DOCUMENT RESUME

ED 247 948 IR 050 830

AUTHOR Niehoff, Robert; Mack, Greg
TITLE Evaluation of the Vocabulary Switching System. Final Report.
INSTITUTION Battelle Memorial Inst., Columbus, Ohio. Columbus Labs.
SPONS AGENCY National Science Foundation. Washington, D.C. Div. of Information Science and Technology.
PUB DATE Aug 84
GRANT NSF-IST-7911190; NSF-IST-8111497
NOTE 209p.;
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS *Classification; Comparative Analysis; *Databases; *Search Strategies; *Subject Index Terms; Thesauri; Use Studies
IDENTIFIERS End Users; National Science Foundation; Natural Language; *Online Search Skills; Search Intermediaries; Switching Theory; *Vocabulary Switching Systems

ABSTRACT

This report describes recent efforts to modify, test, and evaluate an experimental online database consisting of 15 vendor-supplied, controlled subject vocabularies or thesauri. The experimental database, called the Vocabulary Switching System (VSS), is designed to enhance search strategies and ultimately retrieval performance for users of online bibliographic databases by integrating vocabularies into common VSS files, thereby allowing the user access to about 315,000 possible search terms. VSS assists users with free-text or controlled vocabulary searches and single or multiple database searches. Work described includes: (1) evaluation of numerous switching strategies against the 15 resident vocabularies of VSS; (2) controlled experiments involving end-users and intermediaries; (3) system redesign and current status; (4) system testing and evaluation; (5) results of a 1983 survey of online users and comparison with a similar 1979 survey; and (6) conceptual designs for a future switching system. Numerous tables, figures, and appendices are provided. (THC)

# FINAL REPORT

on

# EVALUATION OF THE VOCABULARY SWITCHING SYSTEM

NSF GRANTS IST-7911190
and IST-8111497

for

NATIONAL SCIENCE FOUNDATION
DIVISION OF INFORMATION SCIENCE
AND TECHNOLOGY

August, 1984

by

Robert Niehoff
and
Greg Mack

This material is based upon research supported by the National Science Foundation under Grant Numbers IST-7911190 and IST-8111497. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors, and do not necessarily reflect the views of the National Science Foundation.



BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

## PREFACE

This report describes work performed under NSF Grant Nos. IST-7911190 and IST-8111497, and covers the periods October 1, 1979 to September 30, 1981 and January 15, 1982 to October 31, 1983, respectively. No attempt is made to report, chronologically, the activity under these two efforts. However, the work performed can be generally described as follows:

| | |
|---|---|
| 1979-1981 | Major software upgrade; expansion of experimental system from 6 to 15 controlled vocabularies; evaluation of various experimental switching stacks; conceptualization of expression - level switching. |
| 1982-1983 | Major feature implementation (related terms); new user interface (menu access); field evaluation of VSS; update of 1979 on-line survey. |

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the National Science Foundation, Division of Information Science and Technology. Additional support was provided by Battelle's Software Product Center and the Computer Center.

Vocabularies were provided gratis or at nominal cost to the project by:

- Data Courier, Inc.
- Management Contents
- Chemical Abstracts Service
- National Aeronautics and Space Administration
- The Institution of Electrical Engineers
- Engineering Information, Inc.
- Battelle's Iron Information Center
- Psychological Abstracts
- Educational Resources Information Center
- National Library of Medicine
- Biological Information Services
- Department of Energy

All of these organizations were helpful and supportive of the project.

Special recognition is due the following individuals for their important contributions: Dr. Barry Brinkman, Brenda Mackay, Helen Pestel, Fred Todt, Leslie Arnold, Lou Meyer, and Robert Kohn. Also, we are especially grateful to the many participants who took part in our experiments. Finally, a very special thank you is given to Debbie Schuster who prepared this manuscript and labored through countless drafts and revisions.

## EXECUTIVE SUMMARY

The Vocabulary Switching System (VSS) is an experimental system designed to enhance search strategies and ultimately retrieval performance for those who use online bibliographic data bases. VSS contains 15 indexing and retrieval vocabularies from 12 different suppliers. By fully integrating these vocabularies into common VSS files the VSS user has access to about 315 thousand possible search terms. VSS assists users with free-text or controlled-vocabulary searches and single or multiple data base searches.

This report describes research efforts undertaken to: (1) modify and expand VSS from an earlier more primitative version, and (2) evaluate VSS in end-user and intermediary communities in real-life situations. In addition, a 1979 survey of online users was updated to determine if any shifts or trends have occurred in user patterns or preferences over time.

During this research the system was expanded from 6 to 15 vocabularies with each vocabulary being assigned to one or more of four modules. The VSS modules and vocabularies are:

**Physical Science Module:** Chemical Abstracts Concept Edit File; Department of Energy Thesaurus; Subject Headings for Engineering; Inspec Thesaurus; Iron Center Thesaurus (trilingual), and NASA Thesaurus.

**Life Science Module:** BIOSIS Master Index Authority File (2 files created); Medical Subject Headings (3 files created); Chemical Abstracts Concept Edit File.

**Social Science Module:** ERIC Thesaurus, Psycological Abstracts Thesaurus.

**Business Module:** ABI Inform Thesaurus; Management Contents Thesaurus.

Also during this period, VSS was modified to handle related terms and both sanitized and unsanitized versions of all lead terms. Finally, a new, menu-driven interface was designed for use in the evaluation phase.

End user evaluations of VSS were conducted at three separate remote sites using two different evaluation methodologies. VSS performed quite well at one site using an iterative methodology (one search intermediary for each

user query) but it was only-marginally better than the non-VSS searches at two other sites using a parallel methodology (two search intermediaries for each user query). The parallel methodology seemed to be the least reliable of the two methodologies because search intermediaries tended to interpret the end user query differently.

Some 65 professional search intermediaries also evaluated VSS independently of end users. The participants thought VSS was easy to learn, use and understand and, quite naturally, rated the system's performance in direct proportion to the amount of usable output it generated for their query. Ratings of 3.5 to 4.1 on a 5-point scale were achieved when the amount of usable output reached 6 or more terms.

About 4 1/2 times as many participants thought VSS would make their jobs easier as thought VSS would make their jobs harder. About 3/4 of all participants thought highly of the concept of subject switching as a search tool, and about 1/2 of all participants thought VSS was very valuable, valuable, or "interesting but needs more work". It was concluded that subject switching has a potentially wide appeal to information professionals but VSS itself needs more work to address their needs and concerns.

The online user's survey confirmed reported growth figures for the online industry, that is, growth at a 30 to 40 percent rate compounded annually. Also, end users are paying for a greater percentage of their own searches in 1983 compared to 1979. The usage pattern of seven major retrieval services showed little change over the four year period, 1979 to 1983 and the survey showed that multiple data base searching is increasing with time in the direction of 4 or more data bases per search request. Results suggested that a user's need for VSS increased as their need for multiple data-base searches increased.

It seems certain that user acceptance of VSS will be quite high if the system contains (1) more vocabularies (2) more of the syndetic relationships available in the host vocabularies, (3) a means of attaining and maintaining the most current versions of host vocabularies, and (4) a more direct approach to searching VSS. These improvements are all well within the state-of-the-art of data base and thesaurus technology.

## TABLE OF CONTENTS

## TABLE OF CONTENTS
(Continued)



## LIST OF TABLES

## TABLE OF CONTENTS
(Continued)

Page

## LIST OF TABLES
(Continued)



## LIST OF FIGURES



## LIST OF APPENDICES

## 1.0 INTRODUCTION

This report describes recent efforts to modify, test and evaluate an experimental online data base consisting of 15 vendor-supplied, controlled subject vocabularies or thesauri. The experimental data base is called the Vocabulary Switching System (VSS).

The work described herein includes:

- Evaluation of numerous switching strategies against the 15 resident vocabularies in VSS
- Controlled experiments involving end-users and intermediaries
- System redesign and current status
- System testing and evaluation
- Results of a 1983 survey of online users and comparison with 1979 data
- Conceptual designs for a future switching system.

## 2.0 BACKGROUND

Man's stored base of computer searchable bibliographic records has reached enormous size. According to Cuadra (1) there are over 2,000 online data bases worldwide, about half of which are bibliographic or textual. The bibliographic data bases alone contain hundreds of millions of citations. DIALOG, one of the largest online services for bibliographic data bases, contains more than 170 data bases and over 75 million records as of January, 1983. Their largest data base contains over 7 million citations; the smallest about 4 thousand (2).

Other large online services include BRS (over 70 data bases); Pergamon Infoline (over 20 data bases and over 22 million records); SDC's ORBIT (over 70 data bases and over 55 million citations); and ESA/IRS (over 50 data bases and over 25 million citations). All totalled, over 250 online services are available worldwide.

With this scale as a reference, several observations can be made:

(1) Relevant information probably exists in some online data base somewhere on virtually any topic known to man

(2) Locating the relevant information on any particular topic may vary from extremely easy to virtually impossible and could be very costly depending on a number of factors: the user, the topic, the intermediary (if used), the online service, the data base, the index, the nature of the data base records, the language of the source documents, how the retrieval records are displayed (online or offline), data base and telecommunication charges, time of day spent searching, and modem baud rate, just to mention a few.

(3) The replacement of existing information retrieval technology (i.e., inverted files, free-text and controlled-descriptor indexing, Boolean query formulation) with newer concepts and technology is likely to be evolutionary for many years ahead simply because of the scale and investment in these existing systems.

Bradford's Law of Scatter* is easily understood and accepted by virtually everybody, in many cases intuitively. If a comprehensive online search can be defined as locating a significant percentage of the relevant information on a particular topic, then, depending on how one defines significant, such a search usually translates into a search of more than one data base. In addition, if the search is a subject search, then a multi-field search is usually specified, including: (1) titles, (b) abstracts, (c) descriptors, (d) identifiers, and (e) subject classifications, if they exist. These fields are usually searched with a combination of free-text and controlled descriptor terms although some searches use one or the other approach exclusively.

When more than one data base is involved in a comprehensive online subject search, differing indexing policies and approaches must be recognized if the search is to be optimized for each data base. This is especially true if the search strategy involves controlled descriptors. Several aids and approaches are available, either commercially or experimentally, to improve subject retrieval from online bibliographic data bases.

The most common search aid is the master index data base, a file containing combined inverted indexes of several data bases. The primary function of a master index data base is to guide the user to appropriate specific data base(s) based on posting frequencies for a stated topic. Several major retrieval services have master indexes:

| Retrieval Service | Master Index |
|---|---|
| Dialog | Dialindex |
| BRS | Cross |
| SDC | Data Base Index |
| ESA/IRS | Questindex |

---

*Relevant information will be scattered over many sources, although a high percentage will be concentrated in just a few sources.

Thus, while the master index directs a user to certain data bases, such an index does little to help the user state the query in the most appropriate system or indexing language for optimal retrieval under that topic.

A second aid is the online thesaurus. Some retrieval services permit a browse of the technical thesaurus but only for the data base being searched. If multiple data bases are to be searched, each thesaurus must be searched separately, if one exists. However, BRS now offers an integrated, controlled vocabulary data base containing subject headings from five social science vocabularies. This service, called TERMS by BRS, is similar to VSS.

A third approach for improved retrieval is to rank order the retrieved citations according to some internally stored algorithm (e.g., combinatorial term weighting or probability estimation). The user is given a list of citations in ranked order where the highest weighted or most probably useful citations are presented first. The work of Doszkocs and Cooper are examples of term weighting and probabilistic approaches to information retrieval, respectfully.

Doszkocs (3) has built a front end to MEDLINE, TOXLINE and other NLM data bases. This front end, called Current Information Transfer in English (CITE), permits queries in natural language, ranked output based on combinatorial term weights, and relevance feedback including automatic query modification. Terms in the user's query are weighted automatically by the system based on inverse collection frequencies of such terms. Output is ranked according to the sum of the weights of the query terms contained in a given document.

On the other hand, Cooper (4) has pursued probability theory in information retrieval and presently bases his probability computation on the entropy of distribution. This approach is called the maximum entropy principle and claims that it overcomes the difficulties of other probability approaches because it avoids certain simplifying assumptions concerning statistical independence. However, he admits that Boolean retrieval is so firmly entrenched in the retrieval world that even "front-ending" a maximum entropy system onto an existing Boolean system would take "some time" to undo the entrenchment. Thus, non-Boolean systems, although theoretically sound, are largely experimental and their acceptance seems to be some time off.

In yet another approach, already retrieved document sets can be further analyzed for word frequency patterns on any field specified by the user. The result is a list of words (occurring in the field specified) in decreasing order according to their frequency. Such lists provide additional clues to the user for fine tuning a search. The European Retrieval Service, ESR/IRS, offers such a search feature, called ZOOM.

The ZOOM feature is useful for identifying terms which co-occur with those used in the original search strategy. It frees a user from the task of actually reading abstracts, titles, descriptors and other fields for additional clues with which to fine tune a search. However, ZOOM cannot identify co-occuring words as synonyms or generic terms and the results are dependent on the quality of certain critical fields (abstracts, titles, descriptors) and the number of citations being analyzed.

Other transparent interfaces and linked systems provide partial solutions to the problem of accessing dissimilar files for similar information. The CONIT system by Marcus (5) provides only truncated keyword and full-phrase searching. The NIH-EPA linked chemical data base system (6) is keyed exclusively to substances and chemical registry numbers (CRN). That is, once the CRN is known, several different files can be accessed since they all use the CRN as access key. The Chemical Substances Information Network, like the NIH-EPA system is keyed to chemical substance linkages (7). Again, this system operates in a limited domain.

In the final analysis, linkage to dissimilar and heterogeneous data bases is an important endeavor if we are to fully utilize our national information resource, and take full advantage of the potential that the computer offers. Bates (8) describes no less than 29 useful search tactics for improving the query process. She pleads for searching decision rules that will minimize cognitive strain and begs for anything that will help searchers work faster [and better]. The existence of stress and strain in this process is confirmed by Standara (9), who found the presence of "peak levels of pressure" in three critical sub-stages of the search process: (1) the selection of systems, data bases and search strategies, (2) evaluating interim results (while connect time is still running), and (3) presenting results and costs to the user. This finding reportedly was based on the opinion of searchers who collectively had performed 10,000 searches.

## 3.0 PURPOSE

A major purpose of this research was to investigate a particular approach for improving search strategy preparation, and ultimately retrieval performance, for those on-line searches involving more than one data base. The approach was to integrate existing controlled vocabularies into one data base and to exploit the enormous intellectual effort represented by such a composite file for the purpose of improving search strategies.

## 4.0 OBJECTIVES AND SCOPE

The combined objectives of the two research grants were to:

(1) Improve and expand earlier subject switching concepts and designs

(2) Test and evaluate VSS in controlled and uncontrolled environments

(3) Determine the influence of certain technical factors on VSS performance, including: vocabulary size, number of vocabularies, vocabulary similarity/dissimilarity, and type of switching strategies employed

(4) Determine the influence of certain human factors on VSS performance, including: education and experience of participants, their search proficiencies in various subject areas, and end user and intermediary evaluations of VSS

(5) Determine the extent of user satisfaction for searches performed with and without VSS enhancement

(6) Characterize a futuristic intermediary-free search environment involving VSS.

The scope of work was limited to a system consisting of 10 to 15 controlled vocabulariesavailable in machine-readable form and at a reasonable cost to the project. Formal testing was limited to 7 field sites.

A few scope modifications were made to the original objectives and scope during the actual course of the research. First, a futuristic, intermediary-free search scenario was not performed because the hours budgeted for this effort were actually needed in bringing up a revised and modified VSS. However, one additional activity was undertaken that was not originally planned. An update of the previous user survey of online searching behavior and patterns was conducted. This update was considered important from the standpoint of determining the shift, if any, in users on-line searching patterns and preferences with time.

The scope of field testing was modified for one primary reason: it was decided to acquire data from a larger segment of online users than was possible under the limitation of 7 field sites. To accommodate this scope

modification, 5 out of the 7 field sites where formal testing was to have occurred were dropped in favor of less formal evaluation involving a much broader audience. This modification resulted in an evaluation of VSS by 65 intermediaries instead of an estimated 14. The larger audience also provided more insight into the problem than would have been possible with a smaller audience.

## 5.0 OVERVIEW OF VSS (PRESENT VERSION)

VSS is a stand-alone, on-line, experimental data base containing the subject descriptors or main entries and all the syndetic relationships found in controlled vocabularies. Therefore, actual postings or frequency totals are not included in this data base because its origin is not the bibliographic record or conventional inverted file.

Currently, VSS resides on Battelle's CDC computer and uses a combination of Fortran and BASIS-IR software (Version 4.0) for data management and retrieval. Access to the data base, is via the TYMNET telecommunications network or direct dial service.

In this experimental mode, the on-line user wishing to enhance a search strategy via VSS must first log onto the Battelle's computer, then interrogate the VSS data base and examine the output, and finally log off. Any resultant search strategy, enhanced or not, must be re-entered in the usual fashion by dialing up a retrieval service and data base combination appropriate for the search. In a production mode, the VSS data base would simply reside as another data base within each retrieval services' host computer and would run with their software. This would eliminate the extra login step now required in the experimental model.

Like most other on-line data bases, VSS contains inverted files, data base records (for display or printing), and system commands. Unlike most other data bases, VSS contains only search terms or keywords useful for performing on-line searches in other data bases.

Since the inverted files and data base records are created from existing, machine-readable controlled vocabularies, these vocabularies are described first, followed by the inverted files, the concept file (data base records), system commands, and the switching options.

### 5.1 VSS Vocabularies

VSS contains all of the lead-term entries and syndetic relationships found in 15 controlled vocabularies. These vocabularies (see Table 1) are organized into four major categories, and they are called modules, accessible only by module. No inter-module access is permitted.

The physical sciences vocabularies, with the exception of the IRON Thesaurus, were acquired under an earlier NSF grant and represent late 1978 acquisitions. The IRON Thesaurus and all other vocabularies shown in Table 1 were acquired in late 1979 and early 1980 under a separate NSF grant. For the most part, these vocabularies were donated to the project or acquired for a nominal fee due to the research nature of the work.

All vocabularies were acquired in machine readable form via magnetic tape. However, in order to minimize the actual cost to each supplier, no format restrictions were imposed upon them. This, of course, placed a burden on the VSS staff to handle and process 12 different "as supplied" vocabulary formats.

As shown in Table 1, the smallest vocabulary included in VSS is Management Contents--the largest is MeSH (Medical Subject Headings) from the National Library of Medicine. This mix of vocabularies was assembled to study switching in various fields of endeavor rather than looking exclusively at any one subject area as was the case in previous research. The actual acquisitions represent the most readily available machine-readable vocabularies of the time.

The Chemical Abstracts Concept Edit File (CA) was placed in both the life and physical science categories because of its scope and impact with other vocabularies in these categories. Also, previous survey results indicated that popular multi-data base searches frequently involved CA in combination with those data bases in the life and physical sciences areas of VSS.

The INSPEC Thesaurus was an important acquisition because of the links (use/used for cross references) it had established between British and American spellings. Thus colour and color, centre and center were useful synonym linkages available only in the INSPEC Thesaurus. The IRON Thesaurus, developed by Battelle for the IRON Information Center was useful because it was trilingual, English, Spanish and Portuguese.

Although the Biosciences Information Service provided only one file, the MAIF, their concept code (C6) seemed to be an important synonym-type structure in which certain biological concepts were also given systematic numeric codes. Therefore, the VSS staff created a second BIOSIS vocabulary, BIOSIS-C, which was an inverse of the main-entry concept-code relationship

TABLE 1 VSS VOCABULARIES

| Modules/ Vocabularies | Thesaurus Relationships (figures given in thousands of terms) Lead Term | Used | Special Scope Notes | BT/* NT | RT | All Other | Total |
|---|---|---|---|---|---|---|---|
| **Business**** | | | | | | | |
| ABI Inform | 9.3 | 0.9 | 0 | 0 | 6.4 | 3.6 | 20.2 |
| Mgmt. Contents | 1.3 | 0 | 0 | 0 | 0 | 0 | 1.3 |
| Sub Total | 10.6 | 0.9 | 0 | 0 | 6.4 | 3.6 | 21.5 |
| **Social Science** | | | | | | | |
| ERIC | 8.1 | 2.9 | 0 | 10.0 | 36.4 | 3.9 | 61.3 |
| Psy. Abs. | 5.9 | 1.8 | 0 | 11.6 | 7.3 | 1.9 | 28.5 |
| Sub Total | 14.0 | 4.7 | 0 | 21.6 | 43.7 | 5.8 | 89.8 |
| **Life Science** | | | | | | | |
| BIOSIS | 16.8 | 6.8 | 4.1 | 0 | 9.2 | 0 | 36.9 |
| BIOSIS-C | 4.1 | 0 | 0 | 0 | 0 | 0 | 4.1 |
| CA | 13.9 | 2.9 | 0 | 0 | 0 | 2.9 | 19.7 |
| MeSH | 73.7 | 36.2 | 63.1 | 0 | 1.1 | 13.0 | 187.1 |
| MeSH-R | 42.2 | 28.3 | 0 | 0 | 0 | 9.4 | 79.9 |
| MeSH-S | 51.6 | 30.5 | 0 | 0 | 0 | 14.0 | 96.1 |
| Sub Total | 202.3 | 104.7 | 67.2 | 0 | 10.3 | 39.3 | 423.8 |
| **Physical Science** | | | | | | | |
| CA | 13.9 | 2.9 | 0 | 0 | 0 | 2.9 | 19.7 |
| DOE | 22.5 | 4.7 | 0 | 59.4 | 27.1 | 4.5 | 118.2 |
| EI | 12.2 | 2.9 | 0.1 | 0 | 0 | 1.9 | 17.1 |
| INSPEC | 9.1 | 4.0 | 0 | 9.8 | 15.9 | 9.3 | 48.1 |
| IRON | 16.8 | 4.1 | 0 | 20.5 | 4.7 | 6.8 | 52.9 |
| NASA | 15.9 | 3.6*** | 0 | 54.8 | 84.9 | 1.6 | 160.8 |
| Sub Total | 90.4 | 22.2 | 0.1 | 144.5 | 132.6 | 27.0 | 416.8 |

* BT - Broader term NT - narrower term RT - related term

** **Business**
- ABI Mgmt. Contents - ABI Inform Thesaurus
- Management Contents Thesaurus

**Social Science**
- ERIC - ERIC Thesaurus
- Psych Abs - Psychological Abstracts Thesaurus

**Life Science**
- BIOSIS - The Master Index Authority File (MIAF) from BIOSIS. In the VSS version of this vocabulary, BC (Biological Codes) and CC (Concept Codes) are invalid.
- BIOSIS-C - A special version of BIOSIS in which CC (concept codes) are valid but the concepts are invalid.
- MeSH - The Medical Subject Headings of the MEDLINE system used at the National Library of Medicine.
- MeSH-R - A special subset of MeSH in which substance names are invalid but their equivalent chemical registry numbers are valid.
- MeSH-S - The inverse of MeSH-R.
- CA - Chemical Abstracts Concept Edit File.

**Physical Science**
- CA - Chemical Abstracts Concept Edit File.
- DOE - Department of Energy Thesaurus.
- EI - Subject Headings for Engineering, from Engineering Information Inc. (formerly Engineering Index).
- INSPEC - INSPEC Thesaurus.
- IRON - A trilingual thesaurus on iron metallurgy. Languages include English, Spanish, and Portuguese.
- NASA - NASA Thesaurus.

*** NASA tape contained only UF relationships, but VSS preprocessing software expected use relationship.

found in the BIOSIS file. In other words, in VSS the main headings in BIOSIS were valid and the concept codes were invalid, while in the BIOSIS-C file the inverse was true. The decision to create a second file to mirror this important relationship made it possible to access VSS with either the concept or the concept code to perform switching.

In a similar fashion, the National Library of Medicine furnished one tape but a special segment on this tape contained MeSH headings mapped to Chemical Abstracts Registry Numbers. This tape was used to create three files, one for regular MeSH headings, one in which CA Registry Numbers were valid (substances invalid), and one in which substances were valid (Registry Numbers invalid).

It can also be seen from Table 1 that some vocabularies contain certain syndetic relationships which others do not. Thus, there were no broader/narrow term relationships in the life sciences area. Likewise, there were no special scope notes in the business and social science areas.

Finally, it should be pointed out, although it is not apparent from Table 1, that the vocabularies themselves represent a mixture of subject heading schemes and authority lists. Thus, they all represent some form of a controlled vocabulary but they are not all exclusively of one type. This is why the system is referred to as the Vocabulary Switching System and why we refer to them as vocabularies rather than thesauri.

A new feature added to VSS during this grant period (1982-83) was the related term. This relationship was ignored in earlier versions of the system because the browse aspect, which a related term provides, was considered much less important than the synonym, generic term (BT/NT), and scope note features. Since those early days, we have changed our opinion about the value, and indeed the purpose, of VSS. The related term is now considered to be an important dimension and potentially useful search term. In addition, VSS is now considered a tool to enhance search strategies not merely a tool to identify synonyms and generically related terms. The usefulness of VSS as a tool for enhancing searches is increased by providing all the syndetic rela-tionships available in the original vocabularies. As can be seen from Table 1, the inclusion of related terms added a considerable dimension to VSS, about 193 thousand new terms altogether.

Battelle processed each acquired vocabulary so as to preserve the syndetic relationships whenever and wherever possible. The syndetic relationships captured by VSS are: main-heading (lead term) entries, conventional scope notes, special scope notes, use (and see) cross references, use for cross references, broader and narrower terms, and related terms.

Special scope notes were created for VSS to handle certain types of synedetic relationships found in the acquired vocabularies which could not be easily handled any other way by our thesaurus software.

The two specific relationships handled by a special scope note in VSS were:

- Term A
  use either term B or term C or ...
- Term A
  use Term B, Term C, Term D ... all OR'ed

Each vocabulary included in VSS is assigned a unique letter code. This code is mapped to a bit table consisting of 24 positions (one position is designated for each assigned code). Therefore, the maximum number of vocabularies that can be handled in any one version of VSS is 24. The vocabulary code preserves the original source of each main entry or syndetic relationship stored and retrieved by VSS. The code is always displayed with all VSS output so that the user knows what controlled-vocabulary search terms are available for each data base.

Each acquired vocabulary was processed via a specially written program into a common VSS input format. (Appendix A shows the current format). It was from this common format that all VSS inverted files and data base records were eventually created.

## 5.2 Inverted Files

VSS inverted files are of four types:

- term file
- word file
- stem file
- stem phrase file

These files are created from the temporary file by executing a series of computer programs and a stemming algorithm. See Appendix B.

The term file is an inverted file containing every valid and invalid main entry term listed in each vocabulary. These terms can be single or multi-word terms or phrases. This file is created by a piece of software which "sanitizes" each term and assigns a pointer (concept number parcel) for use in retrieval of the data. The term file is keyed by term and uses its pointer as a key to retrieval of actual vocabulary records from the concept file, a sequential file keyed by pointer (concept number parcel). The pointer functions like an accession number -- it is the key to retrieval of vocabulary records from the concept file (in essence the data base records).

The sanitization algorithm squeezes all special characters from main entries, such as, hyphens, parentheses, semicolons, commas, etc. This same algorithm is invoked when users enter terms at the terminal. Thus, a whole class of morphologic problems dealing with punctuation are eliminated as a barrier to retrieval. Users need not remember if or how to punctuate a term to use VSS.

Only rarely was it necessary to subject a term to additional sanitization to improve retrieval. The most common case was the use of special codes by vocabulary suppliers (e.g. COSATI category codes). If retained as part of the main entry, special codes would require users to remember them to retrieve the term or to switch on it, or the user would have to truncate every term he/she was unsure of in order to succeed. Therefore, these types of strings, though uncommon, were removed from main entries before their insertion into the term file.

The word file is an inverted file of individual words created by disassembling every lead term in the temporary file into its component words. It is like the term in every other respect.

A major utility of this file is to provide the capability for retrieving portions of multi-word terms or phrases. In other words, if a VSS user starts with a 3-word phrase, he/she can request retrieval of all other phrases which contain any two words used in the original phrase, or any single word used in the original phrase. The system, using an internal scoring algorithm, retrieves all phrases which satisfy the user's requirements. Default is 100 percent.

The stem file is similar to the word file except that all individual words are processed by a stemming algorithm to create roots or stems of the words. This same stemming algorithm is invoked when users specify a stem file search at the terminal. It stems their input the same way it stems the inverted file entries. VSS produces output when matches between a user's stemmed input and a stem file entry are found.

A major use for the stem file is to locate all vocabulary entries with the same root. Thus, computers, computer, computing, computation, computational, all contain the same stem, compu, and all would be considered valid output if the user specified a stem file search on any one of these words.

The stem phrase file is an inverted file created by disassembling each main entry term in the temporary file into its component words, applying the stemming algorithm to each word, and concatenating the resulting stems. The result is an inverted file of stem strings where the string is composed of the stems of each unique word in the original term. Obviously, a stem phrase file entry is only created for terms or phrases containing two or more words. For example, electrical machining would be represented as ELECTRMACH in the stem phrase file.

The purpose of this file is to retrieve a variety of terms where the individual words within the term differ only by ending. Thus, electrical machines, electric machining, etc., all produce the same concatenated stem and would be considered valid output.

## 5.3 Concept File

The concept file is a symbolic keyed file, accessed by the concept number parcel. This file contains what can be considered the data base records, albeit the records are vocabulary records rather than references.

An individual record in this file contains the unsanitized text of the lead term, the vocabulary bit table indicating the source(s) of the term, and pointers to other terms in concept file. The other terms include: co-related terms, narrower/broader terms, related terms, and the USE terms. All VSS output originates from the concept file. The four inverted files merely

provide the means of gaining access to the concept file. Of course, this is all transparent to the VSS user.

The use of unsanitized terms was a new feature added to VSS during this grant period. It adds an important dimension to this system because the VSS user can enter a term incorrectly punctuated and retrieve a correctly punctuated term.

Table 2 shows the relative sizes of inverted and concept files for the four major search modules in VSS. As can be seen, the final file size for all four modules was about 89 million characters. The largest VSS file, is the term file with the exception of the social sciences module, where the concept was slightly larger. The smallest VSS file is the stem file. The largest module was Life Sciences, the smallest was the Business Module. On the average, about 97 characters of storage are required to implement one thesaurus entry with all its relationships according to the specifications of this version of VSS.

TABLE 2 VSS FILE SIZE BY VOCABULARY SET

(No. characters x 106)

| File Type | Business | Life Sciences | Social Sciences | Physical Sciences | Total |
|---|---|---|---|---|---|
| Concept | 0.76 | 9.76 | 1.39 | 7.54 | 19.45 |
| Term | 1.12 | 10.99 | 1.33 | 7.79 | 21.23 |
| Stem | 0.59 | 8.35 | 0.66 | 4.23 | 13.83 |
| Word | 0.73 | 9.14 | 0.80 | 4.88 | 15.55 |
| Phrase | 0.93 | 9.65 | 1.24 | 6.99 | 18.81 |
| Total | 4.13 | 47.89 | 5.42 | 31.43 | 88.87 |

## 5.4 System Commands

There have been no changes to the VSS system commands since the last report (10). Table 3 summarizes the available commands. During this reporting period, the system commands were used almost exclusively by the VSS staff to study the performance of VSS prior to the field evaluation. However, during the field evaluation VSS system commands were made transparent to the user. This was done to relieve the burden on the user of becoming a VSS systems expert before evaluating its usefulness.

## 5.5 Switching Options

Switching options are logical instructions which tell VSS what to do with the user query term and what files to access (in essence what type of switching to perform). They also define the types of records be retrieved and displayed from the VSS concept file. These options (Table 4) are used in conjunction with the system commands to effect a retrieval in VSS.

Option 21 was added to VSS during the reporting period (See Table 4). This option produces related terms, if any exist, to the term entered by the VSS user. It differs from option 9 which produces co-related terms. Related and co-related terms are different relationships in VSS. See Appendix C for a more detailed discussion of the available switching options.

Switching options, like system commands, were used by the VSS staff prior to the field evaluation to study various switching scenarios. During the field evaluation these options were made transparent to the user by designing a menu containing six switching strategies, five of which were pre-defined. When a user selected one of the five pre-defined strategies, VSS automatically invoked the switching options defined for that choice. The sixth strategy, user-defined, required the VSS user to supply his/her own options.

## 5.6 Using VSS

VSS can be used in one of two modes, command or menu-driven. The command mode requires user specification of what VSS must do, while the menu-

# TABLE 3 AVAILABLE SYSTEM COMMANDS

| | COMMAND | FUNCTION PERFORMED |
|---|---|---|
| General | QUIT | Terminates the VSS session |
| | RESET | Erases all previous commands except QUIT and STORE |
| | STORE | Catalogs a switching strategy |
| Modifies VSS Output | DUPLIC | Permits duplicate terms in the output |
| | EXPAND | Output from all preceding operations becomes input for next operation within a switching strategy |
| | FULLV | Displays validity of a term across all vocabularies regardless of VOCCNT parameter |
| | NOBOOL | Suppresses lead terms with imbedded Boolean connectors (e.g. .AND.) |
| | OUT=IN | Output from preceding operation becomes input for next operation within a switching strategy |
| | SCOPE | Permits display of scope notes |
| | SETADJ | Defines the number of adjacent terms to be displayed. |
| | SETOPS | Erases the previous switching strategy |
| | SETPCT | Determines how much of an input phrase must be present in an output phrase, if not 100% |
| | SETVOC | Erases the previously selected vocabularies |
| | VOCCNT | Defines the maximum number of terms desired from each vocabulary |
| Special Purpose | DEBUG | Activates a trace function for system debugged |
| | SETIST | Allows switching to commence at some user designated point within a switching strategy |
| | SETLST | Allows switching to terminate at some user designated end point other than the last option in the strategy |
| | SETPRM | Allows alteration of the array ISPRM |
| | SETRIP | Allows premature termination of a strategy if one or more terms are produced |
| | UNBUG | Deactivates the trace function |

driven mode requires only a few simple choices. The menu-driven mode was a newly implemented feature during this reporting period.

The menu-driven mode is shown in Figure 1. It involves only a few menus plus specification of the number of terms to be displayed. From then on, the user simply enters terms to be switched, one at a time. The VSS system responds to each term entered with a result or a message that switching was unsuccessful. Users can change their menu choices or specification statement at any time.

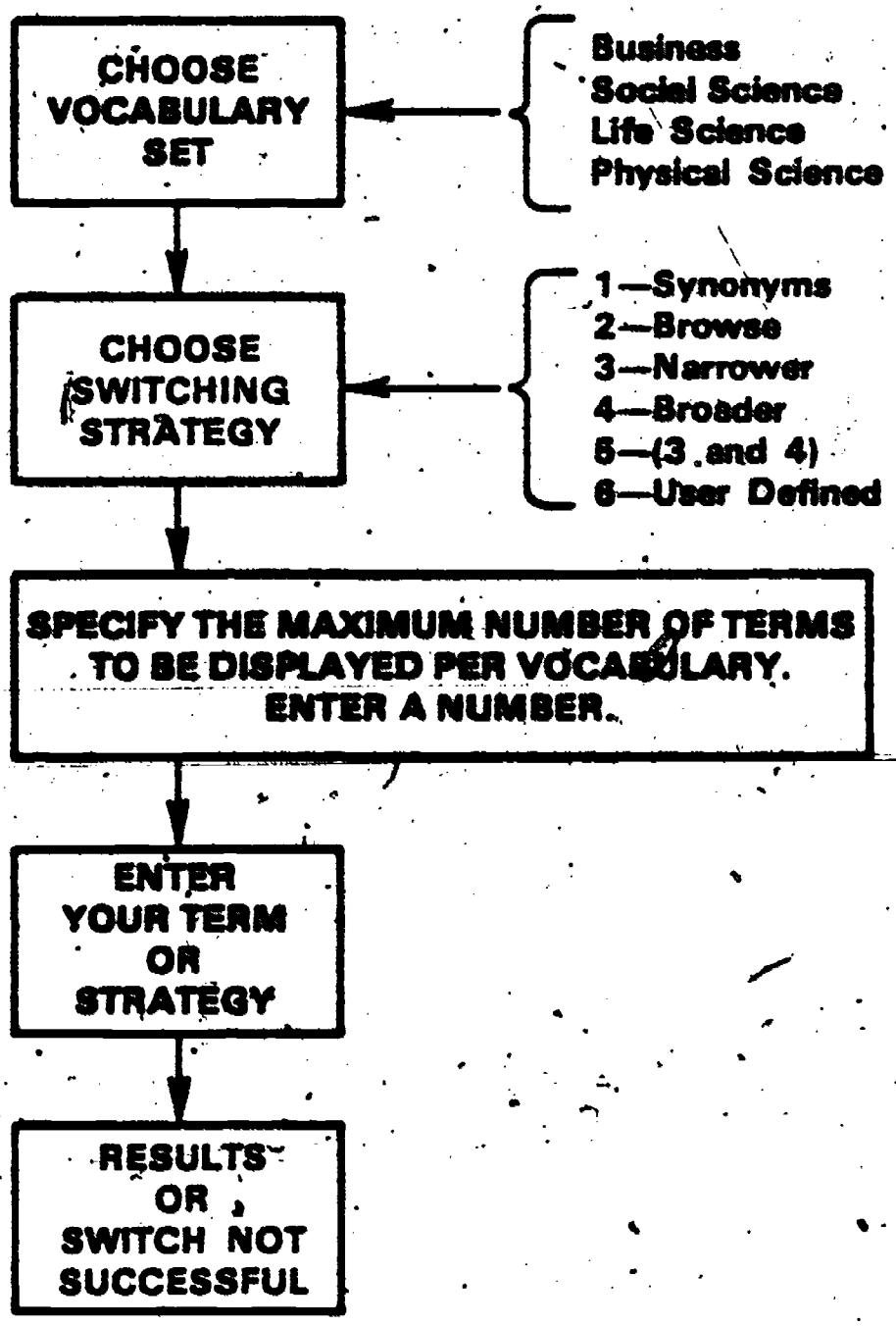
Figure 2 is an example of a typical terminal session using the menu-driven mode. In this example, the user selected all six vocabularies available in the life sciences module, the browse mode of switching and 6 terms per vocabulary. Then the user entered a Chemical Abstracts Service Registry Number, 9005-97-4 for switching. The system was able to execute the browse option and produced a table of results. The table shows the type of switching performed under the column TERM TYPE, the vocabulary containing the term found under the column heading VOCAB, and the term found under the column TERM. Appendix C contains a more detailed explanation of the results and also how to operate VSS in the command mode.

