G. David and Ulrich, J. M., "Technology Transfer--Second Quarterly Report," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, January 1984.

Huffman, G. David, "Technology Transfer--Third Quarterly Report," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, April 1984.

Huffman, G. David, et. al., "Aerospace Research Applications Center. Volume I. Technical Proposal," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, April 1984.

Huffman, G. David, et. al., "The Industrial Applications Study: A Mechanism for Technology Transfer," The Technology Transfer Society, June 25-26, 1984.

Huffman, G. David, "A Proposal for the Support of the Indianapolis Center for Advanced Research's Technology Transfer Program," University of Southern Mississippi, June 1984.

Huffman, G. David, et. al., "Technology Transfer--Annual Report," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, July 1984.

Huffman, G. David, "Technology Transfer Mechanisms," submitted to J. Tech. Trans., January, 1985.

Huffman, G. David, "Software Support Systems for Use in Technology Transfer," J. Tech. Trans., Vol. 9, 1985, pgs. 29-41.

Huffman, G. David, "Economic Development through Technology Transfer--Mississippi's Approach," Annual Technology Transfer Symposium, San Francisco, CA, June, 1985.

Huffman, G. David, Problem Solving Through Technology Transfer, to be published by NASA National Space Technology Laboratories, 1987.

Huffman, G. David, "Semi-Automatic Determination of Citation Relevancy: A Preliminary Report", accepted for publication in Information Processing and Management, 1987.

Jack, R., private communication, NASA Scientific and Technical Information Facility, BWI Airport, MD, 1987.

Janis, F. T., "Economic Development Through Technology Transfer, People Interaction--The Key to Technology Transfer," The Technology Transfer Society, June 20-22, 1983.

Jervis, P., "Innovation and Technology Transfer--The Roles and Characteristics of Individuals," IEEE Trans. on Engr. Management, Vol. EM-22, 1975, pgs. 19-27.

Johnson, F. D., "Technology Transfer--A View of What Works," J. Tech. Trans., Vol. 7, 1983, pgs. 1-4.

Jolly, J. A. and Creighton, J. W., "The Technology Transfer Process: Concepts, Framework and Methodology," J. Tech. Trans., Vol. 1, 1977, pgs. 77-91.

Jung, W., "Barriers to Technology Transfer and Their Elimination," J. Tech. Trans., Vol. 4, 1980, pgs. 15-25.

Jung. W., "Basic Concepts for the Evaluation of Technology Transfer Benefits," J. Tech. Trans., Vol. 5, 1980, pgs. 37-49.

Lefkovitz, David, "PRIMATE: A Microcomputer System for Host Mediation and Private File Text Search," Proceedings of the ASIS Annual Meeting, Volume 19, 1982.

Marcus, Richard S., Kugel, Peter and Benenfeld, Alan R., "Catalog Information and Text as Indicators and Relevance," JASIS, January, 1978.

Marcus, Richard S., "User Assistance in Bibliographic Retrieval Networks through a Computer Intermediary," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-12, No. 2, March/April, 1982.

Meadow, Charles T., "The Computer as a Search Intermediary," On-Line, July, 1979.

Menlo Corporation, "IN-SEARCH" An On-line Database Support Package," 1984.

O'Keefe, T. G., "Evaluations of Technology Transfer: Some Problems and Solutions," J. Tech. Trans., Vol. 6, 1982, pgs. 53-57.

Price, Lynn A., "Thumb: An Interactive Tool for Accessing and Maintaining Text," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-12, March/April, 1982.

Robertson, D., McCracken, D., and Newell, A., "The ZOG Approach to Man-Machine Communication," International Journal of Man-Machine Studies, Vol. 14, 1981, pgs. 461-488.

Robertson, S. E., "The Probabilistic Character of Relevance", Information Processing and Management, Vol. 13, 1977, pgs. 247-251.

Rothwell, R., "Some Problems of Technology Transfer into Industry: Examples from the Textile Machinery Sector," IEEE Trans. on Engr. Management, Vol. EM-25, 1978, pgs. 15-20.

Rouse, Sandra H., and Rouse, William B., "Computer-Based Manuals for Procedural Information," IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-10, August, 1980.

Sagal, M. W., "Effective Technology Transfer--From Laboratory to Production Line," J. Tech. Trans., Vol. 1, 1977, pgs. 7-15.

Salton, G., "A Theory of Indexing," Society for Industrial and Applied Mathematics, Philadelphia, PA, 1975.

Salton, G., Automatic Information Organization and Retrieval, McGraw-Hill Book Company, 1968.

Salton, G., Dynamic Information and Library Processing, Prentice-Hall, Inc. 1975.

Salton, G. and McGill, M. J., "Introduction to Modern Information Retrieval", McGraw-Hill Book Company, 1983.

Saracevic, T., "Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science", J. Am. Soc. Inf. Sci., November-December, 1975, pgs. 321-343.

Shannon, C. E., "A Mathematical Theory of Communication," Bell Systems Tech. J., Vol. 27, pgs. 379-423 and pgs. 623-656, 1948.

Sheahen, T. P. and Ryan, J. F., "Technology Transfer Into Capital-Intensive Industry: The Paper Industry by the Year 2000," J. Tech. Trans., Vol. 7, 1983, pgs. 69-84.

Steinbach, B. Z. and Steinbach, K. H., "Mission and Responsibility of the Technology Transfer Agent," J. Tech. Trans., Vol. 3, 1978, pgs. 67-71.

Strawn, G. W., "Technology Transfer: What is It? What Can It Do for Me?", J. Tech. Trans., Vol. 6, 1982, pgs. 45-51.

Tinsley, E. C., "Computer-Aided Sorting in Industrial Applications Studies," Indianapolis Center for Advanced Research, Inc., Indianapolis, IN, October 1983.

Treu, Siegfried, "Uniformity in User-Computer Interaction Languages: A Compromise Solution," Int. J. Man-Machine Studies, Vol. 16, 1982.

University of Southern Mississippi, Graduate Bulletin, 1984.

van Rijsdergen, C. J. Information Retrieval, Butterworths, 1980.

Vital, Dennis, M. S. Thesis, "A Portable Database Searcher's Software Toolset," University of Southern Mississippi, Dept. of Computer Science, 1985.

Weyer, Stephen A., "The Design of a Dynamic Book for Information Search," Int. J. Man-Machine Studies, Vol. 17, 1982.

Wilde, Dan, private communication, October, 1984.

Witten, Ian H., and Bramwell, Bob, "A System for Interactive Viewing of Structured Documents," Comm. of the ACM, March, 1985.