If switching is unsuccessful, the system simply issues a message to that effect. The user is free to enter another term or change the menu choices and re-enter the same term.

## TABLE 4 AVAILABLE SWITCHING OPTIONS

| Option Code | Switching Option | Option Nickname | Type of Output Produced |
|---|---|---|---|
| 3 | TERM FILE ACCESS - ACCEPT LEAD TERMS ONLY | Exact Matching | Exact Matching |
| 4 | TERM FILE ACCESS - ACCEPT SWITCHABLE TERMS ONLY | Partial Synonym Switching | Synonyms |
| 5 | TERM FILE ACCESS - ACCEPT LEAD + SWITCHABLE TERMS | -- | Exact Matches Plus Synonyms |
| 6 | STEM PHRASE FILE - ACCEPT LEAD + SWITCHABLE TERMS | Stem Phrase Switching | Phrase Variants Based on Stemming |
| 7 | WORD FILE ACCESS - (SETPCT) MATCH REQUIRED | Word File Switching | Phrase Variants Based on Wor |
| 8 | STEM FILE ACCESS - (SETPCT) MATCH REQUIRED | Word Stemming | Word Variants |
| 9 | CONCEPT FILE ACCESS - ACCEPT RELATED TERMS ONLY | Co-Related Term Switching | Co-Related Terms |
| 10 | CONCEPT FILE ACCESS - ACCEPT NARROW TERMS ONLY | Narrower Term Switching | Generic Terms |
| 11 | CONCEPT FILE ACCESS - ACCEPT BROAD TERMS ONLY | Broader Term Switching | Generic Terms |
| 12 | TERM FILE ACCESS - LIST (2*SETADJ) ADJACENT TERMS | Adjacency Switching | Adjacent Terms |
| 13 | WORD FILE ACCESS - LIST (2*SETADJ) ADJACENT WORDS | -- | Adjacent Words |
| 14 | STEM FILE ACCESS - LIST (2*SETADJ) ADJACENT PHRASES | -- | Adjacent Stems |
| 15 | STEM PHRASE FILE - LIST (2*SETADJ) ADJACENT PHRASES | -- | Adjacent Phrases |
| 16 | ACCEPT LEAD + USE + DOUBLE USE TERMS | -- | Synonyms |
| 17 | ACCEPT LEAD + USE + DOUBLE USE + USE-FOR TERMS | -- | Synonyms |
| 18 | ACCEPT LEAD + USE + DOUBLE USE + USE FOR + DOUBLE USED-FOR TERM | -- | Synonyms |
| 19 | ACCEPT LEAD + MULTIPLE USE + MULTIPLE USED-FOR | Exhaustive Synonym Switching | Synonyms |
| 20 | ACCEPT LEAD + MULTIPLE USE + MULTIPLE USED-FOR + RELATED TERMS | -- | Synonyms Plus Co-Related Terms |
| 21 | CONCEPT FILE ACCESS - ACCEPT RELATED TERMS ONLY | Related Term Switching | Related Terms |

FIGURE 1 MENU-DRIVEN MODE OF VSS



```
WELCOME TO *VSS* - VOCABULARY SWITCHING SYSTEM
USING BATTELLE'S DATA MANAGEMENT SYSTEM, BASIS

VSS CONTAINS FOUR VOCABULARY SETS
 1- BUSINESS
    A. ABI              B. MANAGEMENT CONTENTS
 2- SOCIAL SCIENCE
    A. ERIC             B. PSYCH ABSTRACTS
 3- LIFE SCIENCE
    A. BIOSIS           B. CA
    C. MESH
 4- PHYSICAL SCIENCE
    A. DOE              B. CA
    C. EI               D. INSPEC
    E. IRON             F. NASA


PLEASE SELECT 1 OF THE 4 VOCABULARY SETS
  BY ENTERING EITHER 1, 2, 3, OR 4
     ?3

REQUESTED FILE SET IS ONLINE - CONTINUE

PLEASE SELECT THE VOCABULARIES OF INTEREST
 BY ENTERING THE LETTER(S) SEPARATED BY COMMAS;
 FOR EXAMPLE: A,B,C
 3- LIFE SCIENCE
    A. BIOSIS           B. BIOSIS-C
    C. CA               D. MESH
    E. MESH-R           F. MESH-S
      ?a,b,c,d,e,f

VSS PROVIDES FOR 5 SWITCHING FEATURES:
  1- SYNONYMS
  2- BROWSE
  3- NARROWER TERMS
  4- BROADER TERMS
  5- NARROWER/BROADER TERMS
  6- OTHER(USER-DEFINED)

PLEASE SELECT 1 OF THE 6 OPTIONS
  BY ENTERING EITHER 1, 2, 3, 4, 5, OR 6
      ?2

SPECIFY THE MAXIMUM NUMBER OF TERMS
 TO BE DISPLAYED PER VOCABULARY.
 ENTER A NUMBER.
      ?6

PLEASE ENTER SEARCH TERM OR VSS COMMAND
     ?9005-97-4

SWITCH SUCCESSFUL

TERM TYPE     VOCAB     TERM

YOUR TERM     MESH-R    9005-97-4
SYNONYM +     MESH      USE *IODOPROTEINS CASEIN/*ANALOGS ORED
SYNONYM +     MESH-S    PROTAMINE
ADJ-LEAD      MESH-R    9004-94-8
ADJ-LEAD      MESH-R    9004-99-3
ADJ-LEAD      MESH-R    9005-32-7
ADJ-LEAD      MESH-R    9005-66-7
ADJ-LEAD      MESH-R    9005-67-8
```

FIGURE 2 VSS TERMINAL SESSION IN MENU MODE

# 6.0 VSS RESEARCH ACTIVITIES

Research proceeded along two fronts; software development and system evaluation.

## 6.1 Software Development

During the course of work on NSF Grant IST-7911190 (1979-81), VSS software was converted over to a new release of the data management software and now runs under BASIS 4.0. Also, ten new vocabularies were acquired for VSS, necessitating the rebuilding of all VSS files.

During Grant IST-8111497 (1982-83), a completely new user interface was written for VSS to make it easier to use the system and interpret output. This was done in anticipation of a large-scale field test with no user training. Also, the related term relationship found in many of the 15 vocabularies was added to the VSS files, and the concept file was redesigned to handle this new relationship. Finally, all sanitized terms in the concept file (terms without special characters) were replaced with unsanitized terms to more accurately represent the search terms actually used by the data bases accessible via VSS vocabularies.

## 6.2 System Evaluation

Several types of evaluations were conducted with VSS:

- formal evaluation of selected switching strategies
- formal evaluation involving end users, intermediaries, VSS, and publicly available data bases
- informal evaluation involving information brokers, librarians, information science and library school graduate students, and data base vendor staff

In addition, the 1979 On-line Users Survey was updated. Each of these evaluation efforts and the on-line survey is described separately in subsequent sections of this report. Each section contains a description of the test methodology and the results obtained.

## 7.0 EVALUATION OF SWITCHING STRATEGIES

With the addition of ten new vocabularies (a total of 15 unique vocabularies altogether) and with the clustering of all vocabularies into one of four major modules by subject matter, it was felt that a new set of term-level analyses, similar to those conducted earlier (10), was needed. These analyses help determine the performance of various switching strategies.

The objective of this evaluation was to determine the performance of several switching strategies with the goal of utilizing only the best ones in formal and informal field evaluations of the system. A further objective was to compare the performance of each major module to determine the effects, if any, of subject area, vocabulary size, and number of vocabularies on switching performance.

### 7.1 Methodology

A switching strategy consists of one or more switching options (see Table 4) and zero to several system commands (Table 3). These options and commands are set-up by the user in a specified sequential order for execution by VSS. Each strategy to be analyzed was numbered to facilitate reporting.

Seven switching strategies were defined and evaluated within each of the four VSS modules. However, the life sciences module was broken down into three modules, small, medium and large, in order to study the effect of module size on switching performance. Therefore, 42 unique analyses were performed (7 strategies x 6 modules).

The seven switching strategies consisted of the following switching options and system commands:

| Strategy Number | Type of Strategy | Switching Options (in order of execution) | System Commands |
|---|---|---|---|
| 1 | browse | 3,4,19,9,6,10,7,8,12 | scope, setpct=50, fullv |
| 2 | browse | 3,4,6,7,8,20 | same as (1) |
| 3 | broader | 3,4,19,11 | scope, fullu |
| 4 | narrower | 3,4,10,19 | scope, expand, fullv |
| 5 | synonym | 3,4,19,6,9 | same as (4) |
| 6 | browse | 3,4,6,7,8,19,9 | same as (4) |
| 7 | browse | 8,19 | same as (4) |

Each strategy was designed to test different combinations of switching options.

Some strategies invoked relatively simple switching. For example, strategy 7 involved only word stems and synonyms. Other strategies were rather complex and included many features available in VSS.

The analyses were performed by passing a selected set of search terms through each of the seven pre-defined strategies and comparing VSS output to the input. Four different sets of search terms were used as input, one tailored for each of the four major modules in VSS: business, social sciences, life sciences, and physical sciences.

These term sets were derived from actual queries submitted to searchers who routinely provide online retrieval services. The business terms were provided by the Public Library of Columbus and Franklin County (PLCFC); social science terms were obtained from the Mechanized Information Center (The Ohio State University) and PLCFC; life science terms were provided by the Health Sciences Library (The Ohio State University); and physical science terms were acquired from the Battelle Library. Appendix D shows each of the four term sets used for the analyses. Since output was voluminous, the term sets were randomly sampled to estimate performance for each of the 42 unique combinations mentioned earlier. These randomly selected terms are identified in the term sets shown in Appendix D.

Switching strategy performance was estimated by determining the amount of relevant VSS output compared to the input search term and the amount of CPU time necessary to perform a particular switching strategy.

The analyses were conducted in the following manner:

(1) An entire term set was used as input for each of the seven strategies, for the appropriate VSS module (e.g. business terms were used as input to seven switching strategies in the business module, life science terms became input to the same seven strategies for three life science modules, etc).

(2) All VSS output was saved for subsequent analysis. Each of the four input term sets was randomly sampled to estimate switching performance for each set. A 10 percent sample was taken from the physical science term set and a 20 percent sample was taken from each of the three remaining sets.

(3) Relevance judgements were made bye the project staff for all VSS output relative to the input terms in the respective samples. This provided an estimate of the quality of each switching strategy. Quantity of output for each strategy was also compiled. An overall performance value was the product of quality and quantity.

(4) CPU time was compiled for each analysis and included as a performance measure.

A typical example of the type of analysis performed is shown in the example below:

Example

TERM-LEVEL ANALYSIS

| INPUT TERM | VSS OUTPUT | RELEVANCE DECISIONS | PERFORMANCE |
|---|---|---|---|
| DIALYSIS. | DIALYSIS | relevant | 50% relevant output |
| | CC 64712 | relevant | |
| | DIALYSANCE | non-relevant | |
| | DIALYSATE | non-relevant | |
| NEXT TERM | NEXT OUTPUT | more decision | X% relevant output |
| | TOTAL OUTPUT | TOTAL RELEVANCE | overall % relevant |

It must be pointed out that once an output result was judged relevant or non-relevant, that decision was consistently applied across all seven switching strategies which produced that identical output. In other words, if DIALYSATE was judged irrelevant for switching strategy 1 in the life science module, then by convention it was irrelevant if strategies 2 through 7 also produced this output where DIALYSIS was the input term.

## 7.2 Results

A composite of all 42 analyses is shown in Table 5. This table lists the overall performance value determined for each individual analysis as well

as the overall ranking of strategies and VSS modules. The overall performance value was the product of output quantity and estimated quality based on relevance. For example, the performance value of 75 for switching strategy 1 within the VSS business module represents the product of an output calculated at 89% and a relevance estimate of 84% (89 x 84 = 75)."

It can be seen from this table that switching strategies 1, 2, and 6 produced consistently higher performance scores across the board than any of the other four strategies. These three strategies scored particularly high in the business and social sciences modules which can be explained by the fact that these strategies were among the most robust. The fact that strategy 2 performed slightly better than strategy 1 in each module is undoubtedly due to the fact that it excluded some features which tend to produce non-relevant output, including such features as stemming and adjacency options.

Several factors contributed to the low ranking of strategies 3 and 4. First, broader/narrower terms were not available in business or life sciences modules, hence, figures in Table 5 for these modules reflect only the synonym switching taking place. However, in both the social sciences and physical sciences modules where broader/narrower terms were available, performance values were still the lowest of all seven strategies. This can partially be explained by a second factor, the difficulty of making a relevance decision for broader and narrower terms in the absence of the end user. Of all the syndetic relationships available, the hierarchical one is the most dependent on an end-user judgement. In this study, we erred on the side of non-relevance when there was any doubt.

Looking at VSS modules, it can be seen that the social science module produced consistently higher scores than all other VSS modules for all but two switching strategies (1 and 2). Only the business module outperformed the social science modules for strategies 1 and 2. Likewise, the business module produced consistently higher scores than either the life science or physical science modules.

TABLE 5 PERFORMANCE OF VSS MODULES

| | Life Sciences* | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 75** | 68 | 28 | 23 | 23 | 39 | 2 |
| 2 | 76 | 68 | 36 | 25 | 25 | 40 | 1 |
| 3 | 40 | 60 | 15 | 14 | 10 | 29 | 6 |
| 4 | 40 | 57 | 15 | 14 | 10 | 24 | 7 |
| 5 | 51 | 67 | 30 | 18 | 19 | 35 | 4 |
| 6 | 54 | 73 | 31 | 23 | 23 | 36 | 3 |
| 7 | 44 | 65 | 25 | 18 | 19 | 28 | 5 |
| Module Ranking | 2 | 1 | 4 | 5 | 6 | 3 | |

*small module - BIOSIS, BIOSIS-C, MeSH
medium module - BIOSIS, BIOSIS-C, MeSH, MeSH-R, MeSH-S
large module - BIOSIS, BIOSIS-C, MeSH, MeSH-R, MeSH-S, CA

**the performance measure is the product of VSS output (in %) and estimated relevance of output (in %)

These results suggest that the more similar the vocabularies with respect to subject matter and syndetic constructs, the better they will function as switching vocabularies. This result is what one would expect. In this research, the two social science vocabularies were similar in subject matter and syndetic constructs; the business vocabularies were very similar in subject matter but differed considerably in syndetic structure. Both the life science and physical science vocabularies differed in subject matter and structure.

An unexpected result was the inverse relationship between module size and performance. In this particular case, the "small" life science module performed consistently higher than the "large" life science module across the seven switching strategies. We suspect that at least two factors contributed to this anomalous result: (1) the vocabularies became increasingly difficult to evaluate with increasing module size because of the abstract nature of BIOSIS concept codes and CA registry numbers (we erred conservatively when unsure about a code or registry number), and (2) an apparently higher degree

of incompatibility among these vocabularies, than originally thought, both from content and syndetic viewpoints.

Tables 6 through 11 summarize the performances for each of the VSS modules studied. Column A in each table shows the maximum possible coverage. This number represents the number of terms submitted for possible switching multiplied by the number of target vocabularies specified multiplied by the vocabulary count specified. In other words, if two terms, A and B, are submitted for switching across two vocabularies, $V_1$ and $V_2$, and a vocabulary count of two is specified (meaning the user will accept up to two switched outputs per vocabulary), the maximum possible coverage, if switching were 100 percent efficient, is 8 (2x2x2).

For this analysis, the maximum possible coverage was different for each module, according to the schedule below:

| Module | Number of Terms | Number of Vocabularies | Vocabulary Count | Coverage |
|---|---|---|---|---|
| Business | 143 | 2 | 1 | 286 |
| Social Science | 113 | 2 | 1 | 226 |
| Life Science | 222 | 3-6 | 1 | 666-1332 |
| Physical Science | 666 | 6 | 1 | 3996 |

Column B, in Tables 6 to 11, records actual coverage, that is the actual amount of output generated by VSS for a particular strategy. Column C shows the percent coverage [(A ÷ B) X 100] achieved by a particular switching. Coverage ranged from 98 percent to 19 percent, with the social science module exhibiting the best coverage percentages among the four modules.

Columns D and E record the results of relevance assessments made on all the VSS output generated by the sampled input. Sample size was 20 percent of the input term sets for business, social and life sciences, or 28, 22, and 44 terms respectively. For physical sciences, the sample size was 10 percent, or 66 terms. Relevant output from these sampled input terms was recorded in column D, non-relevant terms in column E.

Column F shows the estimated percentage of relevant output for each strategy/module combination. The range was 100 to 40 percent relevance.

Adjusted coverage, column G, is the product of percent coverage (column C) and percent relevance (column F) and was used as the benchmark for comparing strategies and modules (Table 5).

The final column, H, shows the cumulative CPU time to execute a complete set of terms through the modules. In other words, it took 20.5 CPU seconds to perform switching on 143 business terms in the business module for strategy 1, or an average of 0.14 CPU seconds per input term. It can be seen that broader/narrower term switching (strategies 3 and 4) were the most efficient while browse switching with stemming and phrase parsing occurring early in the strategy was roughly 100 times slower, and therefore, extremely inefficient.

Table 12 summarizes the detailed performance of each switching strategy by VSS vocabulary module. In these tables the progressive effect of each switching option can be observed for each strategy, and comparisons of individual options across the various modules are readily made.

It can be seen that exact matching, that is, a character-by-character comparison of users' terms with inverted file terms, contributed a significant portion of the actual output produced by VSS. Of course, exact matching is nothing more than an indication of the subject overlap among controlled vocabularies, ignoring homography. Naturally this result is a constant across all switching strategies because the vocabularies remained constant throughout the reporting period.

In the case of the business and social sciences modules, exact matching produced the greatest output, 37 and 49 percent respectively, in all strategies tested except for strategy 7 which did not incorporate this option. This result indicates the degree to which these vocabularies are compatible compared to life and physical sciences.

Both life and physical science modules produced considerably less output with the exact matching option than either the business or social science modules, indicating a tendency toward subject dissimilarity.

The word option in VSS was the next best producer of VSS output in strategies where it was used with the exception of strategy 6. This, of course, says nothing about the quality (relevance) of output produced by this

TABLE 6 PERFORMANCE OF BUSINESS MODULE

| Switching Strategy | (A) Total Coverage Possible | (B) Actual Coverage | (C) % Coverage | (D) R | (E) NR | (F) % R | (G) Adjusted Coverage CxF | (H) Cum Time (Sec) |
|---|---|---|---|---|---|---|---|---|
| BWDS 1 | 286 | 254 | 89 | 47 | 9 | 84 | 75 | 20.5 |
| BWDS 2 | 286 | 241 | 84 | 45 | 5 | 90 | 76 | 19.2 |
| BWDS 3 | 286 | 115 | 40 | 30 | 0 | 100 | 40 | 0.9 |
| BWDS 4 | 286 | 115 | 40 | 30 | 0 | 100 | 40 | 0.9 |
| BWDS 5 | 286 | 146 | 51 | | | 100 | 51 | 3.7 |
| BWDS 6 | 286 | 169 | 59 | 30 | 3 | 91 | 54 | 99.3 |
| BWDS 7 | 286 | 167 | 59 | 24 | 8 | 75 | 44 | 13.7 |

TABLE 7 PERFORMANCE OF SOCIAL SCIENCE MODULE

| Switching Strategy | (A) Total Coverage Possible | (B) Actual Coverage | (C) % Coverage | (D) No. R | (E) No. NR | (F) % R | (G) Adjusted Coverage CxF | (H) Cum Time (Sec) |
|---|---|---|---|---|---|---|---|---|
| PWDS 1 | 226 | 221 | 98 | 34 | 15 | 69 | 68 | 19.3 |
| PWDS 2 | 226 | 216 | 96 | 32 | 13 | 71 | 68 | 18.0 |
| PWDS 3 | 226 | 135 | 60 | 26 | 0 | 100 | 60 | 1.5 |
| PWDS 4 | 226 | 127 | 57 | 25 | 0 | 100 | 57 | 4.1 |
| PWDS 5 | 226 | 142 | 67 | 24 | 0 | 100 | 67 | 3.8 |
| PWDS 6 | 226 | 168 | 75 | 30 | 1 | 97 | 73 | 144.1 |
| PWDS 7 | 226 | 167 | 74 | 21 | 3 | 88 | 65 | 42.0 |

R = amount of relevant output based on random sampling
NR = amount of non-relevant output based on random sampling
Cum time = cumulative CPU time for entire strategy

TABLE 8 PERFORMANCE OF SMALL LIFE SCIENCE MODULE

| Switching Strategy | (A) Total Coverage Possible | (B) Actual Coverage | (C) % Coverage | (D) R | (E) NR | (F) % R | (G) Adjusted Coverage CxF | (H) Cum. Time (Sec) |
|---|---|---|---|---|---|---|---|---|
| NWDS 1 | 666 | 451 | 68 | 69 | 101 | 41 | 28 | 65 |
| NWDS 2 | 666 | 415 | 62 | 44 | 32 | 58 | 36 | 64 |
| NWDS 3 | 666 | 173 | 26 | 29 | 22 | 57 | 15 | 2.7 |
| NWDS 4 | 666 | 173 | 26 | 29 | 22 | 57 | 15 | 3.1 |
| NWDS 5 | 666 | 221 | 33 | 39 | 4 | 91 | 30 | 9.8 |
| NWDS 6 | 666 | 255 | 38 | 41 | 9 | 82 | 31 | 217 |
| NWDS 7 | 666 | 251 | 38 | 33 | 16 | 67 | 25 | 57 |

TABLE 9 PERFORMANCE OF MEDIUM LIFE SCIENCE MODULE

| Switching Strategy | (A) Total Coverage Possible | (B) Actual Coverage | (C) % Coverage | (D) R | (E) NR | (F) % R | (G) Adjusted Coverage CxF | (H) Cum. Time (Sec) |
|---|---|---|---|---|---|---|---|---|
| OWDS 1 | 1110 | 623 | 56 | 65 | 93 | 41 | 23 | 203 |
| OWDS 2 | 1110 | 546 | 49 | 54 | 49 | 52 | 25 | 199 |
| OWDS 3 | 1110 | 207 | 19 | 35 | 11 | 76 | 14 | 6 |
| OWDS 4 | 1110 | 207 | 19 | 35 | 11 | 76 | 14 | 6 |
| OWDS 5 | 1110 | 275 | 25 | 44 | 16 | 73 | 18 | 16 |
| OWDS 6 | 1110 | 368 | 33 | 50 | 21 | 70 | 23 | 492 |
| OWDS 7 | 1110 | 341 | 30 | 37 | 29 | 59 | 18 | 83 |

TABLE 10 PERFORMANCE OF LARGE LIFE SCIENCE MODULE

| Switching Strategy | (A) Total Coverage Possible | (B) Actual Coverage | (C) % Coverage | (D) R | (E) NR | (F) % R | (G) Adjusted Coverage CxF | (H) Cum. Time (Sec) |
|---|---|---|---|---|---|---|---|---|
| MWDS 1 | 1332 | 767 | 58 | 66 | 98 | 40 | 23 | 91 |
| MWDS 2 | 1332 | 669 | 50 | 56 | 56 | 50 | 25 | 89 |
| MWDS 3 | 1332 | 253 | 19 | 34 | 32 | 52 | 10 | 11 |
| MWDS 4 | 1332 | 253 | 19 | 34 | 32 | 52 | 10 | 11 |
| MWDS 5 | 1332 | 339 | 25 | 45 | 14 | 76 | 19 | 15 |
| MWDS 6 | 1332 | 439 | 33 | 52 | 22 | 70 | 23 | 517 |
| MWDS 7 | 1332 | 403 | 30 | 43 | 26 | 62 | 19 | 86 |

TABLE 11 PERFORMANCE OF PHYSICAL SCIENCE MODULE

| Switching Strategy | (A) Total Coverage Possible | (B) Actual Coverage | (C) % Coverage | (D) R | (E) NR | (F) % R | (G) Adjusted Coverage CxF | (H) Cum. Time (Sec) |
|---|---|---|---|---|---|---|---|---|
| SWDS 1 | 3996 | 3318 | 83 | 123 | 139 | 47 | 39 | 203 |
| SWDS 2 | 3996 | 3073 | 78 | 105 | 102 | 51 | 40 | 450 |
| SWDS 3 | 3996 | 1157 | 30 | 66 | 3 | 96 | 29 | 20 |
| SWDS 4 | 3996 | 1075 | 28 | 52 | 8 | 87 | 24 | 33 |
| SWDS 5 | 3996 | 1944 | 49 | 90 | 35 | 72 | 35 | 39 |
| SWDS 6 | 3996 | 2645 | 66 | 99 | 82 | 55 | 36 | 5140 |
| SWDS 7 | 3996 | 2525 | 63 | 70 | 85 | 45 | 28 | 306 |

option. However, it can be observed that the word option produced consistently high amounts of output across all VSS modules (Table 12). To fully understand this result, it is important to understand the word option in VSS.

The word option breaks down a user's term into its component words and searches the inverted word file with them. A scoring algorithm internal to VSS applies a user-supplied threshold value to word file hits before displaying any VSS output. The default threshold is 100 percent. For example, if the input term is NUCLEAR POWER, and the default of 100 is acceptable, then all VSS output must have 100 percent of the input term(s) present. Thus, NUCLEAR POWER PLANT SITINGS satisfies the default. Similarly, if a synonym relationship exists, where the invalid entry contains 100 percent of the input term, then the word option in VSS is satisfied and the valid term is displayed as output. The word option in VSS can be a source of prolific output because the threshold can be set lower than 100 percent by the user. Also, terms longer than the input term will satisfy a default of 100 percent provided the VSS term contains all of the input term.

For example, if a user enters ESTROGEN in the life science module, there are many terms within the MeSH, MeSH-R and MeSH-S vocabularies containing this term (ESTROGEN SULFOTRANSFERASE, ESTROGEN 2-HYDROXYLASE, ESTROGEN ANALOGS to mention a few) which could satisfy the word option. Thus, the word option tends to produce volumn output, especially at lower threshold values and in cases where single-word terms are entered. Thresholds of 50 and 100 percent were used for strategies 1 and 6, respectively.

Vocabulary switching behaves in a manner virtually identical to data base retrieval theory, that is, the higher the recall, the lower the precision. The principle difference between VSS strategies 2 and 6 is the output produced via the word option (option 20 is in reality options 19 and 9 combined).

Strategy 2 represents high coverage (recall) because a low threshold (50%) for word and stem options was specified while strategy 6 represents high relevance (precision) because a high threshold (100%) for word and stem options was employed. These two strategies behave as one would expect--high recall, low precision and vice versa.

The related phrase option produced output equal to or exceeding the various VSS synonym options. Also, it is observed that this option produced a

TABLE 12 DETAILED PERFORMANCE OF INDIVIDUAL SWITCHING STRATEGIES

| Switching Options | Type of Switching | Business | Social Sciences | Life Sciences | | | Physical Sciences |
|---|---|---|---|---|---|---|---|
| | | | | Small | Medium | Large | |
| **Switching Strategy (1)** | | | | | | | |
| 3 | exact match | 37 | 49 | 19 | 12 | 13 | 19 |
| 4 | synonym | 2 | 4 | 6 | 4 | 3 | 5 |
| 19 | synonym | 1 | 3 | 1 | 3 | 3 | 1 |
| 9 | co-related | 0 | 0 | 0 | 1 | 1 | 0 |
| 6 | related phrase | 11 | 11 | 7 | 4 | 5 | 23 |
| 10 | narrower | 0 | 0 | 0 | 0 | 0 | 1 |
| 7 | word | 28 | 27 | 27 | 22 | 23 | 24 |
| 8 | stem | 5 | 2 | 2 | 4 | 3 | 5 |
| 12 | adjacency | 5 | 2 | 6 | 6 | 7 | 5 |
| Total Output (in %) | | 89 | 98 | 68 | 56 | 58 | 83 |
| Estimated Relevance (in %) | | 84 | 69 | 41 | 41 | 40 | 47 |
| Overall Strategy Performance (in %) | | 75 | 68 | 28 | 23 | 23 | 39 |
| **Switching Strategy (2)** | | | | | | | |
| 3 | exact match | 37 | 49 | 19 | 12 | 13 | 19 |
| 4 | synonym | 2 | 4 | 6 | 4 | 3 | 5 |
| 6 | related phrase | 11 | 12 | 7 | 5 | 5 | 23 |
| 7 | word | 29 | 29 | 27 | 23 | 25 | 26 |
| 8 | stem | 5 | 2 | 2 | 4 | 4 | 5 |
| 20 | synonyms & co-related | 0 | 0 | 1 | 1 | 1 | 0 |
| Total Output (in %) | | 84 | 96 | 62 | 49 | 50 | 78 |
| Estimated Relevance (in %) | | 90 | 71 | 58 | 52 | 50 | 51 |
| Overall Strategy Performance (in %) | | 76 | 68 | 36 | 25 | 25 | 40 |

TABLE 12 (Continued)

| Switching Options | Type of Switching | Business | Social Sciences | Life Sciences: Small | Life Sciences: Medium | Life Sciences: Large | Physical Sciences |
|---|---|---|---|---|---|---|---|
| **Switching Strategy (3)** | | | | | | | |
| 3 | exact match | 37 | 49 | 19 | 12 | 13 | 19 |
| 4 | synonyms | 2 | 4 | 6 | 4 | 3 | 5 |
| 19 | synonyms | 1 | 3 | 1 | 3 | 3 | 1 |
| 11 | broader term | 0 | 4 | 0 | 0 | 0 | 5 |
| Total Output (in %) | | 40 | 60 | 26 | 19 | 19 | 30 |
| Estimated Relevance (in %) | | 100 | 100 | 57 | 76 | 52 | 96 |
| Overall Strategy Performance (in %) | | 40 | 60 | 15 | 14 | 10 | 29 |
| **Switching Strategy (4)** | | | | | | | |
| 3 | exact match | 37 | 49 | 19 | 12 | 13 | 19 |
| 4 | synonyms | 2 | 4 | 6 | 4 | 3 | 5 |
| 10 | narrower terms | 0 | 1 | 0 | 0 | 0 | 3 |
| 19 | synonyms | 1 | 3 | 1 | 3 | 3 | 1 |
| Total Output (in %) | | 40 | 57 | 26 | 19 | 19 | 28 |
| Estimated Relevance (in %) | | 100 | 100 | 57 | 76 | 52 | 87 |
| Overall Strategy Performance (in %) | | 40 | 57 | 15 | 14 | 10 | 24 |
| **Switching Strategy (5)** | | | | | | | |
| 3 | exact match | 37 | 49 | 19 | 12 | 13 | 19 |
| 4 | synonym | 2 | 4 | 6 | 4 | 3 | 5 |
| 19 | synonym | 1 | 3 | 1 | 3 | 3 | 1 |
| 6 | related phrase | 11 | 11 | 7 | 4 | 5 | 23 |
| 9 | co-related | 0 | 0 | 0 | 2 | 1 | 1 |
| Total Output (in %) | | 51 | 67 | 33 | 25 | 25 | 49 |
| Estimated Relevance (in %) | | 100 | 100 | 91 | 73 | 76 | 72 |
| Overall Strategy Performance (in %) | | 51 | 67 | 30 | 18 | 19 | 35 |

TABLE 12 (Continued)

| Switching Options | Type of Switching | Business | Social Sciences | Life Sciences: Small | Life Sciences: Medium | Life Sciences: Large | Physical Sciences |
|---|---|---|---|---|---|---|---|
| **Switching Strategy (6)** | | | | | | | |
| 3 | exact match | 37 | 49 | 19 | 12 | 13 | 19 |
| 4 | synonym | 2 | 4 | 6 | 4 | 3 | 5 |
| 6 | related phrase | 11 | 11 | 7 | 5 | 5 | 24 |
| 7 | word | 6 | 7 | 3 | 8 | 7 | 14 |
| 8 | stem | 2 | 3 | 2 | 2 | 2 | 4 |
| 19 | synonym | 1 | 1 | 1 | 2 | 2 | 0 |
| 9 | co-related | 0 | 0 | 0 | 0 | 1 | 0 |
| Total Output (in %) | | 59 | 75 | 38 | 33 | 33 | 66 |
| Estimated Relevance (in %) | | 91 | 97 | 82 | 70 | 70 | 55 |
| Overall Strategy Performance (in %) | | 54 | 73 | 31 | 23 | 23 | 36 |
| **Switching Strategy (7)** | | | | | | | |
| 8 | related phrase | 58 | 73 | 36 | 28 | 28 | 62 |
| 19 | synonym | 1 | 1 | 1 | 2 | 2 | 1 |
| Total Output (in %) | | 59 | 74 | 37 | 30 | 30 | 63 |
| Estimated Relevance (in %) | | 75 | 88 | 67 | 59 | 62 | 45 |
| Overall Strategy Performance (in %) | | 44 | 65 | 25 | 18 | 19 | 28 |

consistent amount of output regardless of its position in a strategy (Table 12, Strategies 1, 2, 5 and 6). Strategy 7 (Table 12) featured the related phrase option by positioning it first and excluding virtually all other options except a synonym option. Strategy 7 confirms what one would expect of the related phrase option, namely high relevance.

Synonyms proved to be less significant than originally thought. Although VSS has a powerful synonym switching capability, less than 8 percent of all switched output can be attributed to these options. Option 4, a synonym option, accounted for 2 to 6 percent while option 19 contributed another 0 to 3 percent.

One explanation for the lower-than-expected performance of the synonym option is the specificity of synonym terms themselves within individual vocabularies. The more unique a "USE" or "USED FOR" cross reference, the less likely it will be found elsewhere--hence, no switching. For example, consider the entry:

Cosmic ray showers and bursts*
UF Auger showers
cascade showers
cosmic ray jets
EAS
extensive air showers
showers, cosmic ray

Even though six synonyms or quasi-synonyms are given for the entry, "cosmic ray showers and bursts", the probability that any one of them will be found in another vocabulary in VSS is very small.

Another explanation for lower than expected synonym switching is the fewer number of vocabularies per module in this version of VSS. Our earlier work with ten integrated energy vocabularies indicated that Option 4 alone was capable of switching at the rate of about 10 percent (11). Options 9 and 20, which also produce synonyms, were not available back then. Therefore, we conclude that the synergy observed earlier was due to the large number of

*Entry taken from INSPEC Thesaurus

vocabularies per module, and that a similar synergy would be observed in the current version of VSS if it too had 10 or more vocabularies per module.

Also, in this analysis a vocabulary count of one was specified for the seven strategies evaluated. This set-up limited output to one term per vocabulary. Had a higher vocabulary count been specified, more synonyms would have been produced.

Neither broader nor narrower term options produced much VSS output (Table 12, Strategy 1, 3 and 4) in the modules which contained such relationships (social and physical sciences). However, considering the conditions under which these options were analyzed, the result was not too surprising. The conditions were as follows:

(1) The vocabulary count was set at one (1)

(2) If an exact match or synonym was identified in a particular vocabulary, the vocabulary count for that vocabulary was already satisfied, by definition

(3) Only if entered terms had broader or narrower terms could VSS attempt to locate such terms (switch) in other vocabularies where exact match or synonym options had failed.

The adjacency option (strategy 1) contributed moderately toward the overall coverage in various modules; however, the relevance of adjacent terms was low.

In summary, the term-level analyses suggested that strategies 1 and 2 would perform better than any of the others tested for this particular version of VSS. Performance also was noticeably better in the business and social science modules than in life or physical science modules, due primarily to a higher degree of subject overlap among the business and social science vocabularies.

As a result of this analysis, a "browse strategy" was designed for eventual use in the field evaluation of VSS. That browse strategy combined most of the options used in strategies 1 and 2, but it excluded broader and narrower term options and included our new option, related terms. The browse strategy that was eventually field tested is shown in Appendix C. Users were encouraged to set the vocabulary count to a high number (5 to 10) to produce

maximum output for each option within the strategy. In a sense, high-recall switching was encouraged because experience showed that the best use of VSS is as a "shopping list". Users can handle relatively long lists of terms and quickly decide which ones are valuable for eventual search strategy enhancements or modifications.

## 8.0 END USER EXPERIMENTS

Several end user tests were conducted in order to evaluate VSS in actual on-line reference retrieval situations. The purpose of the tests was to compare bibliographic citations produced by VSS-modified searches with those produced under normal conditions.

One series of tests utilized a tandem or iterative design, that is, the original search strategy was modified only after the original query had been negotiated, and the on-line search performed. The test involved one end user and one intermediary.

A second series of tests involved a parallel or simultaneous design. Here, VSS was introduced at the start of the user query and involved two intermediaries for each end user. Each intermediary performed a separate search for the same end user, however, one search involved VSS while the other was conducted in the usual manner.

### 8.1 Methodology

The tandem or iterative test was conducted at a site designated as Test Site 1. It was set up in the following way:

> Subsequent to an end-user query and successful completion of an off-line search session, the intermediary involved noted whether the search just completed was a possible candidate for the VSS test. To be a candidate, the completed search had to involve data bases which VSS could handle via one of its 14 vocabularies. When a candidate search was identified, the end user was contacted and asked to participate in the test. Participation meant that the user's original query would be modified and resubmitted. The user was asked to evaluate all output from both searches for relevance. The second search was done at no cost to the end user. Search terms used in the original on-line search strategy were entered into VSS for possible switching. Then the intermediary evaluated VSS output and modified the original search strategy if deemed appropriate to do so., Next, the modified search was rerun against those data bases named in the original search which matched up with VSS vocabularies. Duplicate citations were eliminated between the two searches within

each unique file (data base). Users received two separate outputs for evaluation: citations produced by the original search and those produced by a VSS-modified search.

The parallel or simultaneous test was conducted at two different sites at two different times (about 2 years apart) using two similar but not identical procedures. Consequently, the results are reported separately for each test site, which are referred to as Test Site 2 and Test Site 3. The test was set up in the following way:

> Both parallel tests employed the same basic experimental design as shown in Table 13. Two intermediaries were involved in each search beginning with the end-user presearch interview. On the odd numbered searches, intermediary A used VSS to develop the search strategy, while intermediary B conducted the same search in his/her usual manner unassisted by VSS. On even numbered searches, the intermediaries reversed their roles. This design was employed to neutralize any skill differences among the intermediaries so that aggregate results for a test site would represent an objective comparison of VSS versus non-VSS searches.

TABLE 13 EXPERIMENTAL DESIGN FOR THE PARALLEL TEST INVOLVING VSS

| End User | Search | Intermediary A | Intermediary B |
|---|---|---|---|
| 1 | 1 | VSS | non-VSS |
| 2 | 2 | non-VSS | VSS |
| 3 | 3 | VSS | non-VSS |
| 4 | 4 | non-VSS | VSS |

The manner in which this design was implemented varied at each site because an attempt was made to improve the procedures for the third site based on the feedback from the second site.

Appendix E contains copies of the forms and procedures used at each site. Separate evaluation forms were designed for the intermediary and the end user. Both sites used a written set of procedures. Intermediaries at both sites were given 1 to 2 hour training sessions in the use of VSS. Procedures were carefully explained and reviewed.

Intermediaries were encouraged to practice with the VSS system for some time before commencing the experiment with end users.

The major differences in experimental methodology between the two sites were as follows:

**Procedures**

- Greater interaction between intermediary and end user was permitted for Test Site 3 experiments than for Test Site 2 experiments. This change was in response to complaints raised by Test Site 2 participants (See Step 5 of procedure in Appendix E).

- Different procedures were outlined for Test Site 3 to isolate VSS citations from non-VSS citations and citations common to both (compare steps 8 and 9 in Test Site 2 with the same steps in Test Site 3).

**Forms**

- A different end user evaluation form was designed for Test Site 3 to match the modified procedures used at Test Site 3 for identifying citation sets. Also, by having the end user complete a performance table on his/her worksheet it was felt that the participant would have an appreciation for questions (Questions 7 and 8) dealing with satisfaction of various sets of citations.

Finally, different versions of VSS were used at each test site. Test Site 2 used an earlier, more cumbersome version of VSS, which required the intermediary to set up his/her own switching strategy and to deal with somewhat cryptic VSS output. Related terms were not available with this earlier version.

All of these conditions were different for Test Site 3, that is, switching strategies were predefined, VSS output avoided cryptic coding, and related terms were available.

## 8.2 Results

A total of eight searches was conducted at three test sites over the course of the two grants. Table 14 summarizes the results for all test sites.

Two searches were conducted at Test Site 1. One search involved the "extraction of flavor from fruit juices", the other was a search for "two-component acrylic adhesives". Tables 15 and 16 show the non-VSS and VSS search strategies and results for these two queries.

The searches differed greatly in subject matter, complexity and number of citations printed off-line. However, the data bases searched were similar in both cases, and user satisfaction with the VSS portion of the output was very favorable in both cases. In fact, the user who requested information on adhesives thought the VSS portion of the output was better than the non-VSS portion, although a definitive reason could not be given for the perceived difference in quality. The person requesting information on fruit juices was simply happy to have the extra relevant citations that a VSS-modified search was able to produce. It can be observed that precision suffered in virtually all the VSS searches; however this performance factor did not seem to affect end user satisfaction with the final results.

As would be expected, VSS produced fewer relevant citations in nearly all searches except the CA search for adhesives. (At this point no explanation can be offered for the extraordinary performance of VSS on the search for acrylic adhesives in the CA file.) The reason that VSS produced fewer relevant citations is that it represented the new citations which the non-VSS searches failed to produce. In a sense, what VSS accomplished in this test was to improve the recall of the non-VSS search.

The parallel or simultaneous test was conducted at Test Sites 2 and 3. A total of six searches was conducted, two at Test Site 2 and four at Test Site 3. Tables 17 and 18 show the detailed results of searches conducted at Site 2 and Tables 19 through 22 show the results of searches at Site 3.

It can be seen that VSS's performance at Test Site 2 was lower compared to Non-VSS performance for search question (B) and virtually non-existent for search question (A). However, end user satisfaction was somewhatsurprising. First of all, the end user associated with question (A), Table 17, gave four evaluations as expected, one for each search combination

performed (ERIC-Non-VSS, ERIC-VSS, PSYCH-Non-VSS, PSYCH-VSS). The end user who submitted question (B), Table 18, evaluated only on the basis of VSS and Non-VSS--an obvious communication breakdown somewhere.

Second, the end user who submitted question (A) gave ratings which, in effect, averaged 2.5 on a 4 point scale for both Non-VSS and VSS searches across the data bases. In other words, there was little difference in the user's mind between these two approaches, yet the ratings given by this end user seem to belie the relevance decisions made. With so few relevant documents in either the Non-VSS set or the VSS set, it seems that the rating is higher than deserved.

This is contrasted with search question (B), Site 2 where the user gave non-VSS searches a rating of 3 and VSS searches a rating of 2 on the same 4-point scale. Here, the ratings seem low in relation to the relevance decisions made, especially for the VSS citation set.

The net result is that one end user (Site 2, search A) gave VSS a 2.5 rating based on a performance of one relevant document out of 295 while another end user (Site 2 search B) gave VSS a rating of 2.0 on a performance of 14 relevant documents out of 302.

The intermediaries in this test remarked that VSS took too long and needed more data bases. Also, they felt that test procedures were too structured and the test eliminated their interaction with the end user. They admitted that these queries were complex, that they do not normally conduct multiple data base searches, and that free-text searching was relied on heavily (maximum recall -- reduced precision). Based on these comments, test procedures and system features were modified prior to the commencement of testing at Site 3.

At Test Site 3, the relevance decisions made for the retrieval set labeled "common" must be added to both the non-VSS and VSS retrieval sets to arrive at comparable results reported for Sites 1 and 2 because "common" represented those citations that would have appeared in either set under normal search conditions.

It appears that users at this site were no more consistent at rating satisfaction relative to the relevance decision than the users at Site 2. For example, the user who submitted question 8301 rated Non-VSS and VSS searches

the same, 3 on a 5 point scale, even though the VSS set contained more irrelevant citations than the Non-VSS search (adding the common set to both). In search 8302, VSS received a higher rating than the Non-VSS search for no apparent reason. Searches 8303 and 8304 seemed to represent a good match between a user's rating and the relevance decisions; however, intermediary performance was inconsistent, particularly in search 8304.

Although minor search strategy differences were expected in the parallel design, search 8304 represents a case where the VSS and Non-VSS strategies differed greatly. Both strategies addressed the concept of tackifiers in a similar fashion but the concept of adhesives was treated quite differently (see Table 22). The result was three very different citation sets. The Non-VSS search achieved a higher performance than the VSS search because the concept of adhesives was treated comprehensively in the Non-VSS search but was virtually ignored in the VSS search. In a sense, this search was a comparison of intermediaries rather than the system.

When end users at Test Site 3 were asked to select only one combination of citation sets, either sets A plus B or sets B plus C, where A was the VSS citations, B was the common citations, and C was the Non-VSS citations, the A plus B combination was chosen in 3 out of 4 searches (8301, 8302, and 8303) while the B plus C combination was chosen in only one search (8304). Again, end user evaluations were not always consistent with the results. For example, the correct response in search 8301 should have been B plus C.

Finally, end users in searches 8301, 8302, and 8303 indicated that they would pay nothing extra for the combination selected, while the end user in 8304 indicated that the combination selected would have been worth an additional $15. It so happens that the combination chosen in 8304 was the Non-VSS search. This combination, Non-VSS plus common, yielded 82 percent relevant output (42 ÷ 52 x 100) while the VSS plus common combination yielded only 67 percent relevant output (38 ÷ 57 x 100). The difference in this case was worth more to this end user.

On the other hand, the end user in search 8303 indicated that the combination chosen, a VSS search with 5 relevant citations, was worth no additional money, even though the Non-VSS search yielded 0 relevant citations. We can only conclude that value truly is in the eye of the beholder.

On balance, VSS held its own in these experiments: At Test Site 1, user satisfaction with VSS was high; at Site 2, Non-VSS searches came out slightly ahead of VSS on the user satisfaction scale; and at Site 3, VSS searches came out slightly ahead.

TABLE 14 SUMMARY OF END USER TESTS AT VARIOUS TEST SITES

| | Search Topic | DB's | Aggregate Search Results | | | User Satisfaction | |
|---|---|---|---|---|---|---|---|
| | | | Non-VSS | VSS | | Non-VSS | VSS |
| Test Site 1 | (A) Flavor from fruit juices | EI<br>FTSA<br>CA<br>NTIS | R=51<br>I=125 | R=12<br>I=66 | | No comment | Very satisfied |
| | (B) Acrylic adhesives | EI<br>CA<br>NTIS | R=141<br>I=242 | R=238<br>I=443 | | Less satisfied | More satisfied |
| Test Site 2 | (A) Library Cataloging and Indexing | ERIC<br>PSYCH | R=5<br>I=85 | R=1<br>I=294 | | 3 (ERIC)(1)<br>2 (PSYCH) | 3 (ERIC)<br>2 (PSYCH) |
| | (B) Counseling, psychotherapy etc. | ERIC<br>PSYCH | R=58<br>I=682 | R=14<br>I=316 | | 3 | 2 |
| | | | Non-VSS | VSS | Common(2) | | |
| Test Site 3 | (8301) Chloro-substituted anilines | Medline<br>BIOSIS | R=0<br>I=1 | R=0<br>I=16 | R=2<br>I=4 | 3 (3) | 3 |
| | (8302) Detergents and hard surface cleaners | CA | R=0<br>I=0 | R=0<br>I=1 | R=11<br>I=15 | 3 | 4 |
| | (8303) Acrylate and methacrylate polymerizations | CA<br>EI | R=0<br>I=2 | R=5<br>I=40 | R=0<br>I=0 | 1 | 3 |
| | (8304) Tackifiers in adhesives | CA<br>EI | R=18<br>I=6 | R=14<br>I=16 | R=24<br>I=3 | 4 | 2 |

R-relevant citations, I-irrelevant citations
(1) A 4-point rating scale (1 = very dissatisfied, 4 = highly satisfied)
(2) Common set must be added to both the Non-VSS and VSS results since it represents the set of citations that would have appeared in either search under non-experimental conditions.
(3) A 5-point rating scale (1 = very dissatisfied, 5 = very satisfied)

TABLE 15 DETAILED SEARCH RESULTS FOR TEST SITE 1 — QUESTION (A)

Search Question (A): Separation of flavor from fruit juices.

Data Bases:

| Non-VSS | VSS |
|---|---|
| EI | EI |
| FSTA | CA |
| CA | NTIS |
| NTIS | |

Search Sets:

Set 1
membrane? ? (or)
ultrafiltrat? ? (or)
reverse (F) osmosis (or)
electrodialysis

Set 2
fruit? ? (or)
food? ? (or)
foods? ? (or)
juice? ?

Set 3
flavor? ?

Set 4
filtration (or)
filtering (or)
separation (or)
osmosis (or)
electroosmosis

Search Strategy:

| Non-VSS | VSS |
|---|---|
| The Non-VSS strategy was "A (and not) B" where | The VSS strategy was "C (and not) D" where |
| (A) Set 1 (and) Set 2 | (C) Set 1 (or) Set 4 (and) Set 2 |
| (B) Set 1 (and) Set 3 | (D) Set 1 (or) Set 4 (and) Set 3 |

Non-VSS and VSS strategies were run 3 calendar days apart.

Results: Number of relevant and irrelevant citations, as determined by end user.

| Data Bases | Non-VSS R | I | T | P | VSS R | I | T | P |
|---|---|---|---|---|---|---|---|---|
| EI | 8 | 11 | 19 | 42 | 1 | 10 | 11 | 9 |
| FSTA | 4 | 30 | 34 | 12 | - | - | - | - |
| CA | 33 | 59 | 92 | 36 | 10 | 39 | 49 | 20 |
| NTIS | 6 | 25 | 31 | 19 | 1 | 7 | 8 | 13 |
| TOTAL | 51 | 125 | 176 | 30 | 12 | 56 | 68 | 18 |

R=relevant citations, I=irrelevant citations, T=total citations, P=precision (R ÷ T x 100)

User Satisfaction: User was very satisfied with the extra output produced by the VSS modified search because it identified 12 additional relevant citations. Rating scale was not used.

TABLE 16 DETAILED SEARCH RESULTS FOR TEST SITE 1 — QUESTION (B)

Search Question (B): State-of-the-art summary on two component acrylic adhesives.

Data Bases Searched:

| Non-VSS | VSS |
|---|---|
| EI | EI |
| CA | CA |
| NTIS | |

Search Sets:

Set 1
adhesive? ?

Set 2
epoxy (F) acrylic (or)
urethane (F) acrylic (or)

Set 3
acrylic (or)
methacrylic (or)
methacrylate? ? (or)
acrylate? ?

Set 4
epoxy (F) vinyl (or)
epoxy (F) vinyl

Set 5
elastomeric (or)
rubbery (or)
toughened (or)
reinforced (or)
resilient

Set 6
block (C) copolymer? ? (or)
graft (C) copolymer? ?

Set 7
phase (F) separat?

Set 8
engineering (or)
structural (or)
cross (W) link AB (or)
crosslink AB

Set 9
curable

Set 10
adhesion

Set 11
vinyl (F) compound? ? (or)
vinyl (F) resin? ?

Set 12
elastomer? ? (or)
rubber? ? (or)
toughness (or)
resilience

Set 13
curing

Search Strategies:

Non-VSS: Complex set of nested statements involving the term sets 1 through 9 above. Many unique citation sets were printed off-line.

VSS: Complex set of nested statements involving all 13 term sets shown above. Several unique citation sets were printed off-line.

Non-VSS and VSS strategies were run 4 calendar days apart.

Results: Number of relevant and irrelevant citations, as determined by end user.

| Data Bases | Non-VSS R | I | T | P | U | VSS R | I | T | P | U |
|---|---|---|---|---|---|---|---|---|---|---|
| EI | 28 | 48 | 76 | 37 | - | 6 | 38 | 44 | 14 | 4 |
| CA | 105 | 176 | 281 | 37 | 17 | 222 | 405 | 627 | 35 | 63 |
| NTIS | 8 | 18 | 26 | 31 | - | - | - | - | - | - |
| TOTAL | 141 | 242 | 383 | 37 | 17 | 228 | 443 | 671 | 34 | 67 |

R=relevant citations, I=irrelevant citations, T=total citations, P=precision (R ÷ T x 100), U=unable to decide relevance.

User Satisfaction: User was more satisfied with the VSS search than the non-VSS search, but unable to articulate why. Rating scale was not used.

TABLE 17 DETAILED SEARCH RESULTS FOR TEST SITE 2 -- QUESTION (A)

Search Question (A): The impact of long range planning, futuristic society, technology, and working conditions on the organizational climate of library cataloging and indexing activities.

Data Bases Searched:

Non-VSS and VSS

ERIC, Psychological Abstracts

Search Sets:

Set 1
organizational (adj) structure (or)
organization climate

Set 2
cataloging (or)
library (or)
libraries (or)
catalogs (or)
indexing

Set 3
long (adj) range (adj) planning (or)
futures with society

Set 4
technology (or)
computers (or)
technological (adj) advancement SI
automation (or)
automatic (or)
automated (or)
automatism (or)
cybernetics (or)
human (adj) factors (adj) engineering (or)
human (adj) engineering

Set 5
working with environment (or)
working with condition (or)
(sic) employer (or)
employee (or)
personnel (or)

Set 6 (Cont'd
employer (or)
job with (satisfaction or analysis or enrichment)

Set 6
performance

Set 7
attitudes (or)
interpersonal (or)
morale (or)
humanization

Set 8
human with relations
group (adj) relation #1
group (adj) relations #1
social (adj) relations #1

Set 9
yr BT 70

Set 10
organization .NJ,MN. (or)
organizational change (or)
organizational effectiveness

Set 11
administrative change (or)
administrative organization

Set 12
task analysis (or)
job analysis (or)
job satisfaction (or)
job performance

Set 13
library technical progress

Set 14
technological advancement (or)
futures of society

Set 15
computers .NJ, MN (or)
automation .NJ, MN

Set 16
individual needs (or)
individual psychology (or)
group behavior (or)
group dynamics

Set 17
work attitudes (or)
employee attitudes

Set 18
library automation (or)
library administration

Set 19
academic libraries (or)
public libraries (or)
research libraries

Search Strategy:

Non-VSS: Complex series of nested statements involving sets 9 through 19, followed by "NOT ing" with VSS strategy.

VSS: Complex series of nested statements involving sets 1 through 9.

Results: Number of relevant and irrelevant citations, as determined by end user.

| Data Bases | Non-VSS R | I | T | P | VSS R | I | T | P |
|---|---|---|---|---|---|---|---|---|
| ERIC | 4 | 40 | 44 | 9 | 1 | 32 | 33 | 3 |
| PSYCH | 1 | 45 | 46 | 2 | 0 | 252 | 252 | 0 |
| TOTAL | 5 | 85 | 90 | 6 | 1 | 284 | 285 | 0 |

R=relevant citations, I=irrelevant citations, T=total citations, P=precision (R ÷ T x 100)

User Satisfaction: User rated the ERIC output (both non-VSS and VSS) 3 on a 4-point scale* and Psychological Abstracts output (both non-VSS and VSS) a 2. The user indicated that his expectations for the amount of relevant citations were met in the non-VSS ERIC search, but not in the non-VSS PSYCH ABS search. The additional relevant citation produced by VSS was judged not too useful. The amount of irrelevant citations (noise) didn't matter to this user.

Intermediary Performance (in minutes):

| | Non-VSS | VSS |
|---|---|---|
| Presearch Preparations Time | 47 | 48 |
| ERIC On-line Time | 27 | 26 |
| PSYCH On-line Time | 21 | 7 |

*A 4-point scale (1=very dissatisfied, 4=highly satisfied)

TABLE 18 DETAILED SEARCH RESULTS FOR TEST SITE 2 — QUESTION (8)

Search Questions (8): The effect of cognitive complexity, conceptual systems, and intelligence on counseling, psychotherapy, behavior modification or persuasion.

Data Bases Searched:

Non-VSS and VSS

ERIC and Psychological Abstracts

Search Sets:

Set 1
intelligence

Set 2
cognitive complexity (or)
cognitive style (or)
conceptual with level
S1 (or)
conceptual (adj) systems
S1

Set 3
counseling (or)
counselor (or)
counselors (or)
psychotherapy (or)
psychotherapist S1 (or)
therapy (or)
therapist S1

Set 4
behavior modification (or)
persuasion (or)
persuasive communication

Set 21
cognitive processes (or)
thought processes (or)
cognitive measurement

Set 22
cognitive with complexity

Set 5
LO-EX

Set 6
cognitive-processes (or)
cognitive-complexity (or)
(cognitive with
measurement)

Set 7
cognitive-style

Set 8
conceptual with (level
S1 or complexity or
systems S1)

Set 9
intelligence.DE.

Set 10
counseling.DE. (or)
group-counseling

Set 23
conceptual schemas (or)
(conceptual with (systems
S1 or level S1 or
complexity))

Set 24
intelligence.DE. (or)
intelligence level

Set 11
counsellors.DE.

Set 12
therapist S.DE. (or)
therapy (or)
psychotherapy.DE. (or)
psychotherapists.DE.

Set 13
behavior.DE.

Set 14
persons S.DE. or)
interpersonal influ-
ences (or)
personality.DE.

Set 15
therapist S.DE. (or)
psychotherap S.DE. (or)
counsellors.DE.

Set 25
counseling.DE. (or)
adult counseling (or)
behavioral counseling

Set 26
therapy.MJ. (or)
therapists.DE (or)
psychotherapy.MJ. (or)
counselors.DE

Set 16
cognitive with complexity (or)
conceptual with level (or)
cognitive processes (or)
cognitive measurement S1 (or)
cognitive style

Set 17
behavior modification (or)
behavior change

Set 18
social influences (or)
social change (or)
social behavior

Set 19
persuasive discourse

Set 20
intelligence.MJ.

Set 27
interpersonal competence (or)
personality.DE. (or)
persuasive discourse

Search Strategy:

Non-VSS (PSYCH SS): Complex series of nested statements involving sets 1 through 8.

VSS (PSYCH SS): Complex series of nested statements involving sets 9 through 18, followed by "NOT ing" with non-VSS strategy.

Non-VSS (ERIC SS): Complex series of nested statements involving Set 1, Set 3 and Sets 16 through 19.

VSS (ERIC SS): Complex series of nested statements involving Set 7, Set 13, and Sets 20 through 27, followed by "NOT ing" with non-VSS strategy.

Results: Number of relevant and irrelevant citations, as determined by end user.

| Data Bases | Non-VSS R | I | T | P | VSS R | I | T | P |
|---|---|---|---|---|---|---|---|---|
| ERIC | 17 | 467 | 484 | 4 | 10 | 123 | 133 | 8 |
| PSYCH | 41 | 215 | 256 | 16 | 4 | 179 | 183 | 2 |
| TOTAL | 58 | 682 | 740 | 8 | 14 | 302 | 316 | 4 |

R=relevant citations, I=irrelevant citations, T=total citations, P=precision (R ÷ T x 100)

User Satisfaction: This user completed only one evaluation form instead of one for each data base searched, indicating a breakdown in communication or failure to follow instructions. The user rated non-VSS results (both ERIC and PSYCH) a 3 and VSS results (both ERIC and PSYCH) a 2 on a 4-point scale.* His expectations were met for the amount of relevant citations produced by the non-VSS search. The additional relevant citations produced by VSS were judged not too useful. Irrelevant citations (noise) were considered tolerable for the non-VSS search but annoying for the VSS search.

Intermediary Performance (in minutes):

| | Non-VSS | VSS |
|---|---|---|
| Presearch Preparation Time | 30 | 41 |
| ERIC On-line Time | 27 | 17 |
| PSYCH On-line Time | 9.5 | 18 |

*A 4-point scale (1=very dissatisfied, 4=highly satisfied

TABLE 19 DETAILED SEARCH RESULTS FOR TEST SITE 3 -- QUESTION (8301)

Search Question (8301): Inhalation and Dermal Toxicity of Chloro-substituted Analines.

Data Bases Searched:

Non-VSS and VSS

Medline, Biosis

Search Sets:

| Set 1 | Set 3 | Set 4 |
|---|---|---|
| Chloroanaline?? (or) | Breathing (or) | Inhal? (or) |
| Dichloroaniline?? | Inhal? (or) | Dermal (or) |
| | Dermal (or) | Dermis (or) |
| Set 2 | Skin (or) | Dermic (or) |
| | Cutaneous (or) | Cutaneous (or) |
| RN=106-47-8 (or) | Cyanosis (or) | Skin (or) |
| RN=95-51-2 (or) | Cyanotic (or) | Methemoglobin ? (or) |
| RN=95-76-1 | Methemoglobinemia | Exposure?? (or) |
| | | Exposed (or) |
| | | Cyanosis (or) |
| | | Cyanic |

Search Strategies:

Non-VSS: Set 1 (and) Set 3

VSS: Set 1 (or) Set 2 (and) Set 4

Results: Number of relevant and irrelevant citations, as determine by end user.

| Data Bases | Non-VSS R | I | T | P | U | VSS R | I | T | P | U | Common R | I | T | P | U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MEDLINE | 0 | 1 | 1 | 0 | 0 | 0 | 5 | 5 | 0 | 0 | 1 | 2 | 3 | 33 | 0 |
| BIOSIS | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 11 | 0 | 0 | 1 | 2 | 3 | 33 | 0 |
| TOTAL | 0 | 1 | 1 | 0 | 0 | 0 | 16 | 16 | 0 | 0 | 2 | 4 | 6 | 33 | 0 |

R=relevant citations, I=irrelevant citations, T=total citations, P=Precision (R ÷ T x 100)
U=unable to decide relevance

Strategy Formulation Time: VSS: 19.5 minutes Non-VSS: 94 minutes

User Satisfaction: 3 out of a maximum of 5 for both VSS and Non-VSS. The VSS was selected over Non-VSS because it gave more references.

Interpretive Comments: VSS - was pleased that registry numbers are in VSS. Non-VSS - most of the time was used examining the Medline thesaurus with which the searcher was unfamiliar.

TABLE 20 DETAILED SEARCH RESULTS FOR TEST SITE 3 -- QUESTION (8302)

Search Question (8302): Applications of an Formulations for Detergents (Mainly Laundry) and Hard Surface Cleaners. No Patents Search Strategy.

Data Bases Searched:

Non-VSS and VSS

Chemical Abstracts

Search Sets:

Set 1

Detergents? ?(L)Builder? ? (or)
Detergent? ?(C)Formulat ? (or)
Detergent? ?(C)Review? ?(C)(or)
PY-1983 (or)
Detergent? ?(L)Builder? ?

Set 2

English or Not a Patent

Set 3

Detergent? ? and Review? ? (or)
Detergent? (or)
Hard (w)Surface? ?(w)Clean? (or)
PY-1983

Set 4

Formulation? ? (or)
Recipe? ? (or)
Builder? ?

Search Strategies:

Non-VSS: Set 1 (and) Set 2

VSS: Set 2 (and) Set 3 (and) Set 4

Results: Number of relevant and irrelevant citations, as determined by end user.

| Data Bases | Non-VSS | | | | | VSS | | | | | Common | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R | I | T | P | U | R | I | T | P | U | R | I | T | P | U |
| CA-FILE 311 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 11 | 15 | 26 | 42 | 1 |

Strategy Formulation Time: VSS: 38 minutes Non-VSS: 19 minutes

User Satisfaction: 4 for VSS and 3 for Non-VSS out of a maximum of 5.

Intermediary Comments: VSS was of very little help for this search question.

TABLE 21 DETAILED SEARCH RESULTS FOR TEST SITE 3 -- QUESTION (8303)

Search Question (8303): Termination mechanism of acrylate and methacrylate radical polymerizations (information usually obtained by kinetic studies of the polymerization reactions).

Data Bases Searched:

Non-VSS and VSS

Chemical Abstracts, Engineering Index

Search Sets:

Set 1

Ethyl(w)acrylate? (or)
Homopolyethyl(2w)acrylate? ? (or)
Polyethyl(w)acrylate? ?(or)
Methyl(2w)Methacrylate? ?(or)
Homopolymethyl(w)Methacrylate? ?(or)
Polymethyl(w)Methacrylate? ?(or)
RN=140-88-5(or)
RN=9003-32-1(or)
RN=80-62-6(or)
RN=9011-14-7

Set 2

Acrylate? ?(L)Polym
Acrylate? ?(L)Polymerization? ?

Set 3

Kinetic? ? (or)
Order? ?(or)
Rate? ?(or)
Mechanism?

Set 4

Radical? ?(or)
Terminat?

Set 5

Kinetics? (or)
Order (or)
Rate? (or)
Mechanis? (or)
Thermodynamic?(or)
Theory

Set 6

Latex?? ?(or)
Emulsion? ? (or)
Dispersion? ?
(not) Aqueous Water(w)born?

Set 7

Review? ?

Set 8

Solution (or)
Bulk

Search Strategies:

Non-VSS: Set 1 (and) Set 3 (and) Set 4 (and) Set 6 (and) Set 7

VSS: Set 2 (and) Set 5 (and) Set 8

Results: Number of relevant and irrelevant citations, as determined by end user.

| Data Bases | Non-VSS R | I | T | P | U | VSS R | I | T | P | U | Common R | I | T | P | U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CA | 0 | 2 | 2 | 0 | 0 | 5 | 40 | 45 | 11 | 6 | 0 | 0 | 0 | 0 | 0 |
| ET | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TOTAL | 0 | 2 | 2 | 0 | 0 | 5 | 40 | 45 | 11 | 6 | 0 | 0 | 0 | 0 | 0 |

R=relevant citations, I=irrelevant citations; T=total citations, P=precision (R ÷ T x 100)
U=unable to decide relevance.

Strategy Formulation Time: VSS: 28.5 minutes Non-VSS: 30 minutes

User Satisfaction: 3 for VSS (maximum 5); 1 for Non-VSS.

TABLE 22 DETAILED SEARCH RESULTS FOR TEST SITE 3 — QUESTION (8304)

Search Question (8304): Composition of Tackifiers and Their Use in Adhesives, Especially Hot Melt.

Data Bases Searched:

Non-VSS and VSS

Chemical Abstracts File 311, Engineering Index

Search Sets:

Set 1

Tackif?

Set 2

Tackifier? ?(or)
Tack (or)
Tackiness

Set 3

PSA? ? (or)
Pressure(w)sensitive (or)
Hotmelt? ? (or)
Hot(w)Melt? ? (or)
Phenol? (or)
Atrylate? (or)
Methacrylate? (or)
Styrene? ? (or)
Alpha(w)methylstyrene?

Set 4

Theory (or)
Improv? (or)
Hot(w)me

Search Strategies:

Non-VSS: Set 1 (and) Set 3

VSS: Set 2 (and) Set 4

Results: Number of relevant and irrelevant citations as determined by end user.

| Data Bases | Non-VSS R | I | T | P | U | VSS R | I | T | P | U | Common R | I | T | P | U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CA | 11 | 2 | 13 | 85 | 1 | 2 | 2 | 4 | 50 | 1 | 19 | 1 | 20 | 95 | 4 |
| EI | 7 | 4 | 11 | 64 | 7 | 12 | 14 | 26 | 46 | 7 | 5 | 2 | 7 | 71 | 4 |
| TOTAL | 18 | 6 | 24 | 75 | 8 | 14 | 16 | 30 | 47 | 8 | 24 | 3 | 27 | 89 | 8 |

R=relevant citations, I=irrelevant citations, T=total citations, P-precision (R ÷ T x 100)
U=unable to decide relevance.

Strategy Formulation Time: VSS: 19 minutes Non-VSS: 17 minutes

User Satisfaction: 4 for Non-VSS and 2 for VSS out of a maximum of 5.

Intermediary Comments: Vocabulary was rather unimportant in this search; limiting the terms available was more important. VSS did not provide any concepts not provided by the end user.

## 9.0 EVALUATION OF VSS BY INTERMEDIARIES

A field evaluation of VSS was planned and conducted between March 14 and May 31, 1983 to elicit response to the system from the information community, primarily information brokers, librarians, information and library science students, and faculty. The evaluation placed few restrictions on participants other than using the system during an assigned period and completing an evaluation form.

The objective was to obtain user reaction to VSS in a field setting, that is, their own work environment with little or no interaction from Battelle. The research team was interested in obtaining data and subjective evaluation both at the detailed level, that is, individual switching transactions, and over-all.

### 9.1 Methodology

A list of potential participants was developed from several sources, including "Fee Based Information Services" by Maranjian and Boss (12). Participation was strictly voluntary and all responses were coded to maintain individual and corporate anonymity. Each organization contacted was provided as many usernames and passwords as desired.

Each participant was given a packet containing an instruction booklet (Appendix C), evaluation form, return envelope, and time schedule (usually a one-week test period), plus two free hours of computer connect time to participate in the evaluation.

The evaluation form developed by the research team consisted of five distinct areas: (1) demographic information; (2) a data base proficiency scale; (3) questions dealing with individual searches using VSS; (4) questions dealing with overall reaction to VSS; and (5) open comments. A sample form is shown in Appendix F.

Questionnaire data were analyzed to obtain the following information:

- Characterization of the study participants
- Correlations between demographics of the participant and the VSS overall evaluations

- Correlations between VSS vocabulary module or switching strategy used and the search results
- Correlations between VSS performance ratings and such factors as: (A) proficiency of participant with a VSS vocabulary module, and (B) differences among the participants themselves as individuals
- Insights from open-ended questions about the system.

## 9.2 Results

### 9.2.1 Characterization of the Study Participants

A total of 65 participants took part in the evaluation (Table 23). About 46 percent were employed in the private sector, 28 percent in the government sector, 23 percent in academia, and 3 percent in the "other" category. About 43 percent of all participants were librarians, 28 percent were employed in some capacity for on-line vendors/data base producers, 15 percent were engaged in some aspect of an information/library science school, and 14 percent were information brokers.

Forty-eight of the participants had at least a Masters Degree, and five had PhD's. Forty-one had at least ten years experience in library/information activity while 40 had five or more years experience with on-line systems. The average years of experience in library/information activities was 13.09 with a median of 12.3 years while the average years of experience with on-line systems was 6.27 with a median of 5.44 years.

Participants in the study were asked to rate their own proficiency with the various data bases whose vocabularies were included in this version of VSS. Table 24 shows the distribution of these self-assigned proficiency ratings on a data base by data base basis.

It can be seen that participants, in general, tended to rate themselves as "average" or as having "little or no" proficiency. Participants rated themselves highest on social science data bases (ERIC, PSYCH Abstracts) and lowest on NASA and DOE Recon data bases. The proficiency rating was used later on to study the effect between this variable and the participant's rating of VSS performance on individual search terms.

TABLE 23 BREAKDOWN OF STUDY PARTICIPANTS BY EMPLOYMENT SECTOR

| Employment Sector | N* |
|---|---|
| 1. Private (N=30) | |
| A. Brokers | 9 |
| B. Librarians | 9 |
| C. On-line Vendor/ Database Producers | 12 |
| 2. Government (N=18) | |
| A. Librarians | 12 |
| B. On-line Vendors/ Database Producers | 6 |
| 3. Academic (N=15) | |
| A. Librarians | 5 |
| B. Library/Information Schools or Depts. | 10 |
| 4. Other (N=2) | |
| A. Public Libraries | 2 |

*Number of participants

## 9.2.2 Correlation Between Participant Demographics and Overall VSS Evaluation

One question of interest in the study was whether the effectiveness of VSS is a function of the participant's experience or education. Questions 4 thru 7 on the evaluation form dealt with experience and education levels of the participants while questions 12 thru 14 were concerned with a participant's subjective perception of the performance of VSS. Pearson correlations were computed between education or experience and the performance ratings given.

Our results indicated that the ease of learning, using or understanding VSS (Question 12), and a participant's confidence (Question 13) with VSS were not dependent on their experience or education. However, when asked

TABLE 24 DISTRIBUTION OF PARTICIPANTS' PROFICIENCY FOR THE VSS DATA BASES

| Proficiency Scale | Data Bases | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ABI Inform | MGT Contents | ERIC | PSYCH Abstracts | MEDLINE | BIOSIS | CA | COMPENDIX | INSPEC | NASA Recon | DOE Recon |
| 1 = Little or none | 24* | 21 | 18 | 16 | 28 | 30 | 30 | 23 | 21 | 46 | 43 |
| 2 = Below average | 1 | 2 | 7 | 4 | 6 | 10 | 8 | 5 | 7 | 3 | 5 |
| 3 = Average | 24 | 27 | 21 | 25 | 11 | 11 | 18 | 15 | 16 | 4 | 8 |
| 4 = Above average | 6 | 6 | 12 | 10 | 6 | 7 | 2 | 12 | 13 | 2 | 2 |
| 5 = Expert | 7 | 7 | 6 | 6 | 10 | 4 | 4 | 8 | 6 | 5 | 2 |
| Missing Value | 3 | 2 | 1 | 4 | 4 | 3 | 3 | 2 | 2 | 5 | 5 |
| Average Proficiency | 2.53 | 2.62 | 2.70 | 2.77 | 2.41 | 2.11 | 2.07 | 2.64 | 2.62 | 1.62 | 1.58 |

*frequency distribution

whether VSS would make their job easier (question 14), the results indicated that participants with fewer years experience felt that VSS would make their jobs easier than did more experienced participants (Table 25 A and B). This result was not surprising since experienced participants would have a tendency to think of themselves as having developed efficient search methods and would be less likely to alter them than less experienced participants.

Interestingly, those participants with higher degrees also tended to indicate that VSS would make their jobs easier. Perhaps those with higher degrees also represented less experienced users, as might be expected, and the true relationship was masked by the presence of confounding factors. The data were not analyzed for the effects of two simultaneous factors.

Tabulation of the individual responses to questions 12 through 14 (Figure 3) indicated the following:

- About 90 percent of the participants felt that VSS was easy-to-moderately easy to learn (Figure 3 A)
- About 79 percent felt that VSS was easy-to-moderately easy to use (Figure 3 B)
- About 75 percent felt that VSS was easy-to-moderately easy to understand (Figure 3 C)
- Somewhat lower percentages of participants, 56 percent and 68 percent felt confident with VSS's capabilities and output, respectively (Figure 3 D and E)
- About 40 percent felt the use of VSS would make their job easier, while only 9 percent felt it would make their job harder (Figure 3 F)

The government sector gave VSS its poorest rating: only 11 percent felt that VSS would make their jobs easier (Table 26). On the other hand, about 50 percent of all participants from each of the three other sectors (private, academic, and "other") felt that VSS would make their jobs easier.

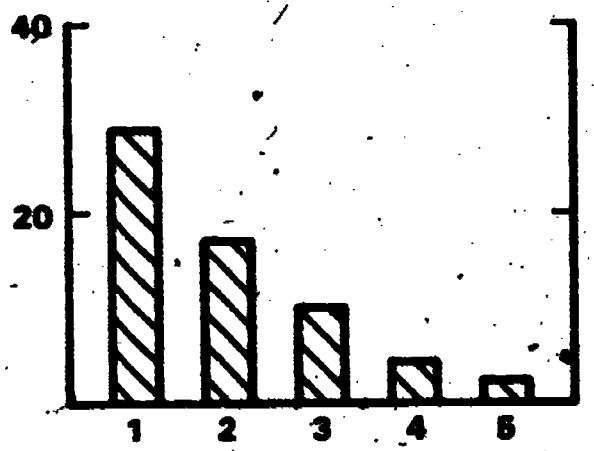
TABLE 25 USER EXPERIENCE VERSUS IMPACT OF VSS ON A SEARCHER'S JOB

(A) Experience in Library/Information Science Field

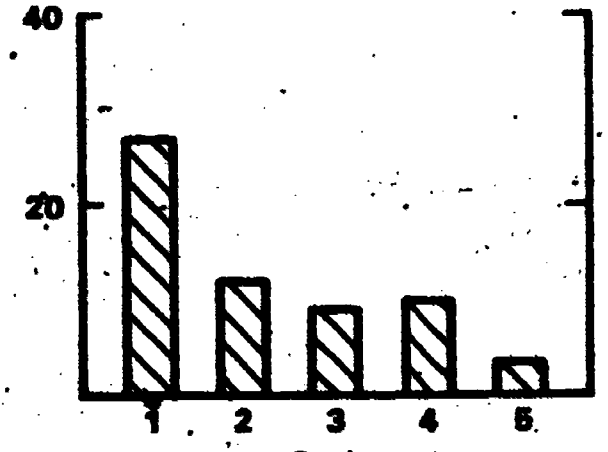
| Years Experience | VSS would make your job of searching multiple data bases: Easier | No Difference | Harder |
|---|---|---|---|
| 0 - 6 | 8 | 7 | 1 |
| 7 - 14 | 10 | 10 | 3 |
| 15 - 20 | 5 | 8 | 0 |
| > 20 | 2 | 6 | 2 |

(B) Experience with on-line systems

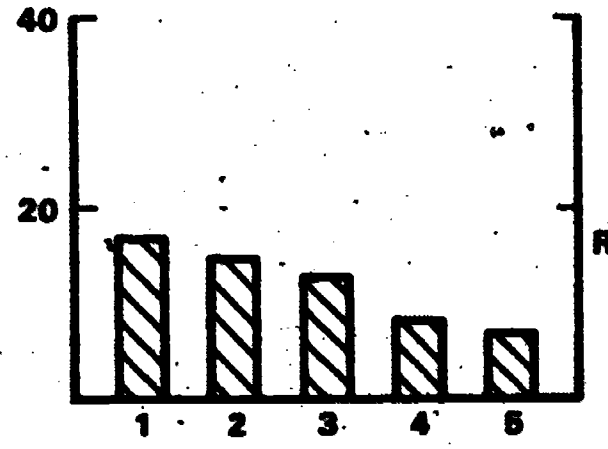
| Years Experience | VSS would make your job of searching multiple data bases: Easier | No Difference | Harder |
|---|---|---|---|
| 0 - 3 | 8 | 7 | 1 |
| 4 - 5 | 9 | 6 | 1 |
| 6 - 9 | 5 | 10 | 1 |
| > 9 | 3 | 8 | 3 |




(A) VSS Easy to Learn

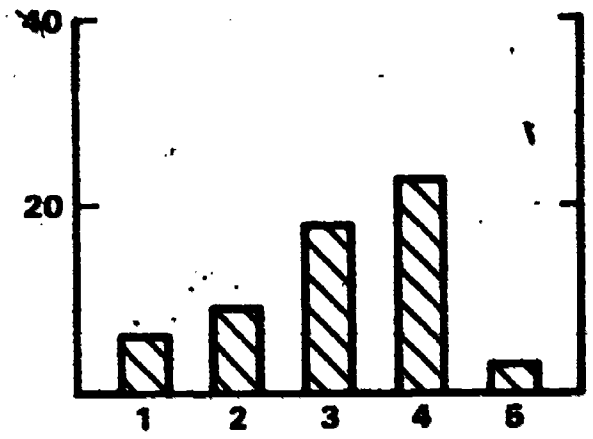



(B) VSS Easy to Use

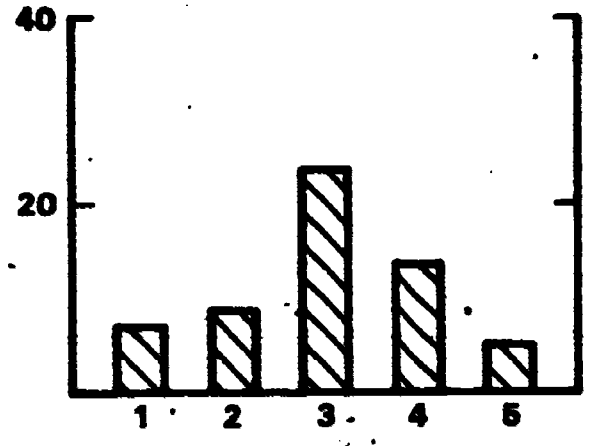



(C) VSS Easy to Understand

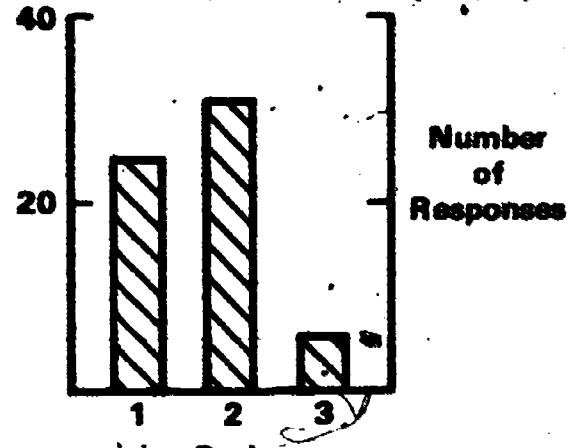
Scale for A, B, C:
1 = Easy
5 = Hard




(D) Confidence With VSS Capabilities




(E) Confidence With VSS Output




(F) How VSS Would Affect Present Job

Scale for D, E:
1 = Confident
5 = No Confidence

Scale:
1 = Easier
2 = No Difference
3 = Harder

FIGURE 3 USER'S RATINGS OF VSS

TABLE 26 IMPACT OF VSS ON JOB PERFORMANCE BY EMPLOYMENT SECTOR

| Sector | VSS would make your job of searching multiple data bases: | | |
|---|---|---|---|
| | Easier | Same | Harder |
| Private | 14 | 11 | 3 |
| Government | 2 | 13 | 3 |
| Academic | 8 | 6 | 0 |
| Other | 1 | 1 | 0 |

VSS allowed each participant a choice of six switching strategies and four vocabulary modules. The objective here was to evaluate VSS strategies and modules against several performance measures, corresponding to:

- Whether the VSS results were usable
- How VSS compared to one's own effort without VSS
- The number of VSS terms that were usable
- The subjective rating of VSS for any given search

Collectively, the participants attempted switching on 623 search terms. Usable results were obtained in 62 percent of the attempts (353 usable, 221 unusable, 49 no response) see Figure 4A. In 48 percent of all attempts VSS was perceived to be "not as good" as what a user's own effort would have been for the same search term, while about 26 percent thought VSS was "better" than what their own effort would have been (Figure 4B). The distribution of usable output is shown in Figure 4C. Table 27 presents a breakdown of the number of usable VSS terms by switching strategy. Strategy 6 produced the greatest percentage of usable terms (92 percent), followed by Strategy 7 (85 percent) and Strategy 2 (65 percent).

TABLE 27 DISTRIBUTION OF SEARCHES BY USABLE VSS OUTPUT BY SWITCHING STRATEGY

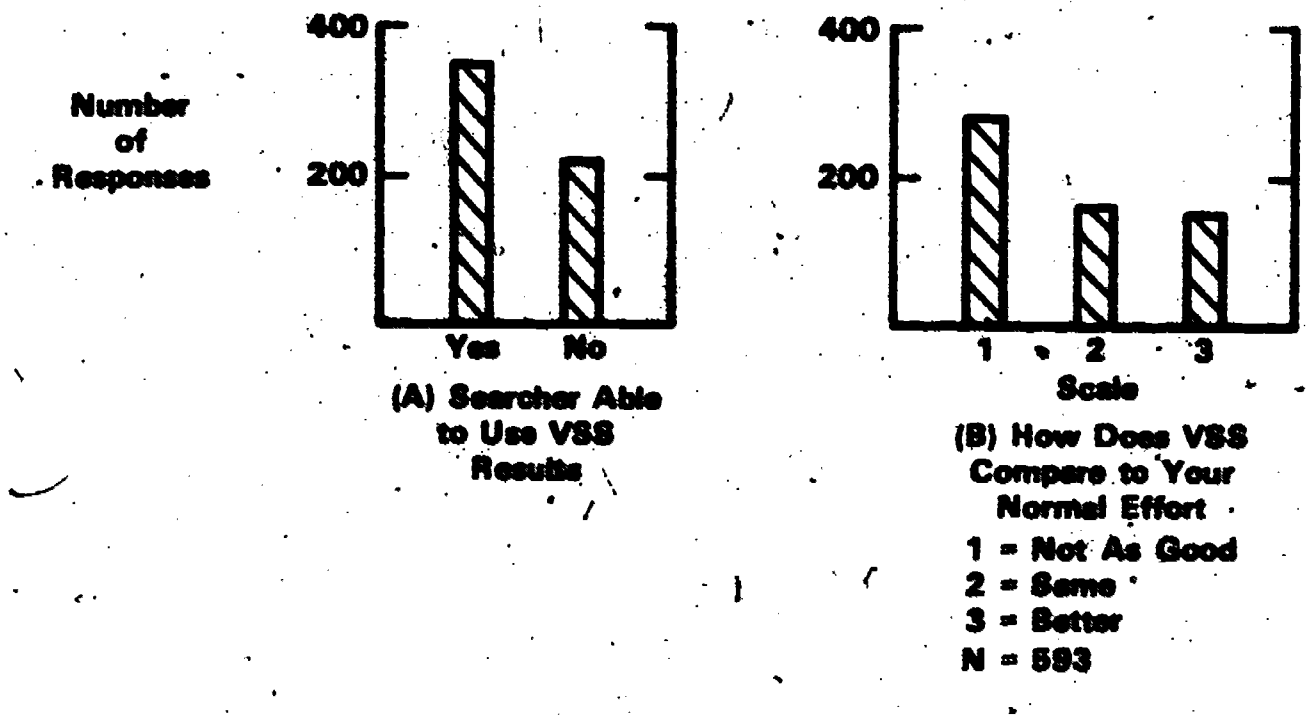
| Switching Strategies* | Number of Usable VSS Terms | | | |
|---|---|---|---|---|
| | 0 | 1 - 5 | 6 - 10 | > 10 |
| 1 synonyms | 20** | 22 | 2 | 0 |
| 2 browse | 69 | 103 | 26 | 13 |
| 3 broader | 8 | 1 | 0 | 1 |
| 4 narrower | 1 | 0 | 0 | 0 |
| 5 BT/NT | 14 | 14 | 6 | 1 |
| 6 user-defined | 2 | 19 | 0 | 5 |
| 7 multiple | 26 | 65 | 30 | 22 |

*Refer to Appendix C for definition of each strategy.
**VSS searches

TABLE 28 VSS RATINGS BY NUMBER OF OUTPUT TERMS BY MODULE

| VSS Module | Number of Usable VSS Terms | | |
|---|---|---|---|
| | 1 - 5 | 6 - 10 | > 10 |
| Physical Science | 2.9 | 3.7 | 4.2 |
| Social Science | 2.6 | 3.9 | 4.1 |
| Business | 2.5* | 3.8 | 4.0 |
| Life Science | 2.7 | 3.1 | 4.0 |
| All Modules | 2.7 | 3.6 | 4.1 |

* VSS rating on a 5-point scale where 1 means VSS was of no help, and 5 means VSS was very helpful.



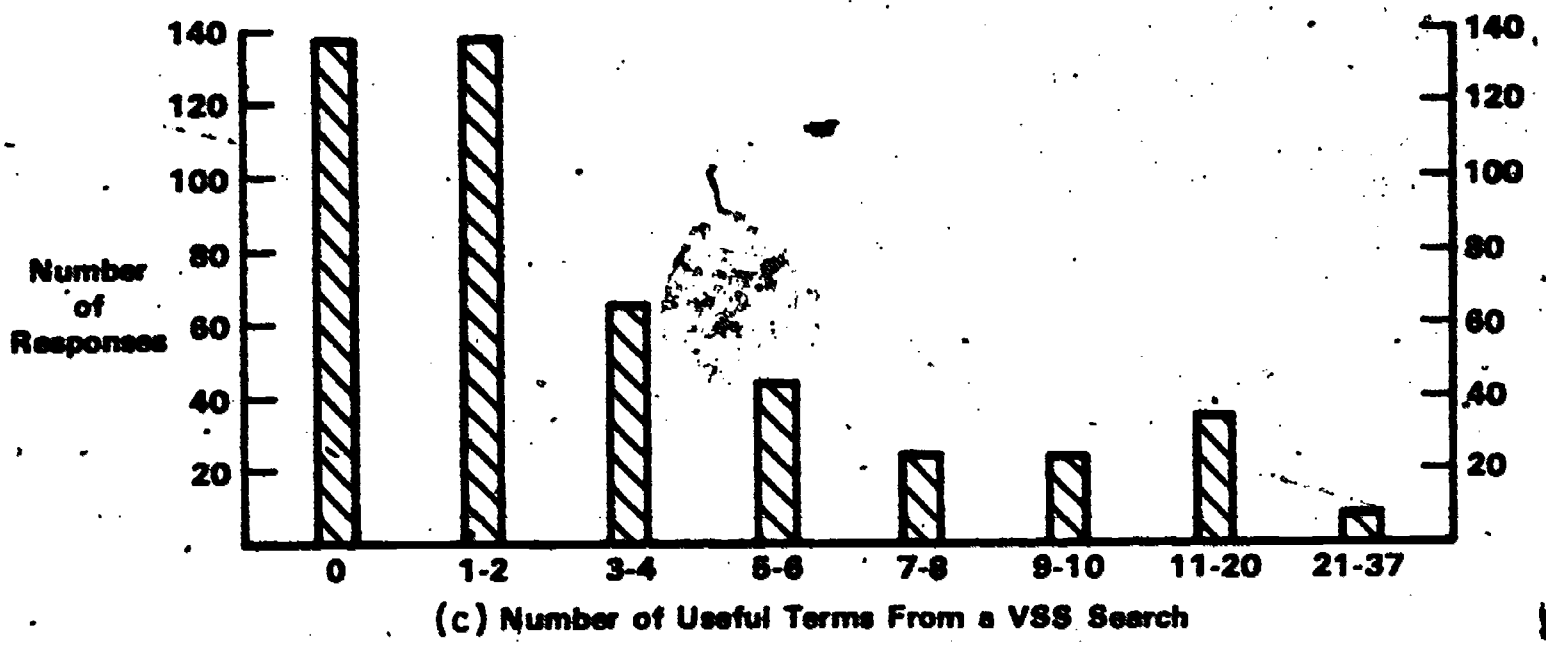




FIGURE 4 VSS PERFORMANCE AS MEASURED BY USEFUL OUTPUT

An important performance measure was the user's subjective rating of VSS, particularly for those searches where usable VSS output was indicated. Table 28 shows, as would be expected, that VSS ratings increase with increasing amounts of usable output. Quite respectable ratings are achieved when usable output reaches 6-10 terms per search request. The overall ratings for VSS were: 3.08 (307 usable searches); 2.35 (593 total searches).

### 9.2.3 Correlations Between VSS Performance and User Proficiency or User Individuality

While education and experience tend to be indicators of one's evaluation of VSS, other factors were of interest. An analysis was performed to identify the importance of other factors, including:

- Participant differences; e.g. differences due to overall experience and education level
- Participant's proficiency with the particular modules used in the search
- Module choice for a particular search

The computer program BMDP3V (Dixon, 1981) was used to perform mixed-model analysis of covariance on the data collected on individual searches. Each participant was asked to rate VSS's performance for each search attempted. A five-point rating scale was used with 1 indicating VSS was of no help and 5 indicating VSS was very helpful. The analysis expressed the rating as a function of the factors previously mentioned.

The following statistical model was used in the analysis:

$$R_k \; R_{k(ij)} = \mu + M_i + P_j + S_{k(ij)} + \beta \cdot PROF_{ij} \qquad (1)$$

where $R_{k(ij)}$ denotes the rating given to VSS's performance on the kth search performed by Participant j with the ith Module.

$M_i$ denotes the effect due to the use of a particular module.

$P_j$ denotes the effect due to the unique qualities of the participant (e.g. general experience, age, and education level).

$S_{k(ij)}$ denotes the non-systematic (random) effects due to the unique characteristics of the particular search.

$PROF_{ij}$ denotes the effect of the participant's proficiency on the particular module used.

$\mu$, $\beta$ constants to be determined by the statistical analysis

This statistical model assumes that the VSS rating for each search is systematically related to several factors, while also being randomly affected by several others. For example, the effects of vocabulary module and participant proficiency are assumed to systematically affect the rating. The model assumes that the average rating assigned to Business module searches (in general) may be different from Social Sciences, Life Science, or Physical Sciences searches. The model also assumes that VSS ratings change at a linear rate with participant proficiency with the particular module used in the search.

The effects due to a participant's or a search's unique characteristics are assumed to be random. That is, the participants and searches are assumed to be randomly selected from a larger conceptual population of possible participants and searches. Thus, the characteristics of a particular search or participant that cause a rating to be high or low cannot be systematically predicted. Rather, their effect is simply to increase the uncertainty in predicting a rating.

The statistical model provides an approach for estimating the magnitude of the effects due to the different factors. The results indicated that all the factors were statistically significant. However, the estimated effects of the factors differed.

The average rating for all VSS searches was 2.25. The effect of a participant's module choice on the VSS rating is shown in Table 29. This table presents the estimated average rating for each vocabulary module. The table indicates that the estimated average rating (over all possible participants and searches) for a Physical Science module search is 2.45. However, this average increases to 3.03 when only the usable searches are considered.

Thus, in general, VSS's performance was best for physical-science-type searches while business and lif-science-type searches received the lowest ratings. However, module choice was not a major factor influencing the outcome of VSS performance ratings.

A participant's self-assessed proficiency (with a given module) was found to be inversely related with VSS ratings. The more proficient the participant (in the use of a particular module) the lower the VSS ratings tended to be. Participants who claimed to be experts on a given module are estimated, on the average, to give ratings 0.65 points lower (on a five point

scale) than those participants claiming to have no proficiency. Specifically, participants who rated themselves "experts" for a particular module rated VSS 1.97 on the average, while those who claimed to have no proficiency rated VSS 2.62 on the average.

One of the most important factors was differences among individual participants. In other words, how is the VSS performance rating affected by the uniqueness of an individual? The estimated variability (standard deviation) in ratings due to differences among the participants was 0.63. Thus, a rating is estimated to vary by as much as ± 1.25 due to differences in

**TABLE 29 ESTIMATED AVERAGE RATINGS FOR VSS SEARCHES BY MODULE**

| VSS Module | User Ratings* | |
|---|---|---|
| | All Searches | Usable Searches |
| Physical Science | 2.45 | 3.03 |
| Social Science | 2.38 | 2.90 |
| Life Science | 2.12 | 2.82 |
| Business | 2.05 | 2.89 |

*5-Point scale

participants. In other words, if the average rating (over all hypothetical participants) is 2.25, a randomly selected participant could give a rating as low as 1.00, while another could give a rating as high as 3.50.

These factors did not explain all the variability in ratings; a considerable amount of rating variation was unexplained. In other words, a considerable amount of variation was due to uncontrolled factors that differed from search to search (e.g. difficulty of search). The size of this variation could cause the ratings to vary by as much as ± 2.00 points on the five point scale. In summary, the relative importance of the factors analyzed (in decreasing order) are: uncontrolled search-to-search differences, differences among participants, differences in a participant's proficiency, and differences in the modules.

### 9.2.4 Summary of Open-ended Questions

There were four open-ended questions on the VSS evaluation form (questions 14 through 17). Appendix G lists the verbatim responses of each participant to each question.

On the question of whether VSS would make a searcher's job easier or not (Question 14), participants who responded "easier" (about 40 percent) felt that it would save time and money because they would not have to consult printed thesauri as much, or at all. Also, they seemed to like the idea of juxtaposition of terms from various thesauri and they felt that they would have a better idea of what terms to use prior to searching any data base.

Participants who responded "no difference" to question 14 (about 51 percent) felt that VSS was time consuming, of little help with free text searches, not needed since BRS's CROS and Dialog's File 411 were sufficient aids, not useful enough because of too few vocabularies, and not needed if a searcher conducts a good presearch interview, and knows his vocabularies.

Those who responded "harder" to question 14 (about 9 percent) felt VSS required a substantial amount of time, would add to the cost of the search and would be another system to learn. Some felt that VSS needed more of the intelligence available in printed thesauri (e.g., scope notes), and/or felt the system needed more work.

When asked about their overall reaction to VSS (Question 15) opinions varied widely from "very favorable" to "not at all impressed".

The three most widely held opinions were:

(1) VSS was very valuable to valuable (about 20 percent)

(2) VSS was useful or interesting but needs more work (about 25 percent)

(3) VSS was cumbersome, tedious or frustrating (about 20 percent)

Other opinions included indifferent or uncommitted, increasingly negative, not too useful, and not worth it (needs work).

The principal underlying cause of user frustration was the decision to provide a menu-driven approach to VSS with no option to provide a more

direct access to the system. This was an understandable reaction in view of the rather surprising fact that many participants used a 300 baud terminal. Menu-driven systems are not well suited to slower terminals. Also, participants had an average of about six years experience with on-line systems and were comfortable with direct-search approaches.

Several causes appear to be related to the user perception that the system needs more work; (a) vocabularies must be the latest available and kept up-to-date in such a system, (b) vocabularies themselves should have more hierarchical and/or synonym relationships to improve performance, and (c) VSS itself needs new menus, direct search, and improved switching algorithms.

On the question of whether subject switching, in general, is a good idea (Question 16), participants responded very positively.

Open comments (Question 17) tended to reiterate replies to earlier questions, however some were new. One person thought there was too much noise in the output, while another thought the stemming algorithm went too far. Several other points were made: (1) thesauri do not stand-alone very well, (2) title and abstract searches with words, phrases and stems work well if one has a creative mind, (3) VSS should have had MeSH tree structures for BT/NT terms (note: they were not supplied to Battelle), (4) scope notes were not working, and (5) VSS is only as good as the vocabularies that went into it.

## 10.0 ON-LINE USERS SURVEY UPDATE

In 1979, as part of a previous NSF grant on VSS (10), a national survey was conducted to determine searching patterns and preferences of on-line users and to elicit their reaction to VSS. There were several key questions in the survey related to this research over and above those questions pertaining to VSS. One dealt with the frequency of searching for information using multiple data bases. Another probed for the reason why searches were conducted against single data bases. Finally, several questions addressed the issue of searching with controlled, uncontrolled, and a combination of controlled and uncontrolled terms.

It was felt that by rerunning the same survey at this time (four years after the first one) a unique opportunity was available to observe any shifts in user search patterns and/or preferrences with time. Very few, if any, surveys in this field are repeated except for salary surveys. A benchmark question for the 1979 and 1983 surveys was the monthly dollar expenditures by individuals and organizations for on-line searches. It has been widely reported and generally accepted that the on-line business is growing at the rate of about 30 percent compounded annually. If our data confirmed this observation, then any other observable shifts in on-line search patterns would have added credibility.

### 10.1 Methodology

The survey was repeated by reissuing the 1979 questionnaire (unmodified). To minimize distribution costs, the survey was distributed as the 1983 National On-line Meeting; held in New York. (See Appendix H for sample questionnaire.) Meeting organizers granted permission for a passive distribution of the survey near the registration desk. Meeting registrants simply helped themselves to a blank survey questionnaire at the time of registration or any other time during the meeting. About 325 questionnaires were distributed in this manner; 38 valid questionnaires were returned, for a 12 percent rate of return. Consequently, sample sizes for 1979 and 1983 surveys varied considerably, 755 versus 38, respectively.

All participation was voluntary and responses were kept confidential. Data were analyzed via simple descriptive statistics. Finally, the 1979 survey data were reanalyzed subsequent to our 1980 report (10) because 66 additional returns were received after the publication cutoff date. This report reflects all the 1979 data, or 755 valid returns.

## 10.2 Results

Benchmarks for comparing the two surveys were the questions about individual and corporate expenditures for on-line search activity. The data for on-line search expenditures (by individuals) confirm various open literature sources that estimate that on-line searching is growing at the rate of 30 percent compounded annually (Table 30c). However, the data also suggest that corporate search activity is growing at an even greater rate, 45 percent compounded annually (Table 31c). Of these two benchmarks, individual search activity is probably the more reliable because individuals generally are in a better position to estimate their own search activity than that of their overall organization. Some respondents admitted their uncertainty about aggregrate corporate search activity; others simply omitted a response to the corporate questions. In any case, there is reason to believe that 1983 survey results, albeit a small sample, have some validity since they closely agree with reported literature regarding growth.

Connect-hour growth for individuals increased at a lower rate, 18 percent compounded annually (Table 30b) and the number of searches per month increased only slightly over the four year period (Table 30a). It can be observed that on-line searches averaged 0.25 and 0.44 hrs/search for 1979 and 1983, respectively (Tables 30 and 31). This observation suggests that one of the factors contributing to on-line growth is lengthier search sessions. Other factors, of course, are price increases and the number of new customers.

Figures 5 and 6 show the four year trends in search activity. There is a noticeable shift, left to right, in number of searches, connect hours, and dollars expended per month in these two figures, confirming the growth dynamics reported in the literature about this business.

# TABLE 30 INDIVIDUAL SEARCH ACTIVITY

(a)

| No. of Searches Per Month | 1979 N | 1979 % | 1983 N | 1983 % |
|---|---|---|---|---|
| 1-10 | 228 | 38 | 5 | 15 |
| 11-20 | 145 | 24 | 11 | 32 |
| 21-30 | 81 | 14 | 8 | 24 |
| 31-40 | 43 | 7 | 2 | 6 |
| 41-50 | 35 | 6 | 3 | 9 |
| 51-100 | 37 | 6 | 3 | 9 |
| 101-200 | 16 | 3 | 2 | 5 |
| 201-500 | 8 | 2 | 0 | 0 |
| >500 | 2 | 0 | 0 | 0 |
| Total | 595 | 100 | 34 | 100 |
| Total Searches | 17,790 | | 1196 | |
| Average Searches/Mo. | 29.9 | | 35.2 | |

(b)

| Connect Hours Per Month | 1979 N | 1979 % | 1983 N | 1983 % |
|---|---|---|---|---|
| 1-5 | 298 | 59 | 13 | 42 |
| 6-10 | 110 | 22 | 4 | 13 |
| 11-15 | 40 | 8 | 3 | 10 |
| 16-20 | 22 | 4 | 4 | 13 |
| 21-30 | 13 | 3 | 0 | 0 |
| 31-50 | 13 | 3 | 5 | 16 |
| 51-100 | 2 | 0 | 2 | 6 |
| >100 | 4 | 1 | 0 | 0 |
| | 502 | 100 | 31 | 100 |
| Total Hours | 4468 | | 533 | |
| Average Hours/Mo. | 8.9 | | 17.2 | |

(c)

| On-line Expenditures Per Month (in dollars) | 1979 N | 1979 % | 1983 N | 1983 % |
|---|---|---|---|---|
| $ 1-50 | 65 | 14 | 0 | 0 |
| 51-100 | 102 | 21 | 2 | 6 |
| 101-200 | 103 | 21 | 3 | 9 |
| 201-300 | 63 | 13 | 5 | 16 |
| 301-400 | 34 | 7 | 3 | 9 |
| 401-500 | 30 | 6 | 3 | 9 |
| 501-1000 | 46 | 10 | 9 | 28 |
| 1001-2000 | 29 | 6 | 3 | 9 |
| 2001-3500 | 7 | 2 | 2 | 7 |
| >3500 | 1 | 0 | 2 | 7 |
| Total | 480 | | 32 | 100 |
| Total Expenditures | 161,760 | | 31,105 | |
| Average Expenditures/Mo. | 337 | | 972 | |

## TABLE 31 CORPORATE SEARCH ACTIVITY

(a)

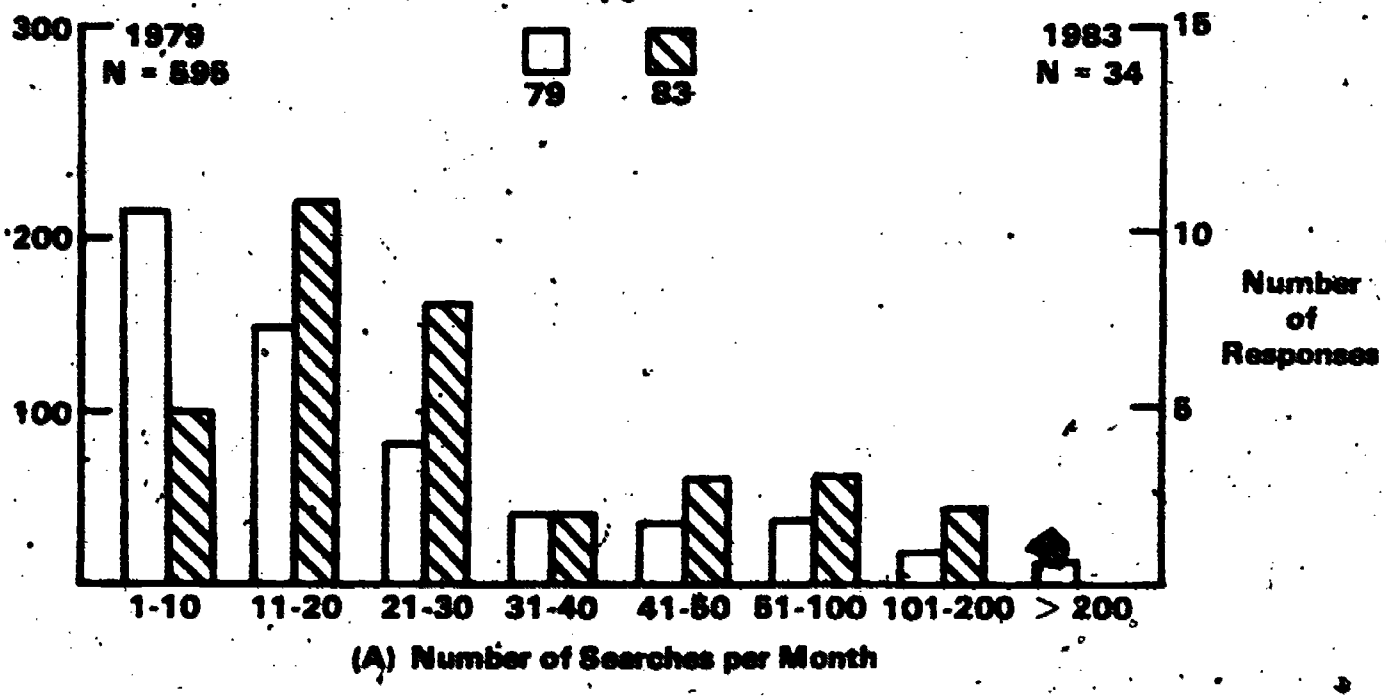
| No. of Searches Per Month | 1979 N | 1979 % | 1983 N | 1983 % |
|---|---|---|---|---|
| 1-10 | 93 | 20 | 2 | 8 |
| 11-20 | 73 | 15 | 2 | 8 |
| 21-30 | 55 | 12 | 3 | 12 |
| 31-40 | 38 | 8 | 3 | 12 |
| 41-50 | 38 | 8 | 3 | 12 |
| 51-100 | 82 | 17 | 6 | 24 |
| 101-200 | 54 | 11 | 1 | 4 |
| 201-500 | 36 | 8 | 3 | 12 |
| >500 | 4 | 1 | 2 | 8 |
| Total | 473 | 100 | 25 | 100 |
| Total Searches | 35,097 | | 5,682 | |
| Average Searches/Mo. | 74.2 | | 227.3 | |

(b)

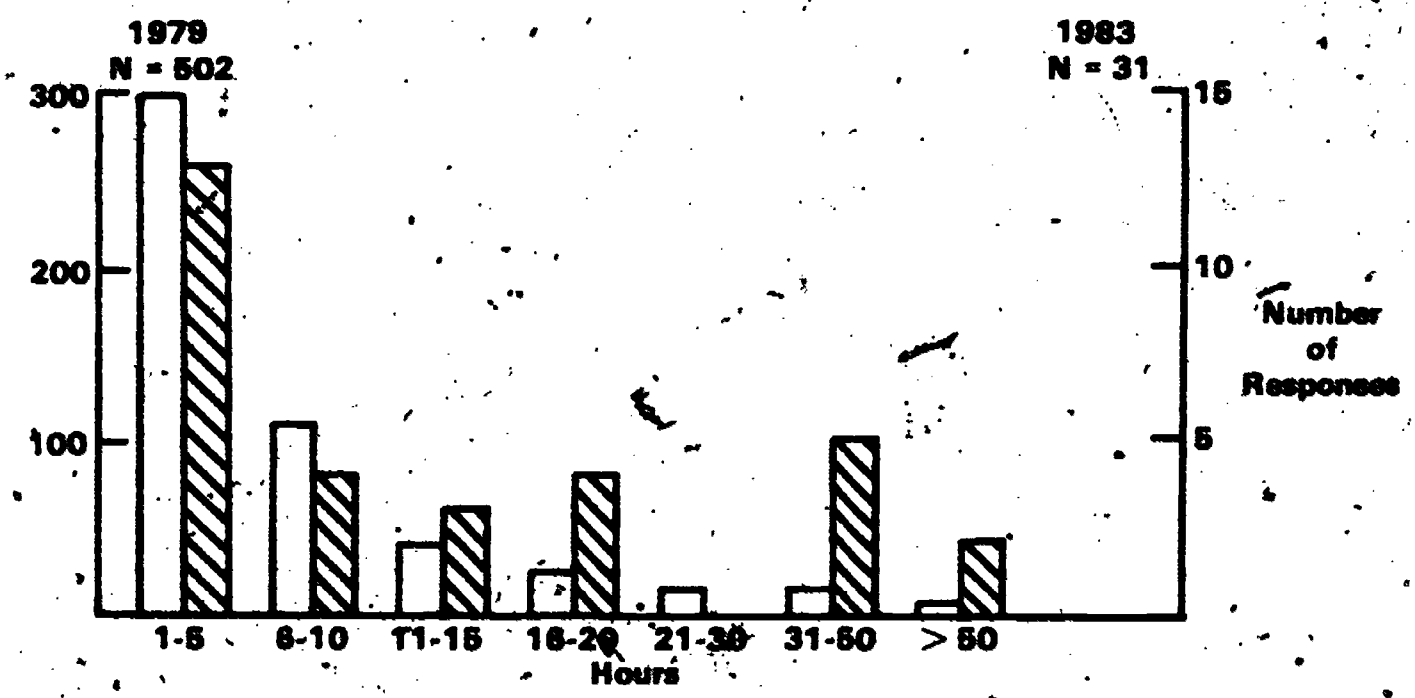
| Connect Hours Per Month | 1979 N | 1979 % | 1983 N | 1983 % |
|---|---|---|---|---|
| 1-5 | 137 | | 3 | |
| 6-10 | 82 | | 1 | |
| 11-15 | 38 | | 3 | |
| 16-20 | 31 | | 4 | |
| 21-30 | 41 | | 2 | |
| 31-50 | 44 | | 3 | |
| 51-100 | 19 | | 4 | |
| >100 | 10 | | 3 | |
| | 402 | | 23 | |
| Total Hours | 8,241 | | 1,483 | |
| Average Hours/Mo. | 20.5 | | 64.5 | |

(c)

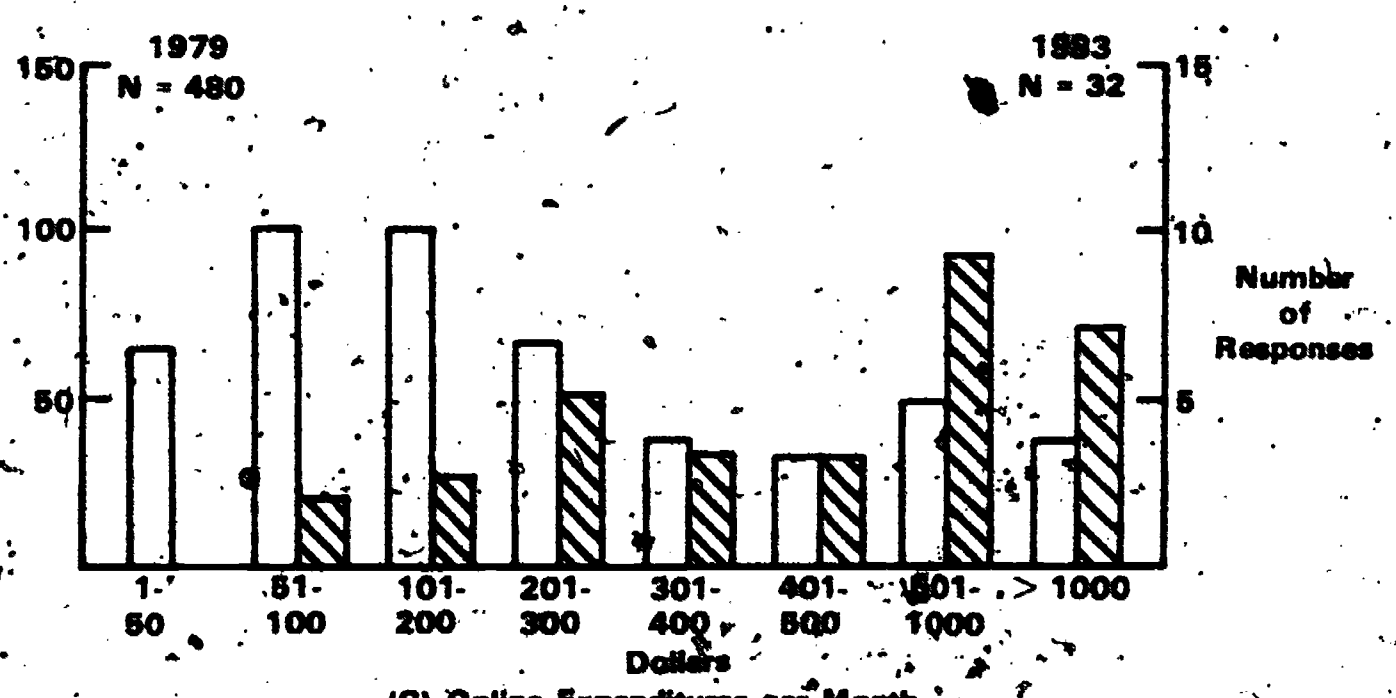
| On-line Expenditures Per Month (in dollars) | 1979 N | 1979 % | 1983 N | 1983 % |
|---|---|---|---|---|
| $ 1-50 | 27 | 7 | 0 | 0 |
| 51-100 | 40 | 10 | 0 | 0 |
| 101-200 | 48 | 12 | 0 | 0 |
| 201-300 | 56 | 14 | 3 | 12 |
| 301-400 | 39 | 10 | 1 | 4 |
| 401-500 | 33 | 8 | 2 | 8 |
| 501-1000 | 66 | 17 | 3 | 12 |
| 1001-2000 | 54 | 14 | 8 | 32 |
| 2001-3500 | 12 | 3 | 3 | 12 |
| 3501-10,000 | 14 | 4 | 2 | 8 |
| >10,000 | 3 | 1 | 3 | 12 |
| Total | 392 | 100 | 25 | 100 |
| Total Expenditures | 391,608 | | 102,940 | |
| Average Expenditures/Mo. | 999 | | 4,118 | |




(A) Number of Searches per Month

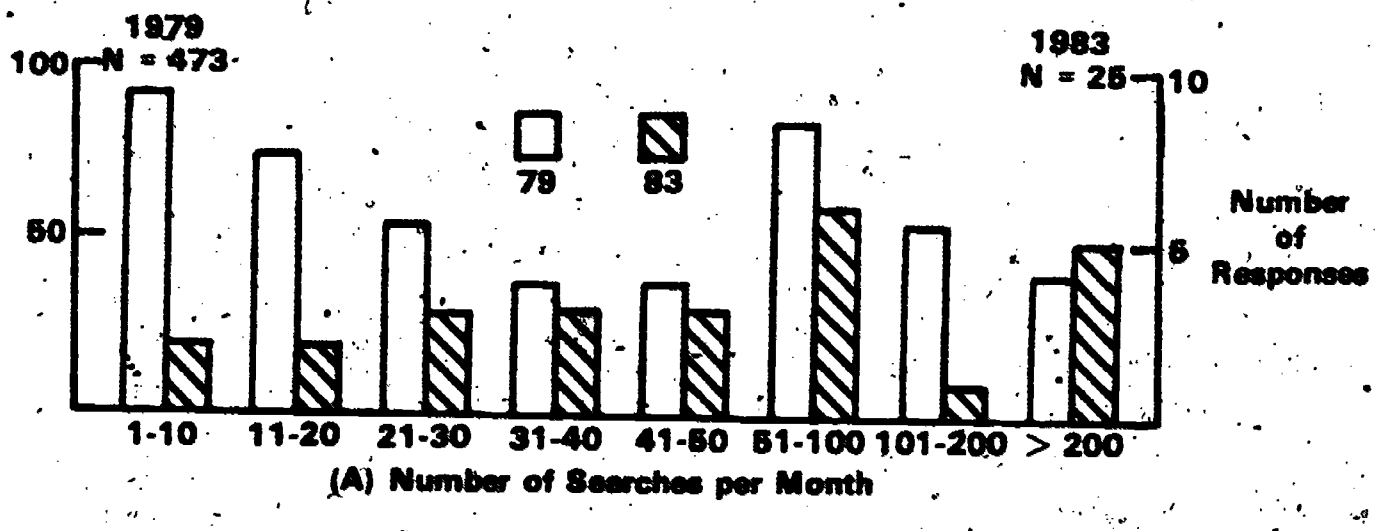



(B) Online Connect Hours per Month

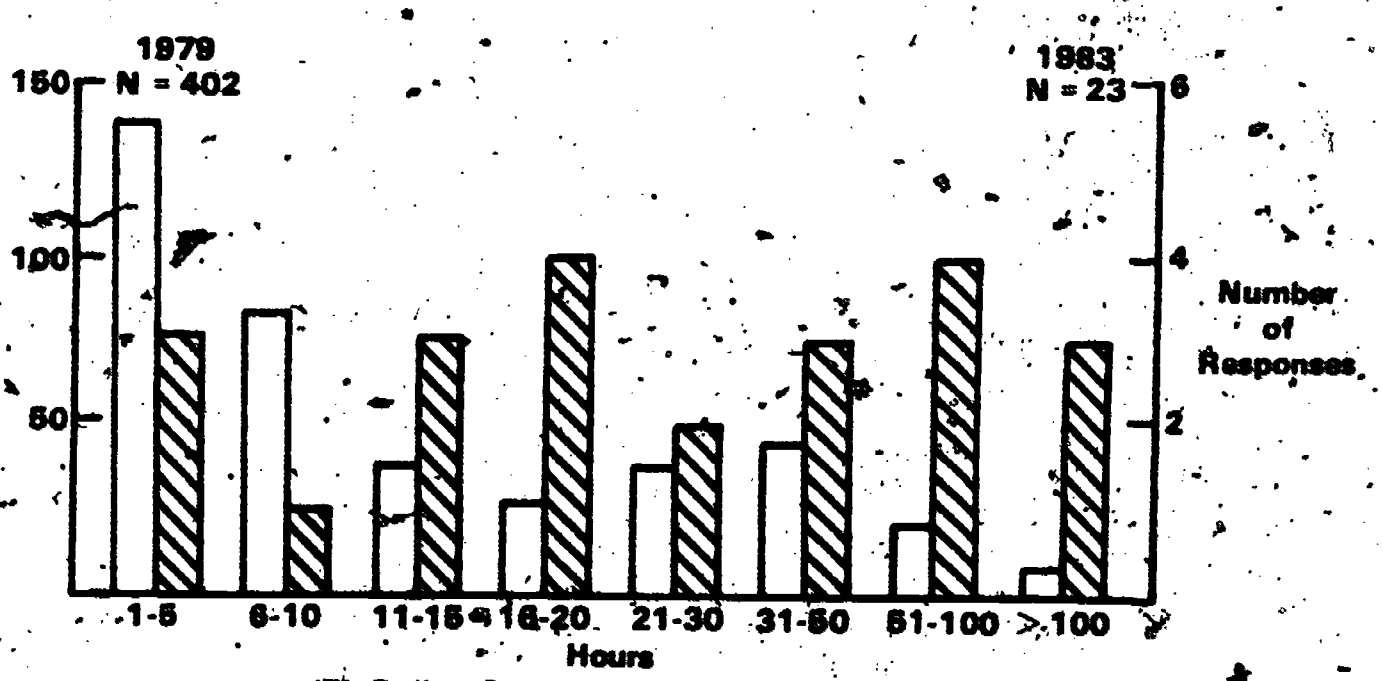



(C) Online Expenditures per Month

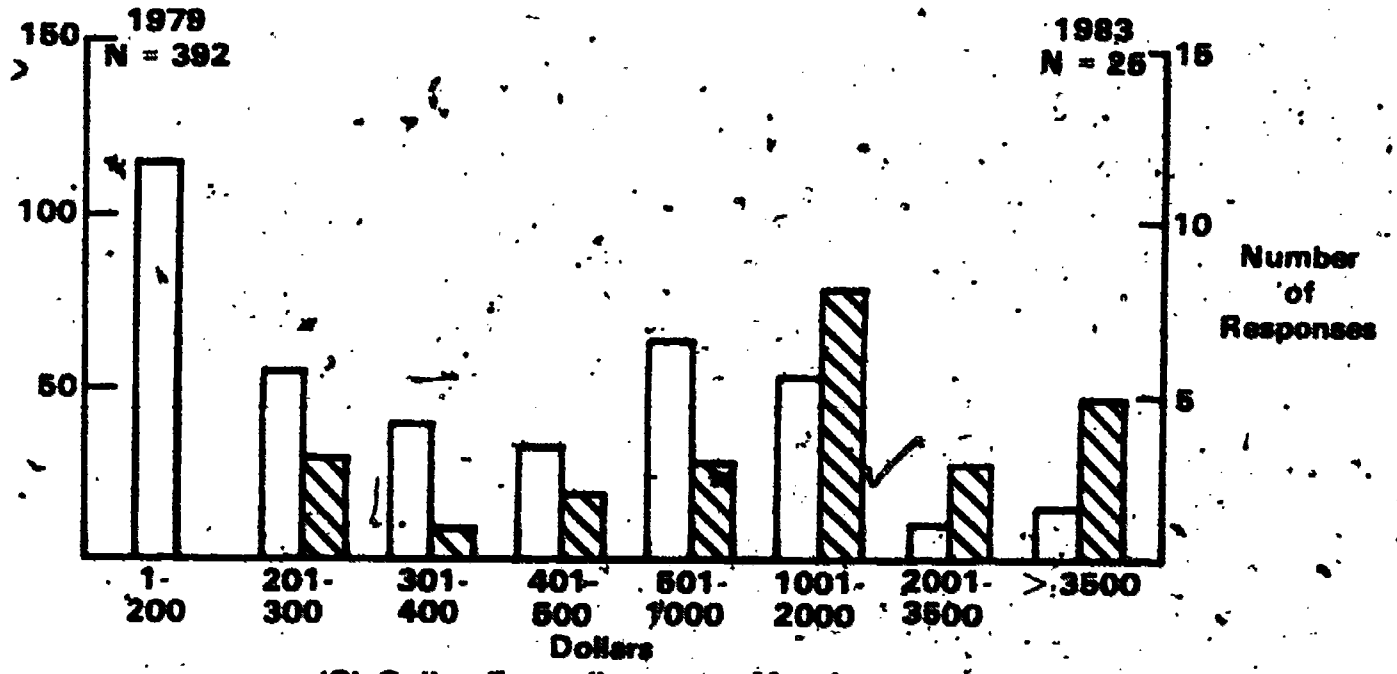
FIGURE 5 INDIVIDUAL ON-LINE SEARCH ACTIVITY




(A) Number of Searches per Month




(B) Online Connect Hours per Month




(C) Online Expenditures per Month

FIGURE 6 CORPORATE ON-LINE SEARCH ACTIVITY

Another aspect about the online business is search activity within various sectors of the user community. Table 32 shows 1979 search activity by four major user sectors. This breakdown was not previously reported in our 1980 report but was produced during our re-analysis of all 1979 data. However, 1983 data were not analyzed in this fashion due to the small initial sample size.

It can be seen that government sector users did more searching in 1979 both at individual and organizational levels than any other sector. However, the for-profit sector user spent more money for searching, on the average, than any of the other sectors.

At the individual level, both government and university sectors spent about the same amount on an "average" search ( $7.75), while for-profit and non-profit sectors spent about $17.89 and $10.82, respectively, for the average search. At the organizational level, university users spent about $8.83 per search, government and non-profit users spent about $11.40, and for-profit users spent about $27.52 per search.

When asked who was paying for on-line searches (Question 6 in the survey), a bimodal distribution resulted (Table 33). This was expected, because if a respondent replied that end users were paying for searches 100 percent of the time, then, of course libraries paid 0 percent of the time.

If blanks are counted as 0 percent, which was virtually always the case on the individual questionnaires, then Table 33 shows that end users are paying for a greater percentage of the searches in 1983 than they were in 1979 and, conversely, libraries are paying for far fewer searches in 1983 than in 1979.

Restrictions which an end user might place on an on-line search, in descending order were: (1) limit the search by date, number of citations, etc; (2) limit the search to some certain data base(s); (3) limit the search by some cost ceiling; and (4) limit the search by some amount of connect time (Table 34). Data for 1983 tend to reverse the order of (2) and (3).

On the question of which on-line retrieval services people are using (Question 8), there has been no change in the ranking of major services over time (Table 35).

TABLE 32 ON-LINE SEARCH ACTIVITY AND EXPENDITURES BY USER SECTOR (1979)

| USER SECTOR | Searches per Month | | | Expenditures per Month (in dollars) | | |
|---|---|---|---|---|---|---|
| | N | mean | median | N | mean | median |
| **INDIVIDUAL SEARCH ACTIVITY** | | | | | | |
| For Profit | 235 | 26.1 | 15.0 | 199 | 467 | 275 |
| Government | 66 | 51.8 | 20.0 | 52 | 398 | 245 |
| Non Profit | 41 | 25.5 | 20.0 | 35 | 276 | 180 |
| Universities | 250 | 28.2 | 14.5 | 175 | 222 | 115 |
| **CORPORATE SEARCH ACTIVITY** | | | | | | |
| Government | 61 | 112.2 | 56.0 | 53 | 1261 | 680 |
| For Profit | 173 | 57.7 | 25.0 | 153 | 1588 | 400 |
| Universities | 211 | 90.2 | 49.0 | 157 | 797 | 390 |
| Non Profit | 31 | 45.8 | 27.5 | 27 | 532 | 207 |

TABLE 33 WHO PAYS FOR ON-LINE SEARCHES

| Percentage of Searches Paid for by A, B or C | On-line Searches Are Paid By | | | | | |
|---|---|---|---|---|---|---|
| | (A) End Users | | (B) Library | | (C) Both | |
| | 79 | 83 | 79 | 83 | 79 | 83 |
| 0-20 | 85* | 4 | 161 | 10 | 32 | 0 |
| 21-40 | 28 | 2 | 38 | 4 | 5 | 0 |
| 41-60 | 26 | 2 | 15 | 1 | 2 | 1 |
| 61-80 | 64 | 3 | 22 | 1 | 1 | 0 |
| 81-100 | 243 | 18 | 262 | 8 | 54 | 4 |
| Blanks | 309 | 9 | 257 | 14 | 661 | 33 |
| 0-60 | 60** | 45 | 38 | 76 | | |
| 61-100 | 40 | 55 | 62 | 24 | | |

* Number of respondents
**Percent of respondents falling into this range for this year.

TABLE 34 RESTRICTIONS PLACED ON ON-LINE SEARCHES

| Percentage of Searches Limited by A, B, C or D | Limit Search by: (A) Date, Quantity, 79 | 83 | (B) Cost Ceiling 79 | 83 | (C) Data Base 79 | 83 | (D) Connect Time 79 | 83 |
|---|---|---|---|---|---|---|---|---|
| 0-20 | 236* | 7 | 286 | 8 | 267 | 12 | 296 | 2 |
| 21-40 | 62 | 4 | 39 | 6 | 49 | 2 | 5 | 0 |
| 41-60 | 103 | 6 | 50 | 3 | 45 | 3 | 14 | 0 |
| 61-80 | 87 | 6 | 41 | 2 | 44 | 2 | 6 | 0 |
| 81-100 | 153 | 9 | 53 | 5 | 121 | 3 | 10 | 1 |
| Blanks | 114 | 6 | 286 | 14 | 228 | 16 | 424 | 33 |
| mean % | 48 | 57 | 28 | 45 | 39 | 67 | 8 | 35 |

*Number of respondents

TABLE 35 ONLINE RETRIEVAL SYSTEM USAGE

| | 79 | 83 | Change |
|---|---|---|---|
| Dialog | 1.27* | 1.21 | -0.06 |
| Elhill | 1.95 | 1.95 | NC |
| BRS | 2.03 | 2.23 | +0.20 |
| Orbit | 2.10 | 2.23 | +0.13 |
| Recon | 2.43 | 2.53 | +0.10 |
| NY Times | 2.49 | 2.56 | +0.07 |
| DDC | 2.72 | 2.87 | +0.15 |

*the mean value of all responses where: 1 = frequently used; 2 = sometimes used; 3 = infrequently used

Individually, Dialog apparently was being used somewhat more frequently in 1983 than in 1979, while for Elhill (NLM) and all others there was either no change or somewhat less usage in 1983 compared to 1979. What is not reflected is the usage of new services since 1979, which are many. Also,

the respondent populations varied between the 1979 and 1983 surveys which would tend to distort the results.

One of the primary questions of interest in this survey was the trend in single versus multiple data base searching patterns. A shift toward greater usage of multiple data bases represents a justifiable case for a search aid like VSS.

Figure 7 shows how multiple data base search patterns are shifting with time. Single data base searches are on the decline -- note the shift from right to left in Figure 7A. Searches involving two data bases show minor shifts in individual categories, but overall, appear to be unchanged (Figure 7B); 3-data-base searches appear to be increasing with time (Figure 7C); and multiple data base searches involving four or more files are definitely increasing (Figure 7D). In fact, respondents who answered this question for the 1983 survey frequently changed the choice on the questionnaire by adding the word "more" to the choice identifying four data bases. Overall, the trend is toward more multiple data base searches.

When asked under what conditions searches were limited to one data base (Question 10), the three most common reasons given in decreasing order were:

(1) Exhaustive search was not required

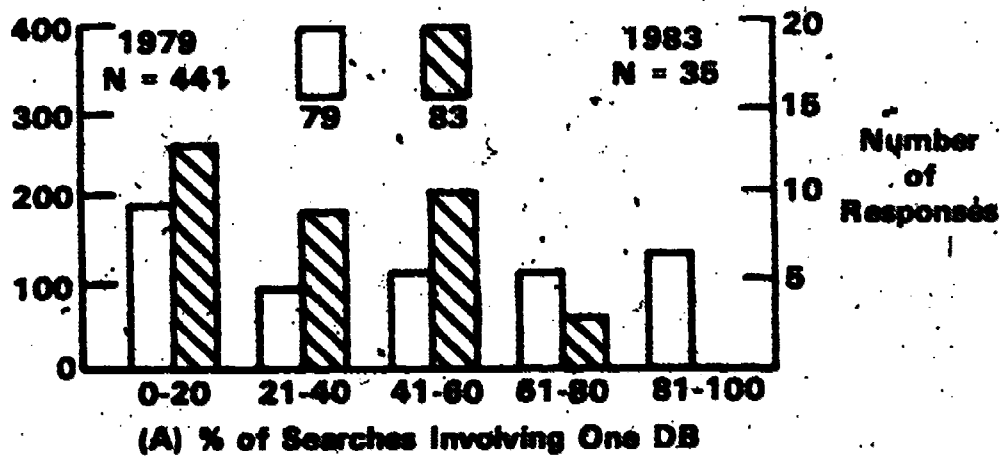
(2) Multi-base searches are too costly

(3) Multi-base searches are too time consuming

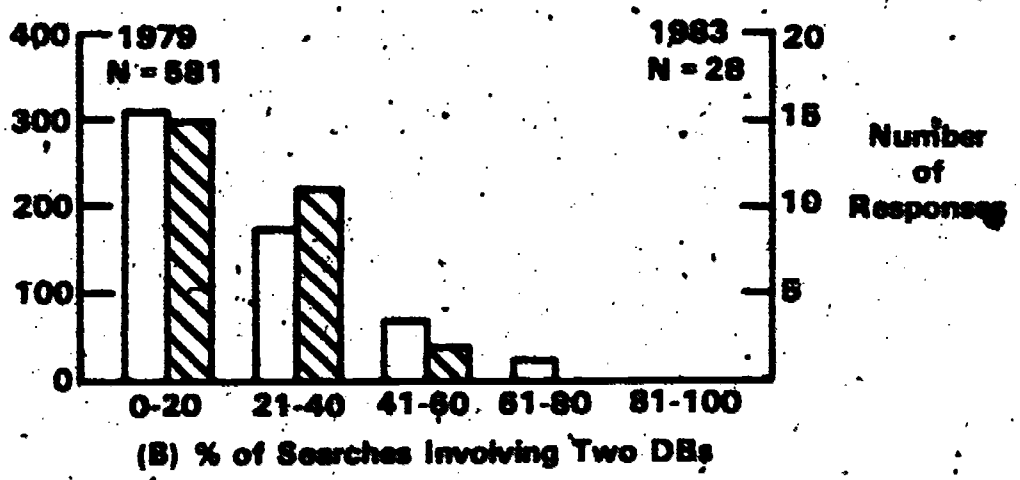
These results ignore "other" as a reason (Table 36).

The fact that some data bases might be too difficult or too unfamiliar (Column d, Table 36) ranked fourth and fifth in 1979 and 1983, respectively. In other words, users are not confining a search to one data base just because others might be unfamiliar or different to use. Overall, there was virtually no change in user response to Question 10 over time.

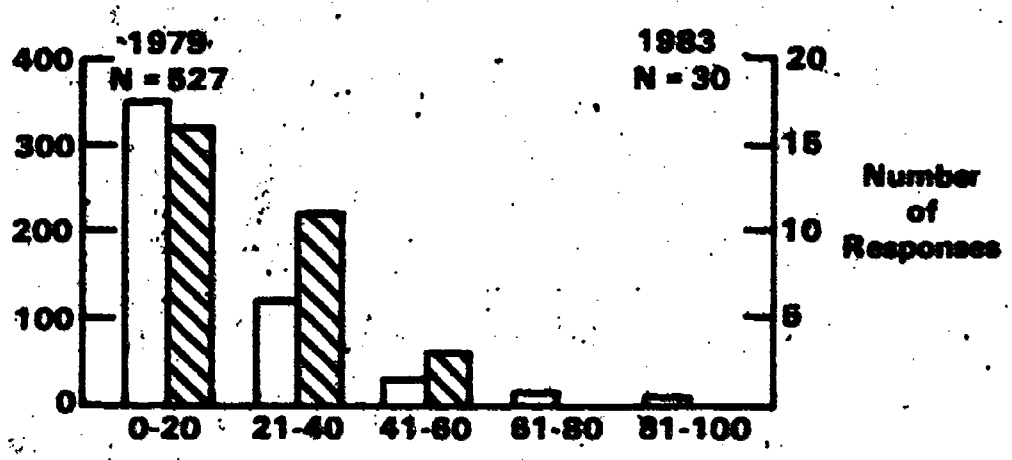
On the question of which subject areas would be most useful for multiple data base searching (Question 11), it is appropriate to examine the results (Table 37) both in terms of the number of respondents per subject area and the mean value of the responses because the usefulness of an area is a function of both factors.



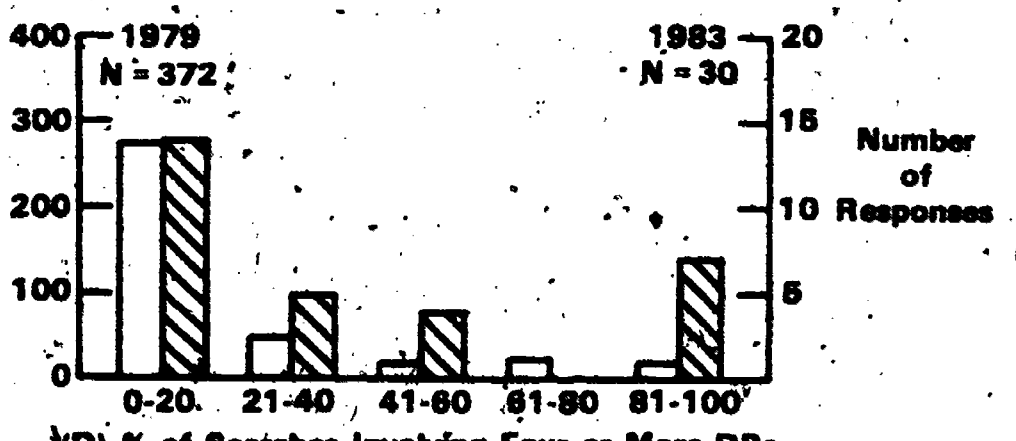

(A) % of Searches Involving One DB




(B) % of Searches Involving Two DBs




(C) % of Searches Involving Three DBs




(D) % of Searches Involving Four or More DBs

FIGURE 7 MULTIPLE DATA BASE SEARCHING PATTERNS

In 1979, respondents ranked the life sciences/medicine area highest among subject areas where multiple data base searching is (or would be) most useful. This area also received the greatest number of responses, 259. Other areas receiving over 200 responses are shown in Table 37 in decreasing order based on the mean.

In areas where the response was between 100 and 200, education received the highest ranking, while in areas receiving less than 100 responses, agriculture was ranked the highest. Mathematics was the least popular area (N=13) and the area least likely to be useful for multiple data base searches (mean=2.38).

The 1983 survey was too small to provide a useful comparison with 1979 data for Question 11; none-the-less, the distribution of 1983 data is given in Table 37. It can be seen that areas receiving heavy response in 1979 were also receiving the greater percentage of responses in '83.

Data for Question 12 were not compiled for the 1983 survey because there were too few respondents and too many possible data bases and combinations of data bases to warrant the effort. The 1979 data were reported earlier (10) for this question.

Questions 13 and 14 dealt with the three principal methods for conducting subject searches: controlled vocabulary, free text and a combination of both. On these two questions, we had the benefit of additional data for 1983 because the 65 participants in the field evaluation of VSS were also asked these same questions.

Figure 8 (A) shows a definite shift from right to left over time indicating the decreasing popularity of controlled vocabulary searching. Figure 8 (C) shows a very slight shift, right to left, indicating virtually no change in free text searching over time.

On the other hand, Figure 8 (B) shows a gradual shift from left to right indicating the increasing popularity of searching data bases with a combination of controlled vocabulary and free text terms.

Table 38 clearly shows that the users prefer a combination of controlled plus free text searching, followed by free text searching (exclusively) and controlled vocabulary searching (exclusively). Actually, the latter two methods exchanged places over time.

TABLE 36 - REASONS GIVEN FOR SEARCHING A SINGLE DATA BASE

| Rank Order | (a) 79 | 83 | (b) 79 | 83 | (c) 79 | 83 | (d) 79 | 83 | (e) 79 | 83 | (f) 79 | 83 | (g) 79 | 83 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1* | 329** | 32 | 81 | 2 | 5 | 0 | 4 | 0 | 25 | 0 | 1 | 0 | 50 | 1 |
| 2 | 94 | 3 | 217 | 14 | 60 | 6 | 41 | 1 | 64 | 3 | 9 | 1 | 21 | 0 |
| 3 | 16 | 0 | 65 | 6 | 105 | 9 | 54 | 1 | 46 | 4 | 27 | 0 | 7 | 0 |
| 4 | 1 | 0 | 23 | 2 | 42 | 2 | 86 | 6 | 23 | 2 | 34 | 5 | 1 | 0 |
| 5 | 4 | 0 | 16 | 1 | 29 | 2 | 46 | 4 | 16 | 2 | 64 | 7 | 1 | 0 |
| 6 | 0 | 0 | 3 | 1 | 12 | 0 | 8 | 5 | 81 | 6 | 47 | 6 | 1 | 0 |
| 7 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 5 | 1 | 4 | 0 | 2 | 2 |
| blank | 111 | 3 | 349 | 12 | 501 | 19 | 516 | 21 | 495 | 20 | 569 | 19 | 672 | 35 |
| mean | 1.2 | 1.1 | 2.2 | 2.6 | 3.3 | 3.0 | 3.6 | 4.6 | 3.8 | 4.4 | 4.7 | 4.9 | 1.7 | 5.0 |

Reasons

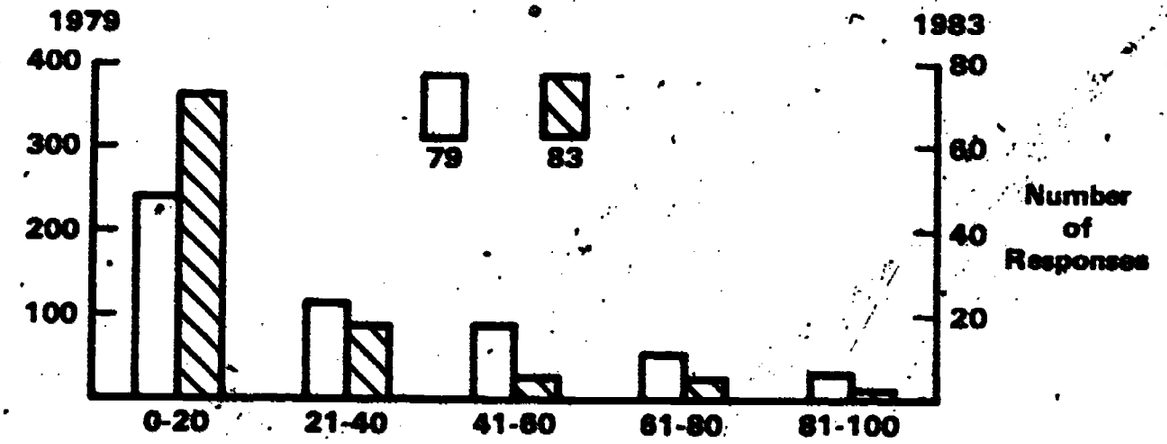
a. Query doesn't require more exhaustive search
b. Too costly to do multiple data base search
c. Too time consuming to do multiple data base search
d. Too difficult to use other, less familiar data bases
e. Other data bases not available to my organization
f. Too many data bases available to know which ones to use
g. Other (Please Specify)

*, user's rank order of reasons
**number of respondents

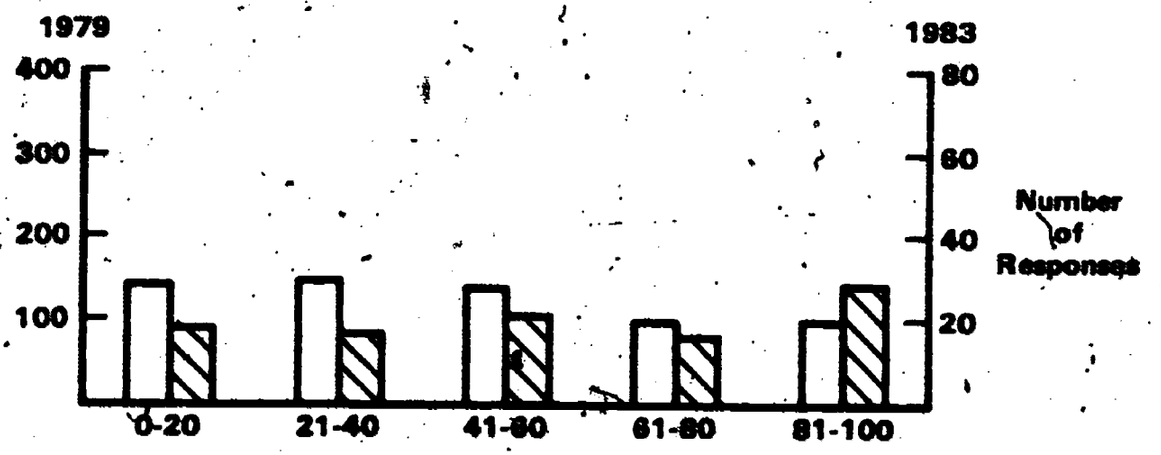
TABLE 37 SUBJECT AREAS MOST USEFUL FOR MULTIPLE DATA BASE SEARCHING

| Subject Area | 1979 | | | | | | 1983 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1* | 2 | 3 | N | Weight | Mean | 1 | 2 | 3 | N | Mean |
| Life Sciences/ Medicine | 135** | 84 | 40 | 259 | 423 | 1.63 | 7 | 5 | 6 | 18 | 1.94 |
| Engineering | 96 | 67 | 66 | 229 | 428 | 1.87 | 4 | 6 | 2 | 12 | 2.16 |
| Chemistry | 80 | 69 | 61 | 210 | 401 | 1.91 | 4 | 3 | 4 | 11 | 2.00 |
| Business/ Economics | 83 | 82 | 65 | 230 | 442 | 1.92 | 8 | 6 | 4 | 18 | 1.77 |
| Education | 56 | 29 | 45 | 130 | 249 | 1.92 | 1 | 1 | 3 | 5 | 2.40 |
| Psychology | 40 | 98 | 41 | 179 | 359 | 2.01 | 4 | 2 | 1 | 7 | 1.57 |
| Energy | 34 | 50 | 46 | 130 | 272 | 2.09 | 0 | 2 | 3 | 5 | 2.60 |
| Environment | 38 | 62 | 71 | 171 | 375 | 2.20 | 5 | 5 | 4 | 14 | 1.93 |
| Agriculture | 31 | 23 | 20 | 74 | 137 | 1.85 | 0 | 1 | 0 | 1 | 2.00 |
| Physics | 11 | 22 | 23 | 56 | 124 | 2.21 | 1 | 1 | 1 | 3 | 2.00 |
| Mathematics | 2 | 4 | 7 | 13 | 31 | 2.38 | 0 | 0 | 1 | 1 | 3.00 |

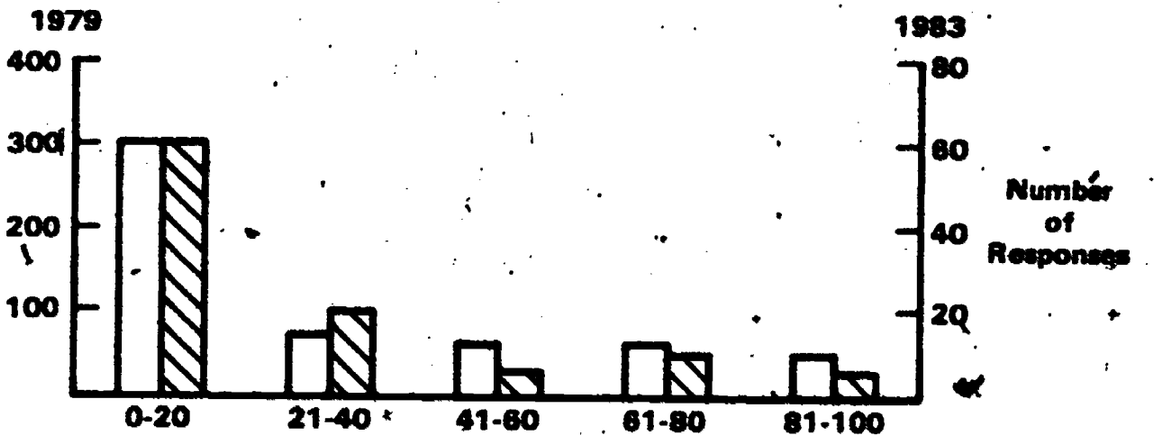
* respondents ranked top three choices
**number of respondents




(A) % of Searches Using Controlled Vocabularies Only




(B) % of Searches Using Controlled Plus Free Text




(C) % of Searches Using Free Text Only

FIGURE 8 USAGE OF VOCABULARIES

TABLE 38 USER PREFERENCES FOR TYPE OF SUBJECT SEARCHING DONE

| Subject Search Method | User Preferences* 1979 | 1983 |
|---|---|---|
| 1. Controlled plus free text | 1.31** (658) | 1.26 (101)*** |
| 2. Free text only | 2.38 (634) | 2.23 (100) |
| 3. Controlled only | 2.28 (632) | 2.49 (98) |

*Scale was 1 to 3 where: 1=most preferred; 3=least preferred
**arithmetic mean
***number of responses

When asked in what percentage of their present searches a user might turn to a subject switching system (Question 15), given that the only exposure a respondent had to such a system was a covering letter containing a simple example, the response was evenly divided in 1979, but is skewed toward the low end of the usage scale in 1983 (Table 39). The only clue as to why this response shifted over time may be due to the overall decline in user preference for controlled vocabulary searches (Table 38).

If there is a direct relationship between the low user preference for controlled-term searches and a low anticipated use of VSS, this suggests a possible misconception about how VSS might be used. The misconception is this: since VSS is essentially a data base of controlled vocabularies, it is only useful for controlled-term searches. This, of course, is not true. VSS output can, and probably should, be used to construct searches with controlled terms, free text terms, and controlled plus free text terms. For example, if a user is interested in information related to the term memory, in the psychological sense, and VSS produces "human information storage", "forgetting", "cognitive processes", "retention", etc., all of these controlled vocabulary terms can be

used as is for a controlled-term search, or they can be broken down into individual words and used as free text terms.

If "retention", "forgetting", etc, are valid terms and the user insists on searching in a free text mode, then VSS has served the same purpose as an abstract, title, or descriptor field, namely as a source of additional search terms. It is evident by some responses to VSS that some people cannot or will not make the transition that controlled vocabularies do not necessarily have to be used for controlled descriptor searches.

TABLE 39 ANTICIPATED USAGE OF VSS

| % VSS Usage | 79 | 83 |
|---|---|---|
| 0-20 | 164** | 13 |
| 21-40 | 138 | 7 |
| 41-60 | 162 | 7 |
| 61-80 | 123 | 6 |
| 81-100 | 106 | 3 |
| > 40 | 56% | 44% |

**Number of responses

Unless a user exhibits a high degree of ingenuity, no inverted list will produce "retention" or "forgetting" for the word "memory". The user in this case is left to his own devices or required to read abstracts, titles and descriptor fields in retrieved document sets to find additional clues. The point is, much intellectual effort at relating concepts to each other has already been expended developing controlled vocabularies. Subject switching merely taps the potential which is present in each controlled vocabulary.

Question 16 probed for the possible migration from single data base searches to multi-base searches assuming users had a search tool like VSS. Table 40 shows that about 1/3 of all respondents would consider converting their single-base searches to multi-base searches at the rate of about 40 percent of all their searches. This response was unchanged with time.

The response to Question 17, "Factors affecting the use of VSS", is shown in Figure 9. "Vocabulary differences" is the reason most cited for a user's anticipated usage of VSS, both in '79 and '83. Cost of using subject switching ranked second followed by the number of data bases being used in a search.

On the question of what users would be willing to pay for a subject switching capability (Question 18), their response is unquestionably very little, compared to current connect-hour charges for commercially available data bases themselves (Table 41). It appears that this question has changed little over time.

TABLE 40 MIGRATION TO MULTIPLE DATA BASE SEARCHING VIA VSS

| % Conversion* | 79 | 83 |
|---|---|---|
| 0-20 | 310 | 21 |
| 21-40 | 105 | [illegible] |
| 41-60 | 126 | 5 |
| 61-80 | 44 | 1 |
| 81-100 | 64 | 4 |
| > 40 | 36% | 30% |

*Percentage of single data base searches that might be converted to multiple data base searches via a VSS capability

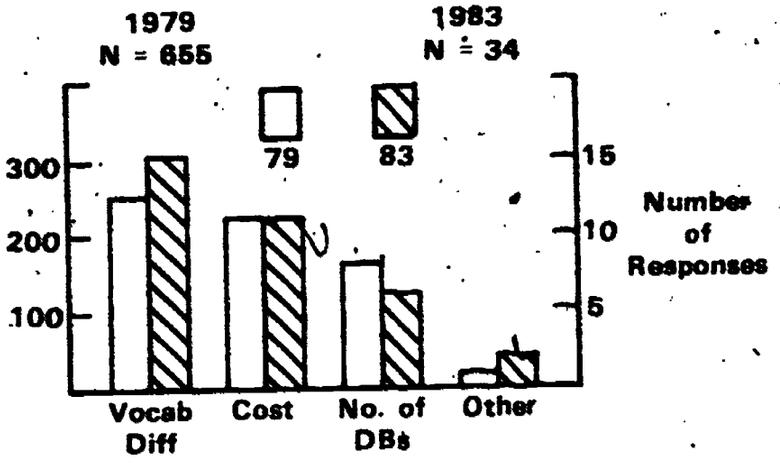



FIGURE 9. FACTORS MOST INFLUENCING ANTICIPATED USE OF VSS

Finally, on the question of just when a user would turn to subject switching for assistance based on the number of data bases being searched, the trend is toward four or more data bases (Figure 10).

TABLE 41 USER'S WILLINGNESS TO PAY FOR SUBJECT SWITCHING

| Dollars per Connect Hour | 79 | 83 |
|---|---|---|
| Nothing | 103* | 5 |
| up to 5 | 405 | 17 |
| 6-10 | 149 | 12 |
| 11-20 | 33 | 0 |
| over 20 | 10 | 1 |
| no response | 54 | 3 |

*Number of respondents

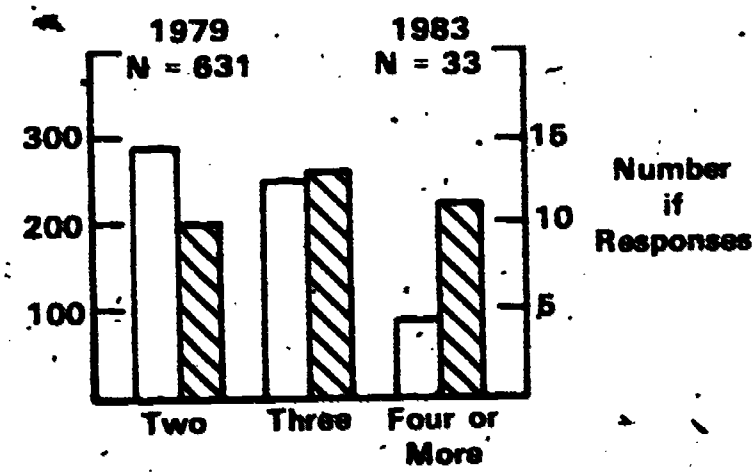



FIGURE 10. NUMBER OF DBS TO BE SEARCHED BEFORE INTERMEDIARY WOULD USE VSS

This coincides with the trend toward a greater number of multiple data base searches involving four or more data bases (Figure 7D). Thus, multiple data base searching trends are increasing in the right direction to justify using a subject switching capability, yet there is evidence to suggest some resistance about using such a system (Table 39). Some of this resistance can be attributed to the fact that a certain percentage of searches don't require an exhaustive, multi-base search (Table 36a). Another factor for possible diminished use of VSS is avoidance of additional time and costs associated with multi-base searching (Table 36b & c) which is primarily a resistance unrelated to the VSS system itself. What users are saying is that a multi-base search involves more of their time, more on-line costs (for each data base to be searched), more off-line print costs, and perhaps more post-search activity such as sorting and organizing of retrieved references in order to eliminate duplicates.

## 11.0 CONCEPTS FOR EXPRESSION-LEVEL SWITCHING

The current VSS system switches on vocabulary terms one term at a time. That is, if the query contains two or more terms and includes Boolean operators, each term must be entered separately into VSS. Following switching, the user must construct his/her own search strategy using the VSS output and appropriate Boolean logic.

To streamline the process, it would be desirable to perform subject switching on the entire search strategy (expression). Such a system was conceptualized but not implemented.

In expression-level switching, VSS would have the capability to "understand" the Boolean expression supplied by a user and would construct the appropriate search logic for each vocabulary specified. As a simple example, assume there are vocabularies; A, B, and C. A search for CAR indicates the following:

AUTO (B)
AUTOMOBILE (A, C)
CAR (C)

This implies that AUTOMOBILE is the controlled term for vocabulary A, AUTO is the controlled term for vocabulary B, and both AUTOMOBILE and CAR are used for vocabulary C. Similarly, assume a search of LARGE yields.

BIG (A)
LARGE (C)

This implies that there is no corresponding concept indexed in vocabulary B.

In expression level switching the user might enter:

BIG .and. CAR  where .and. is a Boolean "AND"

VSS would respond:

BIG .and. AUTOMOBILE (A)

LARGE .and. AUTOMOBILE (C)
LARGE .and. CAR (C)

Note that since the concept of BIGness did not appear in vocabulary B, no search expression using AND logic could be formed. The user could then revert to a single term search or attempt the search on an uncontrolled field (titles, abstracts).

Similarly, if the user had entered:

BIG .or. CAR where .or. is a Boolean "OR"

VSS would respond:

BIG .or. AUTOMOBILE (A)
AUTO (B)
LARGE .or. AUTOMOBILE .or. CAR (C)

In this case lack of a match is not a problem since a Boolean "OR" is satisfied if either term can be found.

In a similar manner, more complex boolean searches could be switched. How to best handle a missing AND concept in a complex query is still an open research question.

## 12.0 SUMMARY AND CONCLUSIONS

VSS is the most advanced tool yet developed for searching on-line bibliographic data bases based on controlled technical vocabularies. The 14 vocabularies in VSS represent an investment of about 52 man-years of creative work by the original vocabulary developers. This stored knowledge base was evaluated as an aid for structuring and enhancing search strategies.

### 12.1 Controlled Vocabularies

The use of controlled vocabularies in today's search environment cannot be denied. The survey data shows that the number of searches involving both controlled and free text terms is actually on the increase, while searches involving one or the other approach, exclusively, are rather unpopular. This finding is reinforced by the fact that users actually prefer, by a wide margin, the combination of controlled plus free text searches over either approach used individually.

The fact that searchers not only use controlled plus free-text terms in an increasing number of their searches but actually prefer this approach tells us that this phenomenon is not simply a case of "blind faith" usage of controlled descriptors just because they are available. If that were the case, the usage response would differ from the preference response.

On the basis of these findings, we conclude that users perceive no superior indexing method in IS&R systems, opting instead for the synergy of two methods combined.

### 12.2 Multiple Data-Base Searching

Multiple data-base searching patterns are changing with time. The trend is toward increased usage of multiple data bases in on-line searches. This trend suggests the need for search aids which transcend data bases or mitigate the differences among them.

In other words, users need navigational aids to search more effectively across different data bases. Better than 75 percent (42 out of 55) of those who participated in the VSS field evaluation responded favorably to

the concept of subject switching, suggesting the wide appeal of a navigational aid.

Probing deeper into the multiple data base searching issue, but from the viewpoint of user inhibition due (a) to the plethora of available data bases or (b) to the possible difficulty of using less familiar files, our survey showed that neither issue was a serious factor for confining a search to a single data base. Instead, users indicated that when searches were limited to a single data base it was primarily because the user did not require an exhaustive search or there was a cost limitation which precluded a broader search across multiple files.

It is concluded that the popularity and user preference for controlled plus free text searching and the favorable response to subject switching as a navigational aid for cross-file searching lends credibility to the approach being pursued in this research.

## 12.3 The VSS System

Feedback from the field evaluation provided valuable insights into the strengths and weaknesses of VSS as a system and concept. Users who thought VSS would make their job easier (about 40 percent) felt that it would save them time and money because they would have fewer printed thesauri to consult. They also liked the juxtaposition of terms from various thesauri and thought VSS would improve their presearch preparation.

About half (45 percent) of all participants expressed an overall positive attitude toward VSS or thought it was interesting but needs more work. Many (about 75 percent) participants responded positively to the concept of subject switching.

However, several weakness were also observed. The menu approach to VSS, with no provision for direct access, led to user frustration. This was exacerbated by the unforeseen high usage of 300 baud terminals in the on-line community. Menu-driven systems are not amenable to slower terminals. Also, once users become familiar with a system, menus are not necessary and thus unacceptable.

Users felt, and the investigators concur, that vocabularies must be kept current with rapidly changing technology in order to meet the needs of searchers and end users. This is a potential problem for vocabulary developers and a real problem for a system like VSS. The VSS vocabularies used in the field evaluation were 3 to 5 years old, due simply to the cost of reformating, reprocessing and rebuilding new files with new versions of the vocabularies.

Users expressed a need for more synonym and hierarchical relationships than were provided. Users also wanted scope and history notes. The former is a shortcoming of vocabulary suppliers and is not easily rectifiable, while the later is a shortcoming of VSS itself and is easily rectifiable.

Another problem was the poor quality of output derived from stemming algorithms and inverted-file adjacency features. These types of switching options created more noise in VSS output than useful terms and were undoubtedly responsible for establishing a negative impression of the system in the minds of some users.

Based on the feedback provided by the users, we believe that with additional work, a system such as this can achieve the potential usefulness that the users and the investigators expected.

## 12.4 Performance of VSS Modules

On one hand, it was shown that the most successful switching, based on a performance measure which takes into consideration the amount of output and its relevance, is a function of the similarity of the vocabularies (see Section 6.0). Vocabularies that are similar in syndetic structure and subject content produce more satisfactory switching results than those that are dissimilar in one respect or another.

In this version of VSS, a term-level analysis suggested that the social science module would give consistently better performance than the other vocabulary modules, followed by the business module, the physical science module, and finally, the life science module.

There was a direct relationship between module performance and subject similarity among the vocabularies. One measure of subject similarity

is the rate of exact matches found during switching. Also, it was observed that the social science vocabularies were very similar to one another with respect to their syndetic structures.

On the other hand, module performance rankings were quite different in the field evaluation by intermediaries (see Section 9.0). Here, intermediaries were asked to rate VSS on a scale of 1 to 5 for each search term entered (1 meant VSS was of no help, 5 meant VSS was very helpful). In a mixed model analysis of covariance the module rankings were as follows: physical sciences (2.45), social sciences (2.38), life sciences (2.12) and business (2.05).

The principal reason for the big difference in module rankings between the formal analysis and the field evaluation was the inclusion of related terms in the field-tested version of VSS. These terms were not present in the version used to conduct the formal analysis. The top two modules rated by intermediaries, physical and social science, contained vocabularies that were rich in related terms.

The performance difference described above shows just how important the related term is, and why VSS is now best viewed as a tool to provide a shopping list approach to searching.

Although the formal analysis of switching strategies served a useful purpose in planning pre-defined strategies for subsequent use in field experiments, user ratings represent a better indicator of VSS system performance. It also appears that for maximum usefulness, all of the original syndetic structure should be incorporated into a system like VSS, including related terms, scope and history notes, subject category codes (e.g., COSATI codes) and any other special relationships provided by the vocabulary supplier.

## 12.5 End User Experiments

VSS was evaluated at three sites in actual retrieval situations to estimate its effect and impact on end user satisfaction. Experiments were designed whereby end users evaluated citations retrieved from a normal search with those produced from a VSS enhancement or modification to the same search.

Two types of experiments were designed, iterative searching and parallel searching. In the iterative searching experiment, a normal search was repeated subsequent to its modification by VSS. In the parallel experiment, two intermediaries worked on the same end user request beginning with the pre-search interview. One intermediary conducted a normal search, the other used VSS.

VSS performed quite well in the iterative searches. In both searches, end user satisfaction was high and VSS produced a significant amount of relevant output. In one search, VSS accounted for 20 percent of all relevant output, in the other search, 62 percent.

In the experiments where a parallel approach was employed, VSS performance was about as good as non-VSS performance but end user evaluations appeared to be inconsistent and illogical. Of the six searches conducted via the parallel design, users gave VSS searches a higher rating in two, Non-VSS searches a higher rating in two. In two searches the ratings were equal. The success of these searches appeared to be dependent on interpretation of the end user question rather than the use of VSS. Therefore, it was concluded that experiments involving two intermediaries probably confounded the issue.

As it turned out, the parallel design was a difficult undertaking because search intermediaries seemed hard pressed for time in their real-life work environments. Many found it difficult to start the experiment and nobody completed all the searches planned (six per site). One site failed to participate at all even though they were briefed on the experiment and trained on the system.

It appears that the iterative design is the most appropriate method for evaluating the effect of VSS because only one intermediary is involved in the search and by modifying his/her original search with VSS terms, a clear distinction between searches is possible.

Several additional observations can be made about these end user experiments.

(1) High-recall type search strategies can lead to some rather low search precisions at times. (e.g., 2 and 3 percent). However, user toleration of noise at these levels may be quite high.

(2) High-recall type search strategies don't always produce high recall. In several instances, VSS produced a significant amount of additional unique relevant citations even though the Non-VSS search was geared for high recall.

(3) Search strategies can be a very complex series of nested statements. It appears that some of these nestings and Boolean combinations are simply done to reduce the document set to a manageable number. Therefore, what is thought to be a valuable data base attribute, namely its retrospective depth, turns out to be a liability in certain high-recall searches, where mental gymnastics are required just to identify a set of citations which the user can (a) afford to print out and/or (b) cope with once it is delivered.

## 12.6 Evaluation of VSS by Intermediaries

VSS was evaluated by 65 information professionals, from brokers to professors, indexers to searchers. The evaluation consisted of a packet of materials explaining VSS and how to access it on-line, a user evaluation form, and a scheduled week in which to log-on and use the system. Private, government, and academic sectors were represented. Overall, user proficiencies on the VSS data bases ranged from just average to virtually little or none.

Generally, the participants thought VSS was easy to learn, use and understand, but they were less certain about its capabilities and output. However, 40 percent thought VSS would make their job easier while only 9 percent felt it would make their job harder. Also, participants with fewer years of on-line experience and those with higher degrees thought VSS would make their jobs easier. This finding was based on a Pearson correlation between experience or education and performance ratings given.

The participants tried 623 searches in VSS. Usable results were obtained in 62 percent of the attempts, and of these, about 22 percent produced six or more useable terms per entered term. Although the overall rating for VSS was 2.25 on a 5-point scale this rating increased to 3.08 when only those searches which produced usable output were considered. Also, the average rating increased directly with the amount of usable output. In searches where 6-10 usable terms were produced the ratings averaged 3.6, and where more

than 10 usable terms were produced, the rating averaged 4.1 on a 5-point scale.

It is obvious that a system such as VSS is only as good as the vocabularies in it. It's performance is directly related to the number, specificity and currency of such vocabularies. The larger the system, and the more up-to-date the vocabularies, the better the performance. It appears that high performance ratings and high user acceptance are well within the grasp of such a system.

A mixed model analysis of covariance was also employed to express the VSS rating assigned by a participant (to each of the 623 searches) as a function of (a) VSS module chosen, (b) the participant's proficiency with a particular VSS module, and (c) the unique qualities of each individual participant. The model assumed that the average ratings assigned to searches in one VSS module may differ from those assigned in another, and it also assumed that VSS ratings change linearly with increases in participant proficiency within a particular module. The unique qualities of individual participants were assumed to be random.

Results indicated that all factors were statistically significant but their estimated effects differed. In decreasing order, the relative importance of the factors were: (1) uncontrolled factors (from search-to-search); (2) unique qualities of the participants; (3) the participant's proficiency with a VSS module; and (4) the VSS module chosen.

The effect due to a participant's proficiency was found to be inversely related with his VSS rating. In other words, participants who rated themselves proficient in a given module rated VSS lower on the average than those who claimed to have no proficiency.

The effect due to module or subject area was the least important of the variables studied. A spread of 0.4 points was observed between the highest and lowest rated modules.

Finally, there was a considerable amount of variability in VSS ratings that cannot be explained. This variability represents the uncontrolled factors in the experiment. The size of this variation was estimated at $\pm 2.00$ points.

In open-ended questions, about four and one-half times as many participants thought VSS would make their job easier as thought VSS would make

their job harder. Close to half of all participants thought VSS was very valuable, valuable, or "interesting but needs more work". The concept of subject switching as a search tool was very well received; about three-fourths of all participants responded positively to the approach.

In view of these results, it was concluded that subject switching has a potentially wide appeal to information professionals, but VSS will need several improvements or even major redesign if it is to address their needs and concerns. These needs, briefly summarized are: (1) include more vocabularies; (2) keep the system updated with current versions of vocabularies; (3) include the full syndetic structure of each vocabulary; (4) provide direct access as an alternative to menus; (5) eliminate stemming and adjacency features; (6) allow users to combine vocabularies in any combination desired; (7) make it inexpensive to use.

## 12.7 On-line Users Survey

The user survey provided insights into how searching patterns are shifting with time. Areas dealing with controlled vocabularies and multiple data base searching have already been covered earlier (Sections 12.1 and 12.2).

On-line expenditures by individual searchers have indeed grown at the rate of about 30 percent compounded annually. This agrees with various market study results reported in the literature. The average monthly on-time expenditure per individual searcher was $972 in 1983, compared to $377 in 1979. However, spending patterns differed from one employment sector to another. In 1979, the average monthly on-line expenditures by individuals by sector were: for-profit sector ($467), government sector ($398), non-profit sector ($276), and academic sector ($222). There were no surprises here except to show just how low expenditures for on-line searches in the academic sector are relative to the other sectors. The 1983 survey was not large enough to provide reliable figures by sector.

A greater percentage of end users is paying for his searches today than in 1979. Conversely, library budgets are supporting fewer end-user searches today compared to 1979.

In 1979, respondents indicated that the most common type of limitation (if any at all) placed on a search was by date or number of citations. The next most common limitation was by data base, followed by a specified cost ceiling. In 1983, the most common limitation placed on a search was data base, followed by date or number of citations, and finally cost ceiling.

The usage pattern for seven major retrieval services showed very little change over the past four years. Dialog was still the most frequently used system, DTIC the least frequently used. Also, there was no shift in the relative standings of these major systems based on the usage question. However, numerous new systems were identified, the most frequently mentioned being NEXIS, CAS ONLINE, and DOW JONES.

There has been a shift over the past four years in the subject areas where users feel multiple data base searching would be most useful. In 1979, the three top areas were: life sciences/medicine, agriculture, and engineering, respectively. In 1983, respondents rated psychology, business/economics, and life sciences/medicine their top three choices.

With the simultaneous decline in (a) searches using controlled terms exclusively and (b) the respondents anticipated use of VSS over time, a possible cause and effect relationship is seen between these two questions which leads to the conclusion that there is probably a misconception about how VSS might be used. The misconception is that VSS, being based on controlled vocabularies, must be useful only for controlled-term searches.

On the contrary, VSS is useful for all types of subject searches, controlled or free text because VSS is approaching the breadth and depth of unique words and phrases that title and abstract fields contain. Fields that are rich in technical terms are the ones that searchers turn to when "fine tuning" a search. We believe that VSS is a rich source of technical terms to be used in any one of many imaginative ways during subject searching, regardless of the approach being taken. In other words controlled terms can be used in free text searches and vice versa.

On the migration question, about 1/3 of the users indicated they would expand a single data base search into a multiple data base search greater than 40 percent of the time if they had a system like VSS. However, they would not pay much more than about $10 per connect hour for a system like

VSS. Their need for VSS increased as their need for more data bases per search increased. The survey showed that the trend in multiple data base searching is in the direction of 4 or more data bases per search.

Users see subject switching as a valid, useful concept, but one that they should not have to pay much for. As some of them see it, the on-line vocabulary system has to compete with cheap, off-line, printed versions of the same thing.

There is no question that a system like VSS can be designed with an efficient and streamlined user interface, larger and more up-to-date vocabularies, and even more of them. However, the question of greatly reduced online rates is a marketing and business decision involving the data base (and vocabulary) suppliers and the providers of on-line retrieval services.

It is believed that in time, users would become very efficient using a system like VSS, so the extra cost incurred by using VSS ultimately may be very small relative to the total cost of the search.

The benefits are reduced search-preparation time, improved search strategies and retrieval, and greater usage of existing data bases. Therefore, all parties in the on-line search scenario derive some benefit from a navigational aid such as VSS. If the benefits are substantial, on-line vendors and data base producers could afford to reduce or give away the navigational aid on the theory that more revenue will be generated via greater data base access.

## 13.0 RECOMMENDATIONS

On the basis of this research and the feedback from users who evaluated VSS, the following is recommended:

(1) Build an entirely new model of VSS, based on the relational data base model as a solution to the update problem and therefore, as a means of maintaining current vocabularies in the system.

(2) Expand the breadth and depth of VSS by including many more vocabularies and all of the syndetic relationships available in them.

(3) Streamline the user interface to permit, rapid direct access to VSS files and eliminate non-productive features (e.g. stemming and inverted file adjacencies).

(4) Consider storing VSS on a videodisc in digital form and perform all navigational tasks.

## 14.0 REFERENCES


(1) Total Databases Top 2000, Information Today, Vol. 1, No. 4, April 1984, P. 1.

(2) Spigai, Frances G., "Impediments to Online Database Use: Size and Its Importance in Database Design and Online Customer Support".

(3) Doszkocs, Tamas E. and Rapp, Barbara A., "Searching Medline in English: A Prototype User Interface with Natural Language Query, Ranked Output and Relevance Feedback," Proceedings of the ASIS Annual Meeting Vol. 16, pp. 131-39, 1979.

(4) Cooper, William S., "Exploiting the Maximum Entropy Principle to Increase Retrieval Effectiveness", Journal of the ASIS, Vol. 34 No.1, pp. 31-39, 1983.

(5) Marcus, Richard S., "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries", Journal of the ASIS, Vol. 34, No. 6, pp. 381-404, 1983.

(6) Heller, Stephen R., "Linking Scientific Data Bases--The NIH-EPA Chemical Information System", Online, October 1980, pp. 45-57.

(7) Horowitz, Aren J., et al., "An Intelligent Terminal for Access to Heterogeneous Chemical Information Systems", Proceedings of 43rd ASIS Annual Meeting, Vol. 17, 1980, pp. 83-85.

(8) Bates, Marcia J., "Information Search Tactics", Journal of the ASIS, July 1979, pp. 205-214.

(9) Standera, O. R., "Some Thoughts on Online Systems: The Searcher's Part and Plight", Proceedings of 41st ASIS Annual Meeting, Vol. 15, 1978, pp. 322-325.

(10) Niehoff, Robert, et al., "The Design and Evaluation of a Vocabulary Switching system for Use in Multi-Base Environments", Battelle Columbus Laboratories, Final Report, NSF Grant No. IST 77-04498, February 1980.

(11) Niehoff, Robert T., "Development of an Integrated Energy Vocabulary and the Possibilities for Online Subject Switching", Journal of the American Society for Information Science, Vol. 27, No. 1, 1976, pp 3-17.

(12) Maranjiam, Lorig, and Boss, Richard W., "Fee Based Information Services. A Study of a Growing Industry, R. R. Bowker Co., 1980.

# APPENDIX A
# VOCABULARY PROCESSING AND VSS INPUT FORMAT

At the end of Grant IST-7704498 six full scale vocabularies (DOE, EI, INSPEC, NASA, and two from Chemical Abstracts) were included in VSS. During Grant-IST-79-11190 several new vocabularies were acquired and converted from the vendor's format to a standard format used as inpu to the VSS file-building programs. The new vocabularies were:

ABI INFORM Thesaurus
APA Thesaurus (Psychology Abstracts Data Base)
ERIC Thesaurus
BIOSIS Master Index Authority File
IRON DATA CENTER Thesaurus
MeSH Vocabulary
Management Contents Thesaurus

While several of the format conversions were straightforward, three of the vocabularies deserve special mention.

The IRON DATA CENTER Thesaurus is interesting since it is trilingual, providing Spaniel-Portugese-English translation. This vocabulary has demonstrated the capability of VSS to perform such translation.

The BIOSIS Master Index Authority File proved interesting since it did not resemble a conventional thesaurus structure. In general, the BIOSIS file is free text and the Authority File was set up to indicate how to best search a topic. The following examples indicate how the Authority File was restructured to follow a conventional thesaurus structure.

| MIAF ENTRY | THESAURUS STRUCTURE |
|---|---|
| AARDVARK<br>KW: AARDVARK (20) | AARDVARK |

If KW matches, set up a valid lead term.

| | |
|---|---|
| ABATEMENT<br>KW: ABATE$ (170) | ABATE<br>ABATEMENT use ABATE |

If KW does not match, set up USE reference. Object of USE is set up as valid lead term. Note that truncation indicator is dropped. While this is not desirable for several reasons, it is necessary to allow switching to other vocabularies.

```
ACCESS                                    ACCESS use
  KW:  ACCESS (150) ACCESSI$ (210)          ACCESS, ACCESSI ORed
                                          ACCESSI use
                                            ACCESS, ACCESSI ORed
```

If multiple words in KW field, set up USE reference to scope note.

```
ACETYLCHOLINE                             ACETYL
  KW:  ACETYL & CHOLINE                   ACETYLCHOLINE use
                                            ACETYL .AND. CHOLINE
                                          CHOLINE
```

If split KW, set up USE reference to special "AND" construct. Make sure each part of "AND" construct is a valid lead term.

SEE ALSO entries were isolated as related terms. These are not used in the current version of VSS, but have been captured for possible future inclusion.

The Concept Codes and Biosystematic Codes were mapped into their headings.

```
BC:  -----                                BC07600 use
     MYCOPLASMATALES                        MYCOPLASMATALES
     1969-78:  BC07600 (2630)
```

In addition, a second vocabulary (called the BCCC vocabulary) was defined which maps headings into the Concept Codes and Biosystematic Codes.

```
BC:  ----                                 MYCOPLASMATALES use
     MYCOPLASMATALES                        BC07600
     1969-78:  BC07600 (2630)
```

This approach was chosen with the CODES since any one of several relationships may exist between a lead term and the associated codes.

Therefore, we defined relationships between Codes and Headings rather than between Codes and lead terms.

Although several other minor points were addressed, the above examples indicate how the BIOSIS Authority File was restructured for VSS.

The final vocabulary which deserves special mention is the MeSH vocabulary. The standard MeSH vocabulary was converted in a fairly straightforward manner. However, in addition to the standard vocabulary records, we received a large file of chemical records from NLM.

These records contain mappings among MeSH headings, CA registry numbers, and chemical substance names. With this information and work similar to that performed for BIOSIS, three MeSH vocabularies were created. The first was the standard MeSH vocabulary with the addition that CA registry numbers and chemical substance names were mapped to the appropriate MeSH heading. In the second file the substance names were mapped to the proper CA registry number, and in the third file the registry numbers were mapped to the substance names. This allowed powerful switching for chemical topics.

The current VSS input format is shown in Figure A-1.

FIGURE A-1. VSS INPUT FORMAT

RECORD LAYOUT

| Position | Content |
|---|---|
| 1 | Relationship Code (Alphanumeric, See Below) |
| 2 | Blank |
| 3-4 | Line Sequence Number--for Continuation Cards (e.g., 01, 02) |
| 5 | Blank |
| 6 | Vocabulary Source Code (Alphabetic; Arbitrarily Assigned) |
| 7 | Blank |
| 8-67 | Term |
| 68-80 | Blank |

| Relationship Code | Relationship |
|---|---|
| 0 | LEAD (or MAIN) TERM |
| 1 | SCOPE NOTE |
| 2 | USE |
| 3 | |
| 4 | USED-FOR (UF) |
| 5 | UF+ |
| 6 | SPECIAL SCOPE NOTE |
| 7 | BROADER TERM (BT) |
| 8 | NARROWER TERM (NT) |
| 9 | |
| A | SEEN-FROM (SF) |
| B | RELATED TERM (RT) as designated by supplier |
| C | Subject Category as designated by supplier |
| D | Suggested NT as designated by supplier |
| E | Suggested BT as designated by supplier |
| F | Suggested RT as designated by supplier |
| G | Array NT as designated by supplier |
| H | Array BT as designated by supplier |
| I | Array RT as designated by supplier |
| J | Top terms |
| K | Frequency count |

# APPENDIX B

# VSS FILE BUILDING ROUTINES AND RECORD STRUCTURES

# APPENDIX B

## VSS FILE-BUILDING ROUTINES AND RECORD STRUCTURES

Input to the Vocabulary Switching System (VSS) file building operation consists of all preprocessed vocabulary records in VSS format. (See Figures 1 & 2 for examples of vocabularies in VSS format.) Lead terms must be in alphabetic order. When a lead term appeared in more than one vocabulary, a separate entry was created for each occurrence. These redundant occurrences were further sorted by vocabulary source code.

Following each lead term entry there appeared various relational entries, such as USE, multiple USE, UF (used for), scope-note, BT (broader term), NT (narrower term) and RT (related term). The relational entries were fully reciprocated; that is, a BT entry under one lead term would be matched by a corresponding NT entry of the lead term under its broader term. Consequently, nearly half of the entries in the input (one entry of each reciprocated pair) were not essential to build the vocabulary files. For processing efficiency, both BT and NT entries were accepted as they were encountered, together with USE and special scope note entries.

To form the keys to the term, phrase, stem, and word files, all entries were edited by a term standardization routine to eliminate minor variations due to punctuation and spacing. This routine converted all characters except letters and numerals to spaces, then reduced all multiple spaces to single spaces. No further transformation (such as singularization or removal of prepositions) was employed. Terms that would eventually end up in the concept file were not edited and are referred to as unsanitized terms.

### PASS 1

Entries consisting of term, relational code, and vocabulary code, were processed sequentially. Records for lead term entries were built up in a working area as the various relational entries were processed and then were written to a file keyed by standardized term. Records for relational entries were created or updated and written to the file before processing the next entry.

# FIGURE 1
## PREPROCESSED VOCABULARY RECORDS
## MeSH' THESAURUS (MeSHr)

| A | B | C | Term |
|---|---|---|---|
| 0 | 1 | R | (+)-6-FLUORO-2,3-DIHYDROSPIRO(4H- 1-BENZOPYRAN-4,4↑-IMIDAZOL |
| 0 | 2 | R | IDINE)-2↑,5↑-DIONE |
| 2 | 1 | R | 69880-53-1 |
| C | 1 | R | (+)-8-HYDROXY-4,5,6,7-TETRAMETHYL-1H-2-BENZOPYRAN-1,3(4H)-DI |
| 0 | 2 | R | ONE |
| 2 | 1 | R | 13277-76-4 |
| 0 | 1 | R | (+-)-N-CYCLOHEXYL-N↑-(4-(3-((1,1-DIMETHYLETHYL)AMINO)-2-HYDR |
| 0 | 2 | R | OXYPROPOXY)PHENYL)UREA |
| 2 | 1 | R | 57460-41-0 |
| 0 | 1 | R | (+-)-1-ACETOXY-5,6,6A,7,8,9,10,10A-OCTAHYDRO-9 HYDROXY-6-BET |
| 0 | 2 | R | A-METHYL-3-(5↑-PHENYL-2↑-PENTYLOXY)PHENANTHRIDINE HYDROCHLOR |
| 0 | 3 | R | IDE |
| 2 | 1 | R | 72028-54-7 |
| 0 | 1 | R | (+/-)-ALPHA-HYDRAZINO-3,4-DIHYDRO-2(1H)-ISOQUINOLINEBUTANOIC |
| 0 | 2 | R | ACID |
| 2 | 1 | R | 58489-32-0 |
| 0 | 1 | R | (+/-)-HOMOARGEMONINE |
| 2 | 1 | R | 39013-26-8 |
| 0 | 1 | R | (+/-)-1-((1,1-DIMETHYLETHYL)AMINO)-3-(2-(2-PROPYNYLOXY)PHENO |
| 0 | 2 | R | XY)-2-PROPANOL, HYDROCHLORIDE |
| 2 | 1 | R | 36902-82-6 |
| 0 | 1 | R | (+/-)-1-((1,1-DIMETHYLETHYL)AMINO)-3-(4-(4-(TRIFLUOROMETHYL) |
| 0 | 2 | R | -1H-IMIDAZOL-2-YL)PHENOXY)-2-PROPANOL |
| 2 | 1 | R | 62960-75-2 |

A - relational code

B - card number

C - vocabulary code

Relational Code:

0 = Lead Term

2 = Use

# FIGURE 2
## PREPROCESSED VOCABULARY RECORDS
## IRON THESAURUS

| A | B | C | Term |
|---|---|---|---|
| 0 | 1 | F | A |
| 2 | 1 | F | ARGON |
| 0 | 1 | F | A-L |
| 2 | 1 | F | ALLEGHENY LUDLUM |
| 0 | 1 | F | AACHEN UNIV |
| 4 | 1 | F | INSTITUT FUR EISENHUTTENWESEN |
| 4 | 1 | F | TECHNICAL UNIV OF AACHEN |
| 4 | 1 | F | TECHNISCHEN HOCHSCHULE AACHEN |
| 7 | 1 | F | EUROPE |
| 7 | 1 | F | GERMANY |
| 0 | 1 | F | AB SAHCO |
| 0 | 1 | F | AB COLD-BOUND PELLETS |
| 2 | 1 | F | COLD BOUND PELLETS |
| 0 | 1 | F | AB ST MILJOTEKNIK |
| 0 | 1 | F | ABBATTISTA, F. |
| 0 | 1 | F | ABBOTT, A.F. |
| 0 | 1 | F | ABE, H. |
| 0 | 1 | F | ABE, T. |
| 0 | 1 | F | ABE, Y. |
| 0 | 1 | F | ABEL, O. |
| C | 1 | F | ABEX CORP. |
| C | 1 | F | ABLANDAMIENTO |
| 1 | 1 | F | (SPA) |
| 2 | 1 | F | SOFTENING |
| 0 | 1 | F | ABM ANNUAL CONGRESS |
| 0 | 1 | F | ABONNENC, J. |
| 0 | 1 | F | ABRAHAM, K.P. |
| 0 | 1 | F | ABRAMS, H.B. |
| 0 | 1 | F | ABRASION INDEX |
| 1 | 1 | F | (SPECIFICATIONS) |
| 4 | 1 | F | INDICE DE ABRASAO-PORT |
| 4 | 1 | F | INDICE DE ABRASIVIDAD-SPA |
| 7 | 1 | F | PHYSICAL SPECIFICATIONS |
| 7 | 1 | F | SPECIFICATIONS |

Relational Codes:

0 = Lead Term
1 = Scope Note
2 = Use

4 = Used - For
7 = Array Narrower Term

As each new entry was accepted (except for multiple USE entries) the keyed file was accessed to determine whether or not the term had been encountered previously. If it had been the existing record for that term was read in and stored for updating. If not, a new record was created and a "concept number" was assigned for inclusion in the record. The concept number, merely the next available integer, was assigned whether or not the term was a valid concept. All subsequent references to the term used the concept number rather than the text of the term. These references are described elsewhere in the discussion of concept-number cells.

A special procedure was employed in the case of multiple USE entries. Instead of accessing the individual terms designated as multiple USE, they were used to build a synthetic term consisting of the several USE terms, separated by the operator ".AND.". When the last multiple-USE reference under a lead term had been processed in this manner, the combined expression was treated much the same as a single-term concept. That is a series of input entries consisting of:

TERM A
USE TERM B
USE TERM C

would be treated as

TERM A
USE TERM B .AND. TERM C

and the expression TERM B .AND. TERM C would have an assigned concept number apart from the concept numbers assigned to TERM A, TERM B, and TERM C.

In a like manner, special scope note expressions were handled as if they were single term concepts. Thus, in an entry:

TERM D
USE TERM E OR TERM F

the expression TERM E OR TERM F was considered a concept.

The actions described above created a temporary file containing one record for each unique lead term and a number of additional records for the special scope note or multiple USE concepts. Each record consisted of one concept number parcel (described elsewhere). The first concept number cell contained the concept number for the term and flags indicating those

vocabularies where the term was valid. If the term was not valid, a second concept number cell was used to identify the valid term (the USE cross reference).

Using the same standardized term as a key and following the same processing algorithm, a second temporary file was built containing one record for each unique lead term, multiple use term, and special scope note term. Each record contained a unique unsanitized term for each applicable vocabulary.

**PASS 2**

The two files created in PASS 1, temporary field and unsanitized term file, were used as input to produce a TERM file and a CONCEPT file. Each record in the temporary file was processed in sequence. Lead terms, whether or not they were valid concepts, were used as keys to term file records containing lead-term and USE (both single and multiple) cells in a concept-number-parcel-record. A concept file record was also created for valid concepts only. For this file, the concept number is used as the key. The record includes vocabulary flags for lead-term usage, unsanitized terms of the concept for each vocabulary, and, if appropriate, a concept-number parcel containing BT and NT cells from the CNP of the temporary file.

**PASSES 3 and MAKEUP**

PASS 2 also generated a relatively small number of related concepts, called co-related terms (CRT) in VSS. If for a given term TERM A in the term file one vocabulary said USE TERM B and another vocabulary said USE TERM C, then a "co-related term" relationship was assumed between TERM B and TERM C. Thus, for any vocabulary containing both TERM B and TERM C as valid concepts, each would reference the other as a CRT. If for a third vocabulary TERM A was valid, the CRT relationship would exist between TERM A, TERM B, and TERM C. PASS 3 adds these co-related terms to the CONCEPT file.

MAKEUP also uses the TERM file generated by PASS 2. It finds all USE references in the TERM file and reciprocates the relationship by creating USED

FOR references. As in PASS 2, the CONCEPT file is then updated with the USED FOR terms of this relationship.

**PASSES 4, 4P, 4S, 4W**

The term file was used as input to build three additional files--PHRASE, STEM and WORD files. Each key in the term file was processed in sequence, with the key being transformed into a stemmed phrase, a series of individual words, and a series of individual stems. The algorithm used in these transformations was identical to that used in the Logic module of the switching system.

For each key generated, that is, for each phrase, each word, and each stem, a transaction was issued that contained the key and one cell from the term file concept-number parcel. This was done for each cell in the CNP. The reason for breaking up the term file records in this manner was the later need to merge cells under one key that came originally from many different term-file records. For example, two terms in the term file might be ELECTRICITY and ELECTRONS. Cells for both of these terms would be found in one record in the stem file under the key ELECTR.

The three sets of transactions generated in PASS 4 (phrase, word, and stem) were processed in a similar manner. First, all transactions were sorted, with the primary sort on the key. Then they were processed in order, either adding new concept-number cells or adding vocabulary flags to existing concept-number cells. (Note that the concept-number cell is comprised of a concept number, a relationship code, and a vocabulary-flag subcell. Thus if two cells have the same concept number and the same relationship code, their vocabulary-flag subcells can be merged.)

## CONCEPT NUMBER PARCEL TABLE

Name: CNP

Usage: The CNP format is used in the CONCEPT, TERM, WORD, STEM, and PHRASE files. In all but the CONCEPT file the CNP comprises the entire record format.

Table Description: The CNP is a data structure for the representation of concept numbers and related infofmation.

### GENERAL CNP FORMAT

| Cell | Description |
|---|---|
| 1 | CNP cell count (N) |
| 2....N+1 | Concept number cells (see below). |

### CONCEPT NUMBER CELL FORMAT

The Concept Number Cell is 60 bits in length plus filler as required and it occupies one or two computer words. The first two subcells are left-justified in the first word of the cell and the last two subcells are right-justified in the last word of the cell.

| Number of Bits | Name | Description |
|---|---|---|
| 24 | Concept Number | The concept number is the unique identifier assigned to each identifiable, VALID (postable) index term/thesaurus term in the concept file. It can be used as the key to the concept file. |
| 6 | Concept Type | The concept type indicates the relationship of the concept number to the key of the record containing the CNP.<br>0 = use for<br>1 = co-related term<br>2 = narrower term<br>3 = broader term<br>4 = related term |
| (Variable) | Filler | |
| 6 | Discretionary | This field may be used by different modules for any purpose desired. Care must be taken however to insure that no conflict arises in its usage. |

| Number of Bits | Name | Description |
|---|---|---|
| 24 | VBT (Vocabulary bit table) | This table denotes what vocabulary(s)/thesaurus this concept is valid in. The presence of a vocabulary is indicated by the presence of a "1" bit in the VBT. The vocabularies bits are numbered from left to right from 1 to 24. The vocabulary code corresponding to any given position/bit may be retrieved from the VCT, vocabulary table, where the position (1 to 24) is the index to the VCT. |

## CONCEPT FILE

The Concept File is keyed by concept number and contains (1) the corresponding unsanitized term as it appears in each vocabulary, and (2) a concept number parcel containing concept number cells that point to co-related, narrow, broader, and related concepts.

## CONCEPT FILE FORMAT

Organiztion: SK (symbolic keyed file).
Binary records.
Random access.
Key: Concept number (24 bit unsigned numeric):
Record: The record contains the unsanitized text of the concept for valid vocabulary terms along with vocabulary flags indicating in which vocabularies the unsanitized concept term is found. Also included are pointers that link a concept to other concepts in the concept file. The identified relationships are:

co-related term-CRT
narrower term-NT
broader term-BT
related term-RT
use term

## RECORD FORMAT

| Cell | Length | Description |
|---|---|---|
| 1 | 1 word | Record information cell (see below) |
| 2 | 1 word | Concept information cell (see below) |
| 3 n | Variable, up to 5 wordsBoolean conjunction of lead terms, or special scope note when regarded as a valid concept | Unsanitized term. This is the orginal lead term. |
| n+1 | 1 word | Same as cell 2 |
| n+2 o | Variable, up to 5 words | Same as cell 3 |
| o+1 | 1 word | Number of cells to follow |
| o+2 | 1 word | Concept number cell (CNC) as described elsewhere. One for each CRT, NT, BT, RT, or use term |

## RECORD INFORMATION CELL FORMAT

This cell is 60 bits in length and occupies one computer word.

| Number of Bits | Name | Description |
|---|---|---|
| 6 | Count | Number of unsanitized terms stored |
| 6 | Pointer | Word position of the cell preceding the CNC relative to the record information cell |
| 22 | Filler | |
| 2 | Flag | This flag is set to non-zero when the concept is not a lead term, thus not a TERM file key:<br><br>0 = concept is lead term<br>1 = concept is Boolean expression<br>2 = concept is scope note |
| 24 | VBT | Vocabulary bit table indicating those vocabularies in which the unsanitized term is found |

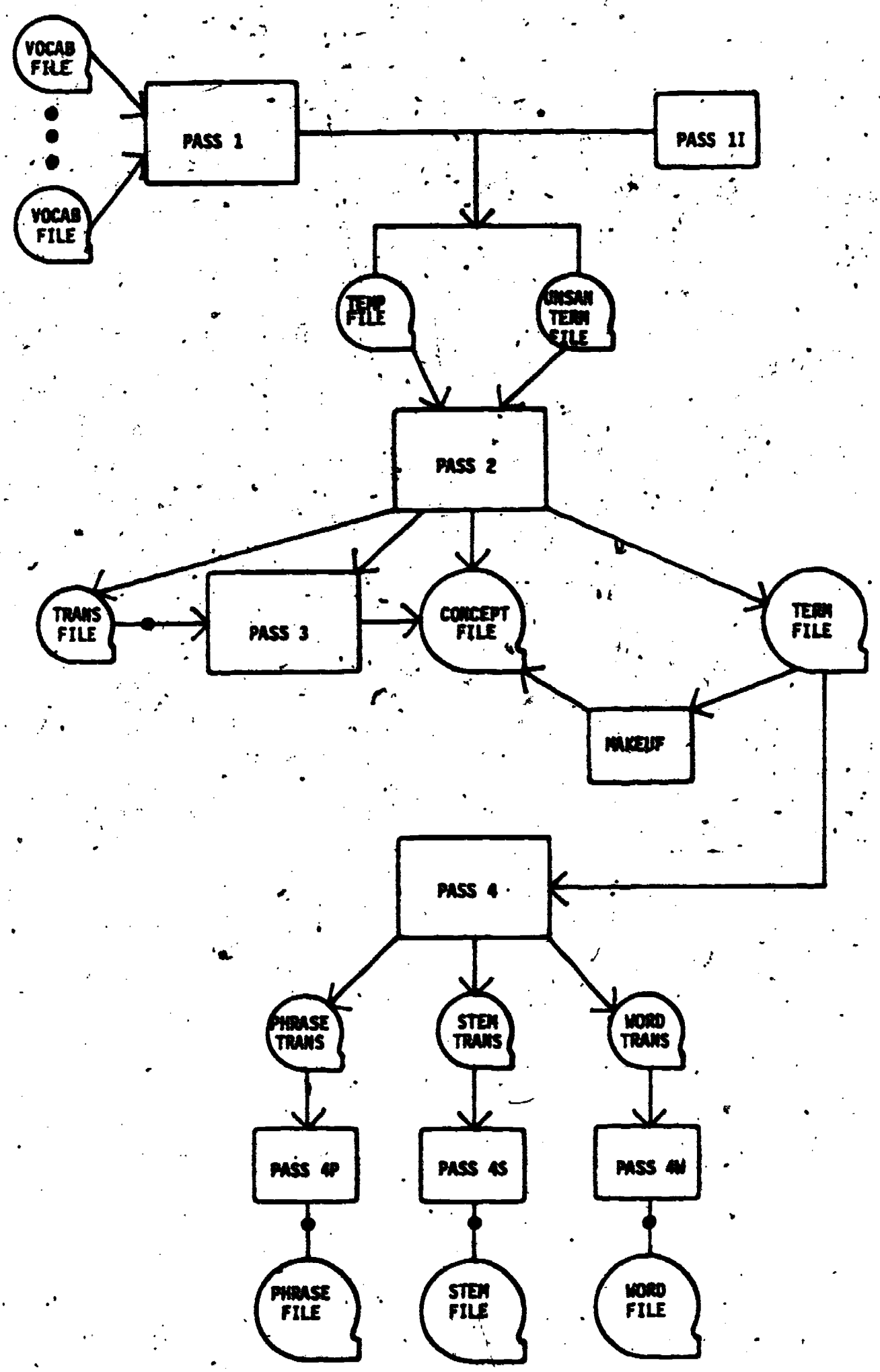
## CONCEPT INFORMATION CELL FORMAT

| Number of Bits | Name | Description |
|---|---|---|
| 6 | Count | Number of words to hold unsanitized term |
| 30 | Filler | |
| 24 | VBT | Vocabulary bit table indicating those vocabularies in which the unsanitized term is found |

TABLE B-1 VSS VOCABULARIES

| Vocabulary Code | Vocabulary Name |
|---|---|
| A | DOE Thesaurus |
| B | ABI Inform |
| C | CA Concept Edit |
| D | ERIC |
| E | SHE Vocabulary (EI) |
| F | Iron Data Center |
| G | Management Contents |
| H | BIOSIS |
| I | INSPEC Thesaurus |
| M | MeSH |
| N | NASA Thesaurus |
| P | Psych Abstracts (APA) |
| R | MeSH' |
| S | MeSH" |
| T | BIOSIS' |

FIGURE B-1 VSS SYSTEM FLOW



# APPENDIX C

## VSS OPERATING INSTRUCTIONS

# VSS INSTRUCTION BOOKLET

March 4, 1983

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

## IMPORTANT INSTRUCTIONS
## PLEASE READ THIS BEFORE USING VSS

Battelle is conducting research under a National Science Foundation grant (ISI-8111497) to evaluate a computer-assisted online search aid called the Vocabulary Switching System (VSS).

Your participation in this study is entirely voluntary. If you decide to participate, your identity will be kept confidential in all reporting activities and/or publications that result from this study. Results will be aggregated and analyzed in such a way as to prevent tracing to any individual or company.

However, we would like to acknowledge all participating organizations by name in a preface in our final report. This is our way of thanking you and letting the readers know just how broad the respondent base was.

Please read and answer the first 10 questions of the VSS Evaluation Form before you commence an actual online session. Also, it is strongly recommended that you read the remainder of the VSS Evaluation Form before login so that you become acquainted with the types of questions to be answered during and after the test period. It is important that Question 11 be completed during the test period after each search is completed. Questions 12 through 17 should be completed at the conclusion of your test period. Return the completed evaluation form in the envelope provided.

You are <u>not</u> confined to running 10 searches on VSS as provided in Question 11. If you can perform more than 10 searches during your test period, please feel free to do so. Just photocopy one of the forms and continue to number your searches.

Your most important guideline is to try to stay within a total of about <u>2 hours of connect time per participant</u> during the assigned test period. You can keep a running total of your own time by noting the connect time when you logout.

Since no formal training is required, just read the login instructions, and other materials enclosed in this packet and begin your evaluation during your assigned week.

The login procedure is tedious; we apologize in advance. If TYMNET responses are slow, it may be due to TYMNET volume or Battelle computer usage. Our computer usage tends to peak at about 10 a.m. and again at 3 p.m. EST. Your response time will be better if you avoid these peak periods. Occasionally you can improve response time by dialing us direct (614) 424-5450. However, the cost of direct dial will be billed to you, whereas a TYMNET call will not.

If you have any trouble with login or a system glitch, call us collect on one of our two HOT LINE numbers:

(614) 424-7843 (Bob Niehoff)
(614) 424-6386 (Helen Pestel)

Your packet contains:

- Cover letter
- Important Instructions
- VSS Login Instructions
- VSS Switching Features
  - VSS Term Types
- VSS Evaluation Forms
- Examples of Subject Switching
- Return Envelope.

Please return:

- Evaluation form(s) as soon as possible after your test period.
- Printouts of your VSS online session if you used a print terminal during the evaluation period (optional).

VSS uses Battelle's BASIS, a data management system for information and data storage and retrieval.

# VSS LOGIN INSTRUCTIONS

(1) Set terminal or modem to half duplex unless you have DEC equipment. For DEC equipment, set "local" mode to on.

(2) Dial a TYMNET number in your area.

Please Type Your Terminal Identifier *

(3) You Enter: your terminal identifier from table below:

TABLE 1. TERMINAL IDENTIFIERS FOR USE WITH TYMNET

| Identifier | Code | Speed | Terminal Type |
|---|---|---|---|
| A | ASCII | 30cps, 120cps | CRT terminals, Personal computers |
| B | ASCII | 15cps | All terminals |
| C | ASCII | 30cps | Impact printers |
| D | ASCII | 10cps | All terminals |
| E | ASCII | 30cps | Thermal printers |
| F | ASCII | 15cps in, 30cps out | BETA terminals |
| G | ASCII | 30cps, 120cps | Bell printers, G.E. Terminet |
| I | ASCII | 120cps | Matrix printers |
| P [cr] | EBCD/ correspondence | 14.8cps | Selectric-type terminals (e.g., 2741) |

Note: A carriage return is keyed only with the P identifier.

Please Log In:

(4) You Enter: BATTELLE [CR]**

p nn
BCL @ Is Online
Destination?

* [ ] System response.

** [CR] = Carriage return

(5) You Enter: CDC

Pause

(6) You Enter: [CR]'s until computer responds

Please Login

(7) You Enter: LOGIN,_____,NSF83,SUP,N

Session charged to...
....
Command-

(8) You Enter: ETL,7777

Command-

(9) You Enter: BASIS,RUN,VSS

Welcome to VSS...
VSS contains

(10) To logout, select END THIS SESSION from the appropriate VSS menu

Command-

(11) You Enter: LOGOUT

**DO NOT**

Request broader or narrower terms (BT/NT) for Business or Life-Science vocabularies; these vocabularies do not use BT/NT relationships in their files.

Request Item 6 from the menu of switching options unless you are prepared to read the attachment to learn how to set up your own switching strategy. Even then, this option may be confusing.

**DO**

Use the browse feature frequently because it has the most switching capability of all the options. However, it does not have BT/NT.

Send us a printout of your online sessions if you use a print terminal.

Set the number of terms to be displayed per vocabulary to relatively high values (5 to 10) to generate sufficient output.

## THINGS YOU SHOULD KNOW

(1) You cannot perform switching across vocabulary sets simultaneously. Instead, you must work with one set at a time. If you want to perform switching in another set, select END THIS SESSION from the appropriate VSS menu. When the system prompts you with COMMAND-, if you enter BASIS,RUN,VSS instead of LOGOUT, you will be given the opportunity to select another vocabulary set defined by VSS without going through the complete login sequence.

However, you can change vocabularies within Life and Physical Science sets by selecting this option from the appropriate VSS menu.

(2) The thesauri used in VSS are now several years old.

(3) VSS contains the following vocabularies:

Business

ABI - ABI Inform Thesaurus

Mgmt. Contents - Management Contents Thesaurus

Social Science

ERIC - ERIC Thesaurus

Psych Abs - Psychological Abstracts Thesaurus

Life Science

BIOSIS - The Master Index Authority File (MIAF) from BIOSIS. In the VSS version of this vocabulary, BC (Biological Codes) and CC (Concept Codes) are invalid.

BIOSIS-C - A special version of BIOSIS in which CC (concept codes) are valid but the concepts are invalid.

MeSH - The medical subject headings of the MEDLINE system used at the National Library of Medicine.

MeSH-R - A special subset of MeSH in which substance names are invalid but their equivalent chemical registry numbers are valid.

MeSH-S - The inverse of MeSH-R

CA - Chemical Abstracts Concept Edit File

Physical Science

CA - Chemical Abstracts Concept Edit File

DOE - Department of Energy Thesaurus

EI - Subject Headings for Engineering, from Engineering Information Inc. (formerly Engineering Index)

INSPEC - INSPEC Thesaurus

IRON - A trilingual thesaurus on iron metallurgy. Languages include English, Spanish, and Portuguese.

NASA - NASA Thesuarus

**VSS VOCABULARIES**

| Vocabularies | Thesaurus Relationship | | | | |
|---|---|---|---|---|---|
| | Lead Term | Use | Special Scope Notes | BT/ NT | RT |
| **Businness** | | | | | |
| ABI Inform | 9* | 1 | 0 | 0 | 6 |
| Mgmt. Contents | 1 | 0 | 0 | 0 | 0 |
| **Social Science** | | | | | |
| ERIC | 8 | 3 | 0 | 5 | 36 |
| Psy. Abs. | 6 | 2 | 0 | 6 | 7 |
| **Life Science** | | | | | |
| BIOSIS | 17 | 7 | 4 | 0 | 9 |
| BIOSIS-C | 4 | 0 | 0 | 0 | 0 |
| CA | 14 | 3 | 0 | 0 | 0 |
| MeSH | 74 | 36 | 63 | 0 | 1 |
| MeSH-R | 42 | 28 | 0 | 0 | 0 |
| MeSH-S | 52 | 31 | 0 | 0 | 0 |
| **Physical Science** | | | | | |
| CA | 14 | 3 | 0 | 0 | 0 |
| DOE | 22 | 5 | 0 | 30 | 27 |
| EI | 12 | 3 | <1 | 0 | 0 |
| INSPEC | 9 | 4 | 0 | 5 | 15 |
| IRON | 17 | 4 | 0 | 16 | 5 |
| NASA | 16 | 19 | <1 | 78 | 132 |

* Number of terms in thousands.

(4) The prompt symbol in BASIS is "?".

(5) The inverted file record length is 50 characters, so if a VSS vocabulary entry is truncated, consult the printed vocabulary to identify the complete term.

(6) If ORED appears at the end of a VSS term, this means that the terms preceeding it should be OR'ed together in Boolean fashion.

(7) If .AND. appears anywhere in a VSS term, the terms should be AND'ed in Boolean fashion. An AND without the periods is a grammatical AND, not a Boolean AND.

## VSS Switching Features*

### 1. Synonyms

This feature invokes a look-up across all vocabularies within a VSS module for occurrences of:

(a) Your term
(b) Synonyms
(c) Corelated

Synonyms are defined as all USE, USED FOR, SEE, and SEEN FROM cross references associated with your term. Co-related terms are those terms which are related by virtue of a common ancestor, e.g., B is co-related to C if the following thesaurus construction is identified by VSS:

Vocabulary I: A use B
Vocabulary II: A use C

Two labels are used at output to identify successful switching; YOUR TERM, and SYNONYM+, shown under the column heading "Term Type". SYNONYM+ is used to label the class of output consisting of both synonyms and co-related terms.

### 2. Browse

This feature automatically invokes many types of VSS switching options in a pre-defined order. The options, called term types at output, used by Browse and their order are:

| Term Type |
|---|
| your term |
| synonym+ |
| rel phras |
| related |
| wd match |
| stm match |
| adj-lead |
| adj-word |

* The amount of switching that takes place within any VSS feature is limited by the number of terms to be displayed per vocabulary. Therefore, a switching feature like BROWSE may never fully execute all of its pre-defined options if the number of terms to be displayed per vocabulary is set too low.

Refer to the next section for further definition of term types (or switching options. If VSS does not perform all the switching defined above, increase the number of terms to be displayed per vocabulary and resubmit your search term. Iterate this process until you are satisfied or VSS simply cannot produce more output.

### 3. Narrower Terms

This feature invokes a search of VSS files for all narrower terms (NT) associated with "your term".

### 4. Broader Terms

This feature invokes a search of VSS files for all broader terms (BT) associated with "your term".

### 5. Narrower/Broader Terms

This feature combines options 3 and 4, in that order.

### 6. Other (User-Defined)

This feature allows you to build your own switching options in any combination and order desired. YOU SHOULD CONSULT THE "HELP" COMMAND TO BECOME ACQUAINTED WITH VSS COMMANDS BEFORE USING THIS FEATURE. SIMPLY ENTER HELP.

When you select this feature, you will be given a menu of VSS switching options in "system-ese".

The "system-ese" menu appears as follows:

```
AVAILABLE SWITCHING OPTIONS INCLUDE:
05 - TERM FILE ACCESS - ACCEPT LEAD + SWITCHABLE TERMS
06 - STEM PHRASE FILE - ACCEPT LEAD + SWITCHABLE TERMS
07 - WORD FILE ACCESS - (SETPCT) MATCH REQUIRED
08 - STEM FILE ACCESS - (SETPCT) MATCH REQUIRED
09 - CONCEPT FILE ACCESS - ACCEPT RELATED TERMS ONLY
10 - CONCEPT FILE ACCESS - ACCEPT NARROW TERMS ONLY
11 - CONCEPT FILE ACCESS - ACCEPT BROADER TERMS ONLY
12 - TERM FILE ACCESS - LIST (2*SETADJ) ADJACENT TERMS
13 - WORD FILE ACCESS - LIST (2* SETADJ) ADJACENT WORDS
19 - ACCEPT LEAD + MULTIPLE USE + MULTIPLE USED-FOR
21 - RELATED TERMS
```

Term types associated with these options are:

| Switching Options | Term Type |
|---|---|
| 05 | term+syn |
| 06 | rel phrase |
| 07 | wd match |
| 08 | stm match |
| 09 | corelated |
| 10 | narrow |
| 11 | broad |
| 12 | adj-lead |
| 13 | adj-word |
| 19 | synonym |
| 21 | related |

To use this feature, simply enter the number corresponding to each option desired, one number per prompt (the prompt symbol is ? in this computer system) until you have entered all selections desired. The selection process is concluded with the word STORE. A typical session might look like this:

```
                ? 5
                ? 19
                ? 9
                ? 10
                ? 7
                ?STORE
Enter a search term for command
                SETPCT
Enter an integer value
                ? 66
Enter a search term for command
                VOCCNT
Enter an integer value
                ? 5
```

## VSS Term Types

(1) adj lead stands for adjacent term. Adjacent term is simply a browse of the inverted file for those terms which are adjacent to the user-entered term. The number of adjacent terms displayed before and after the user-entered term (the window) is 7 by default. If larger or narrower windows are desired, the user must define them via the SETADJ command in VSS. The actual window width displayed is also influenced by the number of terms to be displayed per vocabulary, because the vocabulary count specified by the user takes precedence. This label is used in conjunction with switching option 12 in the user-defined mode of VSS.

(2) adj-word stands for adjacent word. Adjacent word is identical to adjacent-term except that individual words are extracted from the user-entered term and used to browse the VSS inverted word file. This label is used in conjunction with switching option 13 in the user-defined mode of VSS.

(3) broad stands for broader terms (BT). This lable is used to identify all the broader terms associated with "your term". This option works in a limited sense in the DOE Thesaurus; it only produces broader terms at the BT 1 level from the DOE Thesaurus. This label corresponds to switching option 11 in the user defined mode of VSS.

(4) corelated stands for the co-related term. This label identifies those terms which are related by virtue of a common ancestor, e.g., B is co-related to C if the following thesaurus construction is identified by VSS:

Vocabulary I: A use B
Vocabulary II: A use C

This label is used in conjunction with switching option 09 in the user-defined mode of VSS. Note that when a co-related relationship is identified in the Browse feature of VSS the output is simply

labeled synonyms. This labeling anomaly is due to the fact that under the Browse feature, co-related (switching option 09) is embedded in the synonym option. However, in the user-defined mode, option 09 can be selected individually, hence the output can be labeled "corelated".

(5) narrow stands for narrower terms (NT). This label is used to identify all the narrower terms associated with "your term". This option works in a limited sense in the DOE Thesaurus; it only produces narrower terms at the NT 1 level from the DOE Thesaurus. This label corresponds to switching option 10 in the user-defined mode of VSS.

(6) related stands for a related term. Related term is equivalent to the "RT" or related term relationship found in many thesauri. This label is used in conjunction with switching option 21 in the user-mode of VSS.

(7) rel phrase stands for related phrase. Related phrase involves a stemming procedure (right truncation only). The stemming procedure is applied to every word in the phrase enter by the user. Individual stems are combined (concatonated) into a string and this string is used to search a file of strings created in an identical way from all the VSS vocabulary entries. "Hits" from the stem file then cause retrieval of the full unstemmed phrase from another file. Thus, the output phrase contains the stem or root of each word used in the input phrase. This label is used in conjunction with switching option 06 in the user-defined mode of VSS.

(8) stm match stands for stem match. Stem match is identical to word match except that stems of words are used instead of complete words. This label is used in conjunction with switching option 08 in the user-defined term mode of VSS.

(9) synonym stands for synonyms. This type of term includes USE, SEE, USED FOR, and SEEN FROM cross references. It represents an exhaustive look for synonyms in both forward (USE) and backward (UF) directions. It does not incorporate co-related terms. (Note the differente between synonym and synonym+). This label is used in conjunction with switching option 19 in the user-defined mode of VSS.

(10) synonym+ stands for synonym plus. Synonym+ includes USE, SEE, USED FOR (UF), SEEN FROM, and co-related terms. In other words, it is an exhaustive look for synonyms in both the forward (USE) and backward (UF) directions. See co-related for a detailed description of the co-related term relationship. This label is used to identify output associated with "synonym" and "browse" features in VSS. This label is not used in the user-defined mode of VSS.

(11) terms + syn combines two term relationships used by VSS, "your term" and "synonym". Thus, a common label is used to designate two types of terms. This label is used for all output associated with switching option 05 in the user-defined mode of VSS. "Synonym" in this case denotes only the USE or SEE cross references found in various thesauri. This synonym designation is much more limited than either (09 or 10).

(12) wd match stands for word match. In word match, the search term (or phrase) to be switched is broken down into its component words and each word is used to search a VSS word file created in an identical way from all the VSS vocabulary entries. Users must specify how many words of their original term must be matched in order for a switch to be considered successful. The default is 100%. For example, if the search term contains three words and two of the three words must be matched, the user must set the percentage at 66. The system will retrieve all phrases in the file that have at least two of the three words used in the original phrase. However, phrases retrieved by VSS may contain more total words than the original term or phrase entered by the user. In other words,

the percentage specified by the user is applied to the input phrase, not the output phrase. This label is used in conjunction with switching option 07 in the user-defined mode of VSS.

(13) <u>your term</u> is self-explanatory.

# EXAMPLES OF SUBJECT SWITCHING

## from the

## VOCABULARY SWITCHING SYSTEM

```
WELCOME TO *VSS* - VOCABULARY SWITCHING SYSTEM
USING BATTELLE'S DATA MANAGEMENT SYSTEM, BASIS

VSS CONTAINS FOUR VOCABULARY SETS
 1- BUSINESS
    A. ABI          B. MANAGEMENT CONTENTS
 2- BEHAVIOR SCIENCE
    A. ERIC         B. PSYCH ABSTRACTS
 3- LIFE SCIENCE
    A. BIOSIS       B. CHEM ABSTRACTS
    C. MESH
 4- PHYSICAL SCIENCE
    A. DOE          B. CHEM ABSTRACTS
    C. EI           D. INSPEC
    E. IRON         F. NASA


PLEASE SELECT 1 OF THE 4 VOCABULARY SETS
  BY ENTERING EITHER 1, 2, 3, OR 4
     ?1

REQUESTED FILE SET IS ONLINE - CONTINUE

VSS PROVIDES FOR 6 SWITCHING OPTIONS:
  1- SYNONYMS
  2- BROWSE
  3- NARROWER TERMS
  4- BROADER TERMS
  5- NARROWER/BROADER TERMS
  6- OTHER(USER-DEFINED)

PLEASE SELECT 1 OF THE 6 OPTIONS
  BY ENTERING EITHER 1, 2, 3, 4, 5, OR 6.
     ?2

SPECIFY THE MAXIMUM NUMBER OF TERMS
 TO BE DISPLAYED PER VOCABULARY,
 ENTER A NUMBER.
     ?10

ENTER EITHER 1 OR 2 OR 3 OR TERM OR COMMAND.
     ?UNEMPLOYMENT

SWITCH SUCCESSFUL

TERM TYPE     VOCAB     TERM
YOUR TERM     ABI       UNEMPLOYMENT
YOUR TERM     MGMT C    UNEMPLOYMENT
REL PHRAS     ABI       UNEMPLOYABLES
RELATED       ABI       DISGUISED UNEMPLOYMENT
RELATED       ABI       EMPLOYMENT
RELATED       MGMT C    EMPLOYMENT
RELATED       ABI       STRUCTURAL UNEMPLOYMENT
RELATED       ABI       TECHNOLOGICAL UNEMPLOYMENT
RELATED       ABI       UNDEREMPLOYMENT
WD MATCH      ABI       UNEMPLOYMENT BENEFITS
WD MATCH      MGMT C    UNEMPLOYMENT BENEFITS
WD MATCH      ABI       UNEMPLOYMENT COMPENSATION
WD MATCH      ABI       TEENAGE UNEMPLOYMENT
```

WELCOME TO VSS - VOCABULARY SWITCHING SYSTEM
OVER BATTELLE'S DATA MANAGEMENT SYSTEM, BASIS

VSS CONTAINS FOUR VOCABULARY SETS
1- BUSINESS
   A. ABI          B. MANAGEMENT CONTENTS
2- BEHAVIOR SCIENCE
   A. ERIC         B. PSYCH ABSTRACTS
3- LIFE SCIENCE
   A. BIOSIS       B. CHEM ABSTRACTS
   C. MESH
4- PHYSICAL SCIENCE
   A. DOE          B. CHEM ABSTRACTS
   C. EI           D. INSPEC
   E. IRON         F. NASA

PLEASE SELECT 1 OF THE 4 VOCABULARY SETS
  BY ENTERING EITHER 1, 2, 3, OR 4
    ?2

REQUESTED FILE SET IS ONLINE - CONTINUE

VSS PROVIDES FOR 6 SWITCHING OPTIONS:
  1- SYNONYMS
  2- BROWSE
  3- NARROWER TERMS
  4- BROADER TERMS
  5- NARROWER/BROADER TERMS
  6- OTHER(USER-DEFINED)

PLEASE SELECT 1 OF THE 6 OPTIONS
  BY ENTERING EITHER 1, 2, 3, 4, 5, OR 6.
    ?2

SPECIFY THE MAXIMUM NUMBER OF TERMS
 TO BE DISPLAYED PER VOCABULARY.
 ENTER A NUMBER.
    ?25

ENTER EITHER 1 OR 2 OR 3 OR TERM OR COMMAND.
    ?EMOTIONALLY DISTURBED

SWITCH SUCCESSFUL

| TERM TYPE | VOCAB | TERM |
|---|---|---|
| YOUR TERM | ERIC | EMOTIONALLY DISTURBED |
| YOUR TERM | PSYCH | EMOTIONALLY DISTURBED |
| RELATED | PSYCH | ACTING OUT |
| RELATED | PSYCH | CHILDHOOD PSYCHOSIS |
| RELATED | ERIC | EMOTIONAL ADJUSTMENT |
| RELATED | PSYCH | EMOTIONAL ADJUSTMENT |
| RELATED | PSYCH | EMOTIONAL STATES |
| RELATED | PSYCH | EMOTIONS |
| RELATED | ERIC | ANXIETY |
| RELATED | PSYCH | ANXIETY |
| RELATED | ERIC | AUTISM |
| RELATED | PSYCH | AUTISM |
| RELATED | ERIC | BEHAVIOR |
| RELATED | ERIC | BEHAVIOR PROBLEMS |
| RELATED | PSYCH | BEHAVIOR PROBLEMS |
| RELATED | ERIC | CLINICAL PSYCHOLOGY |
| RELATED | PSYCH | CLINICAL PSYCHOLOGY |
| RELATED | ERIC | DEPRESSION (PSYCHOLOGY) |
| RELATED | ERIC | EMOTIONAL MALADJUSTMENT |
| RELATED | ERIC | EMOTIONAL PROBLEMS |
| RELATED | ERIC | HYPERACTIVITY |
| RELATED | ERIC | MENTAL ILLNESS |
| RELATED | ERIC | PERSONALITY PROBLEMS |
| RELATED | ERIC | PSYCHIATRISTS |
| RELATED | PSYCH | PSYCHIATRISTS |
| RELATED | ERIC | PSYCHIATRY |
| RELATED | PSYCH | PSYCHIATRY |
| RELATED | ERIC | PSYCHOPATHOLOGY |
| RELATED | PSYCH | PSYCHOPATHOLOGY |
| RELATED | ERIC | PSYCHOTHERAPY |
| RELATED | PSYCH | PSYCHOTHERAPY |
| RELATED | ERIC | SCHIZOPHRENIA |
| RELATED | PSYCH | SCHIZOPHRENIA |
| RELATED | ERIC | SELF MUTILATION |
| RELATED | PSYCH | SELF MUTILATION |
| RELATED | ERIC | SPECIAL HEALTH PROBLEMS |
| NO MATCH | ERIC | EMOTIONALLY DISTURBED CHILDREN |
| STM MATCH | PSYCH | EMOTIONALLY DISTURBED .AND. SPECIAL EDUCATION |
| ADJ-LEAD | PSYCH | EMOTIONAL TRAUMA |
| ADJ-LEAD | PSYCH | EMOTIONALITY (PERSONALITY) |

```
VSS CONTAINS FOUR VOCABULARY SETS
 1- BUSINESS
    A. ABI              B. MANAGEMENT CONTENTS
 2- BEHAVIOR SCIENCE
    A. ERIC             B. PSYCH ABSTRACTS
 3- LIFE SCIENCE
    A. BIOSIS           B. CHEM ABSTRACTS
    C. MESH
 4- PHYSICAL SCIENCE
    A. DOE              B. CHEM ABSTRACTS
    C. EI               D. INSPEC
    E. IRON             F. NASA


PLEASE SELECT 1 OF THE 4 VOCABULARY SETS
  BY ENTERING EITHER 1, 2, 3, OR 4
      ?3

VSS PROVIDES FOR 6 SWITCHING OPTIONS:
  1- SYNONYMS
  2- BROWSE
  3- NARROWER TERMS
  4- BROADER TERMS
  5- NARROWER/BROADER TERMS
  6- OTHER(USER-DEFINED)

PLEASE SELECT 1 OF THE 6 OPTIONS
  BY ENTERING EITHER 1, 2, 3, 4, 5, OR 6.
      ?2

SPECIFY THE MAXIMUM NUMBER OF TERMS
 TO BE DISPLAYED PER VOCABULARY.
 ENTER A NUMBER.
      ?7

ENTER EITHER 1 OR 2 OR 3 OR TERM OR COMMAND.
      ?VIRUSES

SWITCH SUCCESSFUL

TERM TYPE     VOCAB     TERM
YOUR TERM     CHEM A    VIRUSES
YOUR TERM     MESH      VIRUSES
SYNONYM +     CHEM A    VIRUS
SYNONYM +     BIOSIS    VIRUS
SYNONYM +     BIOSIS    USE VIRAL  VIRUS  OR
SYNONYM +     BIOSIS    VIRAL
RELATED       MESH      INCLUSION BODIES, VIRAL
WD MATCH      BIOSIS    VIROLOGY - PLANT HOST VIRUSES
WD MATCH      BIOSIS    VIROLOGY - ANIMAL HOST VIRUSES
WD MATCH      BIOSIS    PHYTOPATHOLOGY - DISEASES CAUSED BY
WD MATCH      MESH      SCRAPIE (68-72)
WD MATCH      MESH      PULMONARY ADENOMATOSIS, OVINE (68-72
WD MATCH      MESH      VERTEBRATE VIRUSES, UNCLASSIFIED
WD MATCH      MESH      C-TYPE VIRUSES (72-77)
WD MATCH      MESH      DEFECTIVE VIRUSES (67-74)
WD MATCH      BIOSIS    CHLAMYDIACEAE
WD MATCH      BIOSIS-C  CC34510
WD MATCH      BIOSIS-C  CC33508
WD MATCH      BIOSIS-C  CC33506
STM MATCH     MESH-S    VIRUS-NEUTRALIZING FACTOR
STM MATCH     BIOSIS-C  CC34504
```

```
WELCOME TO *VSS* - VOCABULARY SWITCHING SYSTEM
USING BATTELLE'S DATA MANAGEMENT SYSTEM, BASIS

VSS CONTAINS FOUR VOCABULARY SETS
 1- BUSINESS
    A. ANI              B. MANAGEMENT CONTENTS
 2- BEHAVIOR SCIENCE
    A. NEIC             B. PSYCH ABSTRACTS
 3- LIFE SCIENCE
    A. BIOSIS           B. CHEM ABSTRACTS
    C. MESH
 4- PHYSICAL SCIENCE
    A. DOE              B. CHEM ABSTRACTS
    C. EI               D. INSPEC
    E. IRON             F. NASA


PLEASE SELECT 1 OF THE 4 VOCABULARY SETS
  BY ENTERING EITHER 1, 2, 3, OR 4
     ?4

VSS PROVIDES FOR 6 SWITCHING OPTIONS:
  1- SYNONYMS
  2- BROWSE
  3- NARROWER TERMS
  4- BROADER TERMS
  5- NARROWER/BROADER TERMS
  6- OTHER(USER-DEFINED)

PLEASE SELECT 1 OF THE 6 OPTIONS
  BY ENTERING EITHER 1, 2, 3, 4, 5, OR 6.
     ?2

SPECIFY THE MAXIMUM NUMBER OF TERMS
 TO BE DISPLAYED PER VOCABULARY.
 ENTER A NUMBER.
     ?10

PLEASE ENTER A SINGLE SEARCH TERM OR COMMAND

     ?HEAVY WATER

SWITCH SUCCESSFUL

TERM TYPE     VOCAB     TERM
YOUR TERM     DOE       HEAVY WATER
YOUR TERM     EI        HEAVY WATER
YOUR TERM     INSPEC    HEAVY WATER
YOUR TERM     NASA      HEAVY WATER
RELATED       DOE       COOLANTS
RELATED       NASA      COOLANTS
RELATED       DOE       DEUTERIUM COMPOUNDS
RELATED       EI        DEUTERIUM COMPOUNDS
RELATED       INSPEC    DEUTERIUM COMPOUNDS
RELATED       NASA      DEUTERIUM COMPOUNDS
RELATED       DOE       MODERATORS
RELATED       CHEM A    MODERATORS
RELATED       INSPEC    MODERATORS
RELATED       NASA      MODERATORS
RELATED       DOE       TRITIUM COMPOUNDS
RELATED       INSPEC    TRITIUM COMPOUNDS
RELATED       DOE       DUAL TEMPERATURE PROCESS
RELATED       DOE       DEUTERIUM
RELATED       EI        DEUTERIUM
RELATED       INSPEC    DEUTERIUM
RELATED       NASA      DEUTERIUM
RELATED       DOE       TRITIUM
RELATED       EI        TRITIUM
RELATED       INSPEC    TRITIUM
RELATED       NASA      TRITIUM
RELATED       INSPEC    FISSION REACTOR MATERIALS
WD MATCH      DOE       HEAVY WATER PLANTS
WD MATCH      DOE       SGHWR REACTOR
```

# APPENDIX D

## TERM SETS FOR SWITCHING STRATEGY ANALYSES

# BUSINESS TERMS

1 ACCREDITATION
2 ACQUISITIONS
3 ADVANTAGES
4 ADVERTISING RATES
5 ALCOHOLISM
6 ANXIETIES
7 APPLIANCES
8 ASSOCIATIONS
9 AUTOMOBILE ACCESSORIES
10 BANKS
11 BYTE N DATA
12 BUDGET
13 BUSINESS SYSTEM PLANNING
14 BUY SELL AGREEMENTS
15 BUYING
16 CARPENTERS UNION
17 CASHIER
18 CERTIFICATION
19 CHARTS
20 CIGARETTE INDUSTRY
21 CIGARS
22 CLOSELY HELD CORPORATION
23 CLOTHING INDUSTRY
24 COMMERCIAL N INTELLIGENCE
25 COMMUNICATION
26 COMPETITION
27 COMPUTERS
28 CONFLICT
29 CONSULTANTS
30 CONSUMER COOPERATIVES
31 CONVEYORS
32 COPYING MACHINES
33 CORPORATE PLANNING
34 CORRUPTION
35 COSMETICS
36 COST INDEX
37 COUNSELING
38 CREDIT UNIONS
39 CUSTOMERS
40 DECISION TREE
41 DECISION MAKING
42 DIMENSION
43 DISADVANTAGES
44 DOLLAR VALUE
45 ECONOMIC FORECASTING
46 ECONOMIC INDICATORS
47 ECONOMIC TRENDS
48 ECONOMY
49 ELECTRIC POWER
50 ELECTRIC TRANSMISSION LINES
51 EMPLOYEE PROBLEMS
52 EMPLOYEE RECREATION PROGRAMS
53 EMPLOYEE STOCK OPTION PLANS
54 EMPLOYEE TURNOVER
55 EMPLOYMENT
56 ESCALATOR CLAUSES
57 EXERCISE
58 FAST FOOD INDUSTRY
59 FEDERAL GOVERNMENT
60 FLOOR COVERING
61 FOOTWEAR
62 FORTUNE
63 FURNITURE
64 GOVERNMENT CONTRACTS
65 GRAPHS
66 GROUND LEASES
67 GROUP PARTICIPATION
68 HARDWARE
69 HEART DISEASE
70 HOTELS
71 INDUSTRIAL PARKS
72 INDUSTRIAL USAGE
73 INFORMATION
74 INPUT OUTPUT
75 INTANGIBLES TAX
76 INTERACTION
77 INTERINDUSTRY ECONOMICS
78 INTERPERSONAL
79 JEWELRY
80 JOHARI METHOD
81 LABOR UNIONS
82 LABOR MANAGEMENT RELATIONS
83 LEASES
84 LIBRARIES
85 LUMBER INDUSTRY
86 MAN MACHINE SYSTEMS
87 MANAGEMENT INFORMATION SYSTEMS
88 MANPOWER
89 MARKETING
90 MERGERS
91 MOVIES
92 MORTGAGE
93 MOTELS
94 MOTOR VEHICLES
95 MEN EMPLOYEES
96 NUTRITION
97 OBJECTIVES
98 OFFICE BUILDING
99 ONLINE SYSTEMS
100 ORGANIZATIONAL STRUCTURE
101 PREFERRED
102 PRESCRIPTION DRUGS
103 PREVENTIVE MEDICINE
104 PROGRAMMED INSTRUCTION
105 PROGRESSIVE INSURANCE
106 QUALITY
107 RADIO ADVERTISING
108 REAL ESTATE APPRAISAL
109 RECORDS N RETENTION
110 RECREATION VEHICLES
111 RESEARCH AND DEVELOPMENT
112 RESUMES
113 RETAIL ADVERTISING
114 RETAIL STORES
115 RETAILING TRENDS
116 RISK MANAGEMENT
117 SALARY
118 SALES
119 SALES TAX
120 SAVINGS AND LOAN
121 SCIENCE PARKS
122 SELLING
123 SIC
124 SMALL BUSINESS
125 SOFTWARE PACKAGES
126 SPORTING GOODS
127 SQUARE FOOTAGE
128 STOCK GIFTS
129 STRESS
130 TAX PREPARATION
131 TEAM WORK
132 TEMPORARY
133 TIRE INDUSTRY
134 TOBACCO
135 TRAINING
136 VIDEO DISPLAY SYSTEMS
137 VIDEOCASSETTE
138 VIDEODISC
139 VIDEOTAPE
140 VIEWTEL
141 VIP CLUBS
142 VOTING TRUST
143 XEROGRAPHIC COPIERS

O-terms randomly selected for estimation of switching strategy performance.

# SOCIAL SCIENCE TERMS

1 ABUSE
2 ADOLESCENT
3 AESTHETICS
4 ALCOHOLISM
5 ANIMAL BEHAVIOR
6 ANIMALS
7 APPARATUS
8 ATTITUDES
9 BABOONS
10 BATTERED WOMAN
11 BEHAVIOR PROBLEMS
12 BIRTH ORDER
13 BOARD OF EDUCATION
14 CEREBROVASCULAR ACCIDENTS
15 CLASSROOM
16 CLERICAL PERSONNEL
17 COMMUNICATION
18 CONFLICT RESOLUTION
19 CONTINUING EDUCATION
20 CORPORAL PUNISHMENT
21 COUNSELING
22 CURRICULUM ENRICHMENT
23 DATA PROCESSING
24 DELINQUENCY
25 DISABILITIES
26 DISCIPLINE
27 DISCIPLINE POLICY
28 EARLY CHILDHOOD
29 EDUCATIONAL PLACEMENT
30 EDUCATIONAL RESEARCH
31 EMOTIONALLY DISABLED
32 EMPATHY
33 ENVIRONMENT
34 EXPULSION
35 EXTRA CURRICULAR ACTIVITIES
36 FEDERAL COURTS
37 GENETICS
38 GROUP DYNAMICS
39 HANDICAPS
40 HEREDITY
41 INSERVICE EDUCATION
42 INSTRUCTION
43 INSTRUCTIONAL DESIGN
44 INSTRUCTIONAL MATERIALS
45 INTELLIGENCE
46 INTELLIGENCE TESTS
47 INTERACTION
48 INTERMEDIATE GRADES
49 INTERPERSONAL RELATIONS
50 INTERVENING
51 JOB ENRICHMENT
52 KINDERGARTEN
53 LEARNING
54 LIBRARY ACQUISITIONS
55 LIBRARY AUTOMATION
56 LIBRARY COLLECTIONS
57 LIBRARY INSTRUCTION
58 LIBRARY PERSONNEL
59 LEISURE TIME
60 MEDIA SELECTION
61 MENTAL RETARDATION
62 METACOGNITION
63 MINORITY CHILDREN
64 MONKEYS
65 MOTHER CHILD RELATIONSHIPS
66 MOTIVATION
67 MULTIFACTORED EVALUATION
68 MUSIC EDUCATION
69 MUSIC THERAPY
70 NEED SATISFACTION
71 NEWSPAPERS
72 NONBIASED TESTING
73 NONPRINT MEDIA
74 ONLINE SYSTEMS
75 OUTDOOR EDUCATION
76 PARENT ATTITUDES
77 PARENT ATTITUDES
78 PARENT SCHOOL RELATIONSHIP
79 PERIODICALS
80 PLAY THERAPY
81 PRIMARY EDUCATION
82 PRIMATES
83 PROBLEM CHILDREN
84 PROBLEM SOLVING
85 PROGRAM EDUCATION
86 PROGRAM EFFECTIVENESS
87 PSYCHOLOGIST
88 READING ACHIEVEMENT
89 RESEARCH LIBRARIES
90 RESEARCH METHODOLOGY
91 SCHOOLPSYCHOLOGIST
92 SELF PERCEPTION
93 ROLE THEORY
94 SCHOLARLY JOURNALS
95 SCHOOL POLICY
96 SHOPLIFTING
97 SIMULATION
98 SOCIAL ADJUSTMENT
99 SOCIAL PROBLEMS
100 SOCIOECONOMIC
101 SPECIAL EDUCATION
102 SPOUSE
103 STUDENT ORGANIZATIONS
104 STUDENT PLACEMENT
105 STUDENT RIGHTS
106 SUSPENSION
107 SYSTEMS APPROACH
108 TEACHER ATTITUDES
109 TEACHING METHODS
110 TEXTBOOK SELECTION
111 TREATMENT
112 UNIVERSITIES
113 WORK ATTITUDES

# LIFE SCIENCE TERMS

1 ACCIDENT PREVENTION
2 ADRENAL GLANDS
3 ALCOHOL
4 AMBULATION
5 APIDES
6 AMNIOCENTESIS
7 AMOXAPINE
8 ANCRAPINE
9 AMPICILLIN
10 ANALGESICS
11 ANESTHESIA
12 ANTIGENS
13 ANTIHISTAMINES
14 ANTIOXIDANT
15 ASCORBIC ACID
16 ASPIRATION
17 ATHEROSCLEROSIS
18 ATRIAL FIBRILLATION
19 ATRIOVENTRICULAR NODE
20 AUTISTIC CHILDREN
21 BARBITURATES
22 BENDECTIN
23 BILE SALTS
24 BIOCHEMISTRY
25 BIOFEED BACK
26 BLEOMYCIN
27 BLOOD BANKS
28 BLOOD GLUCOSE
29 BONE MATRIX
30 BRAIN DAMAGE
31 BRETYLIUM
32 CADMIUM
33 CAFFEINE
34 CAMPHOR
35 CARBORYL
36 CARCINOMA
37 CARDIOVASCULAR DISEASE
38 CAROTID ARTERY
39 CATHARTICS
40 CEFOPERAZONE
41 CEPHALOSPORINS
42 CEREBRAL CORTEX
43 CESAREAN SECTIONS
44 CHEMILUMINESCENCE
45 CHEMOTHERAPY
46 CHLORINE GAS INHALATION
47 CHLOROPROMAZINE
48 CHOLESTEROL
49 CHOLIC ACID
50 CHROMATOGRAPHY
51 CHRONIC LUNG DISEASE
52 CIRCADIAN RHYTHM
53 COGNITIVE STYLE MAPPING
54 COMPLEX CARBOHYDRATES
55 CONTACT INHIBITION
56 CORONARY ARTERY
57 CYCLOPHOSPHAMIDE
58 CYTAPHERESIS
59 DENTISTRY
60 DEXAMETHASONE
61 DIALYSIS
62 DIGITAL ANGIOGRAPHY
63 DIGOXIN
64 DIMETHYL SULFOXIDE
65 DIMETHYL SULFOXIDE
66 DIURETICS
67 DNA
68 DOPAMINE
69 DRUG ABUSE
70 DRUG HYPERSENSITIVITY
71 ELDERLY
72 ELECTRON MICROSCOPY
73 ELECTRO MICROGRAPHS
74 ELECTROCONVULSIVE THERAPY
75 EMULSIONS
76 ENDOCRINOLOGY
77 ENDORPHINS
78 ENDOTHELIAL CELLS
79 ENERGY BALANCE
80 EPIDEMIOLOGY
81 ERGOT ALKALOIDS
82 ETIOLOGY
83 EUCALYPTUS OIL
84 FATTY ACIDS
85 FENESTRATION
86 FLUID ELECTROLYTES
87 FLUID ELECTROLYTES
88 FLUID THERAPY
89 FOLIC ACID
90 GASTRIC CONTENTS
91 GLUCOSE TOLERANCE
92 GLUTEN FREE DIETS
93 GLYOXYLIC ACID
94 GOLD SALTS
95 GUANIDINES
96 HEALTH PLANNING
97 HEMATOLOGY
98 HEPARIN
99 HERPES
100 HISTOFLOURESCENCE
101 HORMONE SECRETION
102 HOSPITAL INFORMATION SYSTEMS
103 HUMAN SEXUALITY
104 HYDRALAZINE
105 HYPERKALEMIA
106 HYPERTENSION
107 HYPNOSIS
108 HYPOGLYCEMIA
109 HYPOVOLEMIC SHOCK
110 HYSTERECTOMIES
111 IMMUNE RESPONSE
112 IMMUNOGLOBULINS
113 INFLUENZA
114 INHALATION THERAPY
115 INSULIN
116 INTERFERON
117 INTRAOCULAR LENS IMPLANT
118 IODINE 125
119 JAUNDICE
120 KINETICS
121 KETONES
122 KIDNEY TRANSPLANT
123 LACTATION
124 LASER USAGE
125 LIDOCAINE
126 LIDOCAINE
127 LIPIDS
128 LIPOPROTEINS
129 LIQUID CRYSTALS
130 LUNG FUNCTION
131 LYMPH NODES
132 MALEIC ACID
133 MASS SPECTROMETRY
134 MECLOFENAMATE SODIUM
135 MENTAL RETARDATION
136 MENTHOL
137 MERCURY POISONING
138 METHYL SALICYLATE
139 MICROCIRCULATION
140 MICROFILAMENT STRUCTURE
141 MICROWAVE RADIATION
142 MIGRAIN
143 MITOMYCIN
144 MONOCYTES
145 NICOTINE
146 MUTAGENS
147 MYELIN SHEATH
148 NAPROXEN SODIUM
149 NARCOTIC
150 NATURAL KILLER CELLS
151 NEUROMUSCULAR DISEASE
152 NEUROTOXICITY
153 NEW BORN VESTIBULAR FUNCTION
154 NITRATES
155 NURSING EDUCATION
156 NUTRITIONAL SERVICE
157 OBSTETRICS
158 OCCUPATIONAL MEDICINE
159 OPEN HEART SURGERY
160 ORTHODONTICS
161 OXOLINIC ACID
162 PACKED RED BLOOD CELL
163 PANCREATITIS
164 PANNICULITIS
165 PARENT CHILD BONDING
166 PARENTERAL INFUSIONS
167 PATIENT EDUCATION
168 PENICILLIN
169 PENTOBARBITO
170 PH
171 PHENOTHIAZINE
172 PHTHALIC ACID
173 PLATELETS
174 PNEUMONIA
175 POLYCARBOPHIL
176 POPULATION PLANNING
177 POST PARTUM
178 POTASSIUM REPLACEMENT
179 POVIDONE IODINE
180 PREGNANCY
181 PROGRESSIVE PATIENT CARE
182 PROPANOLOL
183 PSYCHOSOCIAL
184 QUALITY ASSURANCE
185 QUINOSINE
186 QUINIDINE
187 RADIATION THERAPY
188 RADIOACTIVITY
189 RENAL HYPERTROPHY
190 RESPIRATORY DEFICITS
191 REYE S SYNDROME
192 SCHIZOPHRENIA
193 SENSORY DEPRIVATION
194 SENSORY OVERLOAD
195 SEPTIC SHOCK
196 SIBLING COUNSELING
197 SKIN ERUPTIONS
198 SODIUM TETRADECYL SULFATE
199 SPEECH DEFECTS
200 SPEECH DISORDERS
201 SUBSTANCE P
202 SUGAR LEVELS
203 SUICIDE
204 SURFACTANTS
205 SUSPENSIONS
206 SUTURE INSTRUMENTS
207 SYNAPTOSOMES
208 TOLUENE
209 TOXIC EFFECTS
210 TOXIC SHOCK SYNDROME
211 TRACHEOTOMY
212 TRANSFUSIONS
213 TREATMENT MODALITIES
214 TRIMEPRAZINE
215 TRINITROPHENYL
216 TUBAL LIGATION
217 TUMOR REGISTRIES
218 UMBILICAL VEINS
219 VASOACTIVE INTESTINAL POLY
220 VITAMIN K DEFICIENCY
221 WEIGHT MAINTENANCE
222 WHOLE BLOOD

# PHYSICAL SCIENCE TERMS

1 ABSORBENT
2 ABSORBING
3 ABUSE
4 ACID SYSTEM
5 ACOUSTIC EMISSIONS
6 ACOUSTICS
7 ACTIVATION ANALYSIS
8 ADHESIVE
9 ADHESIVES
10 AEROSOLS
11 AGGREGATE
12 AIR POLLUTION
13 AIRBURST
14 ALL EMISSION
15 ALLOY
16 ALLOYS
17 ALUMINUM
18 ALUMINUM ALLOY
19 AMYL NITRITE
20 ANALYSIS
21 ANALYTICAL
22 ANALYTICAL METHOD
23 ANTIREFLECT
24 ANTITUMORIGEN
25 ARSENOMETRY
26 ARITHMETIC
27 ARMING DEVICES
28 ARTIFICIAL
29 ASH
30 ASHES
31 ASPHALT
32 ATOMIC
33 BAG
34 BERYLLIUM ALLOY
35 BINDER
36 BINDING MATERIALS
37 BIOCIDAL KINETICS
38 BOILING
39 BITUMEN
40 BLADE
41 BLADES
42 BLOCK
43 BLOOD
44 BONDING
45 BONDS
46 BOOMS
47 BOOSTER EXPLOSIVES
48 BRAKE
49 BRAKING
50 BROMINE CHLORIDE
51 BUBBLE REMOVAL
52 BURNER
53 BURNING
54 BUTYL NITRITE
55 CABINET
56 CABLE
57 CABLES
58 CAKE
59 CAKES
60 CALCIUM SULFATE
61 CALCIUM SULFATES
62 CALIBRATION
63 CANCER
64 CAPACITY
65 CAPITAL
66 CARBON ISOTOPES
67 CARCINOGEN
68 CASING
69 CASTING
70 CELLULOSE
71 CEMENT
72 CERAMIC MATERIALS
73 CEREAL
74 CHAMBER
75 CHARACTERISTICS
76 CHARGE DISTRIBUTION
77 CHELATOMETRY
78 CHEMICAL
79 CHEMICAL ANALYSIS
80 CHLORIDE
81 CHLORINATION
82 CHLORINE
83 CHLORINE DIOXIDE
84 CHROMATOGRAPHY
85 CIGARETTE
86 CIGARETTES
87 CIGARS
88 CINDER
89 CINDERS
90 CLAMS
91 CLAY
92 CLEAN
93 CLINKER
94 CLINKERS
95 CLOSURES
96 COATING
97 COGASINOGEN
98 COCKPIT
99 COILING
100 COLLAGENS
101 COLLECTOR
102 COLLISION
103 COMBUSTION
104 COMBUSTION FIRING
105 COMBUSTION PRODUCTS
106 COMBUSTOR
107 COMMUNICATION TERMINALS
108 COMPACTION
109 COMPARISON
110 COMPLIANCE
111 COMPOSITE
112 COMPOST
113 COMPUTATION OPTICAL
114 COMPUTER ARCHITECTURE
115 COMPUTER PROGRAMMING
116 COMPUTERS
117 CONCRETE
118 CONDUCTIVITY
119 CONNECT
120 CONSOLIDATION
121 CONSUMPTION
122 CONTAINER
123 COOLING
124 COPPER
125 CORDAGE
126 CORROSION
127 COST
128 COTTON
129 COVER
130 CRACKING
131 CRANE
132 CRANES
133 CRYSTALLIZATION
134 CURING
135 CYROGENIC
136 DATA PROCESSING TERMINALS
137 DEFORMATION
138 DEFOULING
139 DEGRADATION
140 DELAY ELEMENTS EXPLOSIVES
141 DELOS ELECTRONICS
142 DERAMS
143 DEPOLYMERIZATION
144 DEPOSITION
145 DERRICK
146 DERRICKS
147 DESCRIMINATION
148 DESIGN ANALYSIS
149 DETERMINATION
150 DIELECTRIC PROPERTIES
151 DIELECTRICS
152 DIESEL ENGINE
153 DIGITAL ARITHMETIC
154 DIGITAL COMPUTERS
155 DIRECT CONVERSION
156 DIRECT ENERGY CONVERSION
157 DISPLAY
158 DISPOSAL
159 DOMESTIC
160 DOWNTIME
161 DRAWING FORMING
162 DRILLING
163 DROP SIZE
164 DROPLET
165 DROPLETS
166 DROPS
167 DUCILITY
168 DUDS
169 DUMPING
170 DUMPS
171 DUST
172 DYNAMIC
173 EARTHFILLS
174 ECONOMIC
175 EFFECT
176 ELECTRIC CHARGES
177 ELECTRICAL
178 ELECTRICAL FIELD
179 ELECTRICAL FIELDS
180 ELECTRICAL INSULATION
181 ELECTRON EMISSION
182 ELECTRON EMISSIONS
183 ELECTROSTATIC FIELD
184 ELECTROSTATIC FIELDS
185 ELECTROSTATICS
186 ELEVATOR
187 ELEVATORS
188 EMBEDD
189 EMBRITTLEMENT
190 ENCAPSULAT
191 ENERGY
192 ENERGY RECOVERY
193 ENGINES
194 ENVIRONMENTAL
195 EPON
196 EPR
197 EPT
198 EQUIPMENT
199 ESTIMATION
200 ETCHING
201 ETHANOL
202 ETHENE TETRAFLUORO
203 ETHYL ALCOHOL
204 ETHYLENE
205 EVALUATE
206 EVALUATION
207 EVAPORTRANS
208 EXHAUST
209 EXPLODING WIRE
210 EXPLODING WIRES
211 EXPLOSION
212 EXPLOSION EFFECTS
213 EXPLOSIONS
214 EXPLOSIVE CHARGES
215 EXPLOSIVE TRAINS
216 EXPLOSIVES
217 EXTRACT
218 EXTRACTING
219 EXTRACTION
220 EXTRUSION
221 EXTRUSIONS
222 FABRICATION
223 FAILURE
224 FAILURE MECHANICS
225 FATIGUE
226 FAULTY
227 FEEDTHROUGH
228 FERROELECTRIC
229 FIBRIN
230 FIBRINOGEN
231 FILL
232 FILLER
233 FILM
234 FILTER
235 FILTERS
236 FIRING
237 FIRE
238 FIREBALL
239 FLARE
240 FLAVOR
241 FLAVORING
242 FLAN
243 FLEXIBLE
244 FLIGHT
245 FLIGHT SIMULATE
246 FLOOR
247 FLOW
248 FLUE
249 FLUE GASES
250 FLUE OR STACK
251 FLUIDIZED BED
252 FLUIDIZED BEDS
253 FLYASH
254 FOAM
255 FOOD
256 FORECASTING
257 FORMATION
258 FRACTIONATION
259 FRACTURE PROPERTIES
260 FREEZING
261 FUEL ADDITIVES
262 FUEL CONSUMPTION
263 FUEL ELEMENT
264 FUEL ELEMENTS
265 FUMES
266 FURNACE BURST
267 GAMMA
268 GARBAGE
269 GASOHOL
270 GEAR TEETH
271 GEARS
272 GENERATE
273 GILSONITE
274 GLASS
275 GLASSES
276 GLUE

# PHYSICAL SCIENCE TERMS (Continued)

277 GRANULITE
278 GRAIN
279 GRAIN DUST
280 GRAVITATIONAL EFFECTS
281 GREASE
282 GRENADE FUSES
283 GRINDING
284 GUIDANCE
285 GUIDED
286 GYPSUM
287 HALOGENATION
288 HARDENING
289 HARDNESS
290 HAZARDOUS
291 HEAT
292 HEAT CAPACITY
293 HEAT RECOVERY
294 HEAT RESISTANT PLASTICS
295 HEAT TRANSFER
296 HEATING
297 HEIGHT OF BURST
298 HELICAL
299 HIGH POWERS
300 HIGH TEMPERATURE
301 HITTING
302 HOIST
303 HOISTS
304 HONEYCOMB
305 HOUSEHOLD
306 HUMIDITY
307 HYDROGEN
308 ICEBREAKER
309 IDOMETRY
310 IMPACT
311 IMPLANT
312 IMPURITY
313 INCINERATOR
314 INDEPENDENT POWER SOURCES
315 INDUSTRY
316 INFARED
317 INHIBITOR
318 INORGANIC
319 INSECTICIDE
320 INSULATING
321 INSULATION
322 INTEGRITY
323 INVESTMENT
324 ISOBUTYL NITRITE
325 ITT CORPORATION
326 LANDFILL
327 LANDFILLS
328 LANDING
329 LASER BEAM
330 LAUNCH VEHICLES
331 LAYER
332 LEACH
333 LEAD
334 LEAD POLYMER
335 LEAD POLYMERS
336 LEAK
337 LEAKAGE
338 LEVITATION
339 LIFE
340 LIGHT
341 LIME
342 LIMESTONE
343 LINE SCANNING
344 LITHIUM ALUMINATE
345 LITHIUM METANIOBATE
346 LITHIUM NIOBATE
347 LITHIUM TANTALATE
348 LOW TEMPERATURE
349 LUBRICANT
350 LUMINESCENCE
351 LUNAR ROCKS
352 MAGNETIC FORMING
353 MALFUNCTIONS
354 MANUFACTURE
355 MANUFACTURING
356 MARKET
357 MATERIAL
358 MATERIALS
359 MATHEMATICAL PREDICTION
360 MECHANICAL
361 MEDICAL DEVICE
362 MELTING
363 METABOLISM
364 METAL
365 METAL OXIDE
366 METAL OXIDES
367 METALS
368 MICROBISMS
369 MINES
370 MISSILE
371 MISSILES
372 MODULAR
373 MOISTURE
374 MORTAR
375 MUNICIPAL WASTE
376 MUNICIPAL WASTES
377 MUTAGEN
378 MYLAR
379 NEGATIVE TEMPERATURE
380 NEUTRON
381 NICOTINE
382 NITRILE RUBBER
383 NOVAE
384 NUCLEAR
385 OCCUPATIONAL RISK
386 OCEAN
387 ODOR
388 ONSITE
389 OPERATIONS
390 OPTICAL
391 OPTICAL INFORMATION PROC
392 ORGANO
393 ORGANOLEAD
394 OXYGEN ISOTOPES
395 OZONE
396 PACKAGE
397 PACKAGING
398 PAPER
399 PAPERMAKING
400 PARTICLE RESUSPENSION
401 PARTICULATE CONSOLIDATION
402 PENETRATION
403 PERFORMANCE
404 PESTICIDE
405 PETROCHEMICAL PLANTS
406 PETROLEUM PRODUCTS
407 PHARMACEUTICAL
408 PHARMACOLOGY
409 PHOTOELECTRIC
410 PHOTOVOLTAIC
411 PICKLING
412 PIPE
413 PIPE TOBACCO
414 PIPES
415 PIPING
416 PITCH
417 PLASTER
418 PLASTERS
419 PLASTIC
420 PLASTICS
421 PLUG
422 PLUGS
423 PLUGS ELECTRICAL
424 POISON
425 POLY
426 POLYESTER
427 POLYESTER PLASTICS
428 POLYESTERS
429 POLYMER
430 POLYMERIC FILM
431 POLYMERS
432 POLYMIDE
433 POSITIONING
434 POTTING
435 POUCH
436 POWDER COMPACTION
437 POWER PLANT WASTES
438 PRECISION
439 PREDICTORS
440 PREPARATION
441 PRICE
442 PROCESS STEAM
443 PROCESSED CIGARETTE LEAF
444 PROCESSES
445 PROCESSING
446 PRODUCTION
447 PROPENES
448 PROPERTIES
449 PROPYLENE
450 PROSTHET
451 PROTECT
452 PROTECTION
453 PROTEINOIDES
454 PROTEINS
455 PULP
456 PULPING
457 PYROLOSIS
458 PYROTECHNICS
459 QUALITATIVE
460 RADAR SIGNAL PROCESSING
461 RADIOACTIVE
462 RADIOACTIVE AEROSOLS
463 RADIOGRAPHY
464 RAILROAD
465 RAILWAY
466 RANKING
467 RATING
468 RAYTHEON
469 REACTIONS
470 REACTOR
471 REACTORS
472 RECLAIM
473 RECONSTITUTED SHEET
474 RECONSTRUCTION
475 RECYCLE
476 RECYCLING
477 REDUCED GRAVITY
478 REELING
479 REENTRAINMENT
480 REFINING
481 REFUSE DERIVED FUEL
482 REFUSE DERIVED FUELS
483 REGULATION
484 REPAIRS
485 RESIDENTIAL
486 RESIDUE
487 RESISTANT
488 RESTORATION
489 REUSE
490 REWORKED
491 RHEOLOGY
492 RICOCHET
493 RISK
494 ROCKET
495 ROCKETS
496 ROCKS
497 ROOM
498 ROPE
499 ROTARY FORMING
500 RUBBER
501 RUST PREVENTION
502 SACCHARIN
503 SALT SPRAY
504 SAMPLING
505 SCRAP
506 SEAL
507 SEAL PLUG
508 SEAL PLUGS
509 SEALS
510 SEAWATER
511 SECONDARY USE
512 SEISMIC DATA
513 SEISMIC WAVES
514 SEMICONDUCTORS
515 SETTLING
516 SHALLS
517 SHATTER
518 SHEAR RATE
519 SHEET RATE
520 SHIELD
521 SHIELDING MATERIALS
522 SHOCK
523 SHOE
524 SHOES
525 SHRINK FORMING
526 SHUTDOWN
527 SIDES
528 SIGNATURES
529 SIMULATE
530 SKIP
531 SKIPS
532 SLIME
533 SLUDGE
534 SMELL
535 SMOKE
536 SNAILS
537 SNUFF
538 SODIUM CHLORIDE
539 SOFTING
540 SOLAR
541 SOLAR ARRAY
542 SOLAR CELLS
543 SOLAR PANELS
544 SOLAR POWER
545 SOLAR SUNARRAY
546 SOLID
547 SOLID WASTE
548 SOLID WASTES
549 SOLIDS
550 SONIL
551 SOUND
552 SOUND FIELDS
553 SPACE MANUFACTURING
554 SPACE SUITS
555 SPARK PLUG
556 SPARK PLUGS
557 SPATTER
558 SPECIFIC HEAT

# PHYSICAL SCIENCE TERMS (Continued)

559 SPECIFICATION
560 SPECTRUM ANALYSIS
561 SPLASH
562 SPONGE
563 SPRAY
564 SPRAYS
565 STACK
566 STANDARD
567 STEAM
568 STEEL
569 STICKING
570 STICKING AGENTS
571 STILLAGE
572 STORAGE
573 STRENGTH
574 STRENGTHENING
575 STRESS
576 STRESS ANALYSIS
577 STRESS CORROSION
578 STRIKING
579 SUBMARINE
580 SUBZERO
581 SUNLIGHT
582 SUPERNOVAE
583 SUPPLY
584 SURFACE
585 SURFACE BURST
586 SURFACE TREATMENT
587 SURFACES
588 SYNTHESIS
589 SYNTHETIC RUBBER
590 TAKE OFF
591 TAR
592 TARGET SIGNATURES
593 TEMPERATURE
594 TERMINALS
595 TEST
596 TEXTILE
597 TEXTILES
598 THERMAL
599 THERMAL RECOVERY
600 THERMAL STRESS
601 THERMODYNAMIC
602 THERMOLUMINESCENCE
603 THERMONUCLEAR
604 THIN FILMS
605 TINSEL
606 TIRES
607 TITANIUM ALLOY
608 TITANIUM OXIDE
609 TOPOGRAPHY
610 TOXIC
611 TOXICITY
612 TRACE ELEMENTS
613 TRAIN
614 TRASH
615 TREATMENT
616 TROPICAL
617 TUBE
618 TUBES
619 TUBING METAL
620 TUNGSTEN
621 TURBO MACHINERY
622 ULTRASONIC
623 UNDERSEA
624 UNDERWATER
625 UNEXPLODED AMMUNITION
626 URANIUM ALLOYS
627 URINE
628 VANE
629 VANES
630 VAULT
631 VENILLE
632 VIBRATION
633 VINYL
634 VINYL RUBBER
635 WALLS
636 WASTE
637 WASTES
638 WASTEWATER
639 WATER
640 WEARING
641 WEATHER
642 WEIGHTLESSNESS
643 WHEEL
644 WINCH
645 WINDOWS
646 WIND
647 WINDING
648 WINDMILL
649 WIRE
650 WIRE ROPE
651 WIRE ROPES
652 WIREROPE
653 WOOD
654 X RAY ABSORPTION ANALYSIS
655 X RAY APPARATUS
656 X RAY ASTRONOMY
657 X RAY CAMERAS
658 X RAY DIFFRACTION
659 X RAY FILM
660 X RAY FLASHES
661 X RAY LUMINESCENCE
662 X RAY PHOTOGRAPHY
663 X RAY SCATTERING
664 X RAY SPECTRA
665 X RAY TUBES
666 X RAYS

# APPENDIX E

## EVALUATION PROCEDURES AND FORMS USED AT TEST SITES 2 AND 3

# TEST SITE 2

Subject Procedures for Conducting Retrieval Experiments

1. Decide which query(s) will become a part of the experiment. Six queries will be evaluated during the month of March.

   (A) Determine if request is well matched with one of the VSS modules. The modules are:

   **Business**

   ABI Inform
   Mgmt. Contents

   **Behavioral**

   Psyc. Abs.
   ERIC

   **Life Sciences**

   MeSH
   MeSH' (Chemical Registry NOS.)
   MeSH" (Chemical Substances)
   CA (CEF)
   BIOSIS

   **Physical Sciences**

   DOE
   CA
   EI (COMPENDEX)
   INSPEC
   NASA
   IRON (Trilingual)
   NTIS*

   (B) Determine willingness of end user to participate in the experiment. To participate, end users will be required to record their relevance decisions at the time the search results are examined. (R = relevant or useful item; I = irrelevant or not useful item; ? = undecided). Battelle will pay for the VSS search and half of the regular search. The end user will be required to turn in evaluated results for photocopying to permit subsequent analysis.

*Not in VSS but optional for searching purposes.

2. If 1(A) and (B) are favorable, start the experiment at the search negotiation (pre-search interviews) stage by bringing in the second intermediary. This interview will consist of an end user and the two search intermediaries. Once the interview is complete end user intervention or interaction will be permitted only for the purpose of reviewing VSS output or the search terms selected without the use of VSS. End users interaction during the online session will not be permitted.

3. At the conclusion of the interview, each intermediary must proceed independently until the searches are completed.

4. One searcher will proceed without VSS, the other with VSS. SOP will be employed for all searches done without the aid of VSS. This may or may not involve the use of printed thesauri. It is important that the regular search be done as near to SOP as possible.

5. Both intermediaries may intervene with their online searches to "fine tune" them, but only if this is warranted and is done as SOP. Avoid the temptation to "fine tune" just to made the experimental results better.

6. Each searcher will record the actual time required to prepare the record strategy prior to logon. Also, comments may be recorded both during and after online search session.

7. The searcher who does a non-VSS-type search will print all results for evaluation by end user and store the search profile.

8. The searcher who does a VSS-type search will store the profile before printing any results. Then, in order to eliminate duplicates, the non-VSS search profile will be subtracted (Boolean NOT) from the VSS search profile. The results will be printed and sent to the end user for evaluation. Also, the intersection between VSS and non-VSS (Boolean AND) will be identified by the VSS searcher, the accession numbers printed out, and the results properly identified. These Acc. Nos. represent the citations common to both strategies but evaluated only once by the end user. NOTE: This procedure will only eliminate duplicates within each data base searched. It will not eliminate duplicate across data bases. A sort prior to printing-out data base citations is desirable for the purpose of eventually eliminating duplicates across data bases.

9. Label all printouts according to the following scheme:

| | | Intermediaries | |
|---|---|---|---|
| | | A | B |
| Search Requests | 1 | VSS | -- |
| | 2 | -- | VSS |
| | 3 | VSS | -- |
| | 4 | -- | VSS |
| | 5 | VSS | -- |
| | 6 | -- | VSS |

In this scheme, each intermediary is identified by a letter code, A or B, and he/she keeps that code throughout the experiment. Each search is numbered sequentially

TEST SITE 2 (Continued)

and each data base clearly identified. Thus, one end user request, completely processed through the experimental methodology, might be labeled as follows:

| Search | VSS Searches |
|---|---|
| 1-A ERIC | 1-B ERIC |
| 1-A Psyc. Abs. | 1-B Psyc. Abs. |

The end user will evaluate all four searches for relevance and will use the notation described in (1.B). This is an important detail. We don't want to interpret 6 different notation schemes for denoting relevance, non-relevance and undecided-type decisions.

10. Have the end user complete the brief questionnaire and return it along with the evaluated results.

11. Return to Battelle:

    (A) Intermediary's worksheet with pre-search preparation time and comments

    (B) Copy of online terminal session

    (C) The list of accession numbers common to both VSS and non-VSS searches (properly labeled)

    (D) The end user's evaluated search results

    (E) The end user's completed questionnaire.

RTN/tln

## TEST SITE 2 (Continued)

### VSS Vocabulary Codes

**BUSINESS**

B ABI INFORM
G Management Contents

**SOCIAL SCIENCE**

D ERIC
P APA (Psych. Abstracts)

**PHYSICAL SCIENCE**

A DOE
C Chemical Abstracts
E EI
F Iron (Trilingual)
I INSPEC
N NASA

### Suggested Strategies

| | | |
|---|---|---|
| 20 | 20 | 3 |
| 6 | 6 | 4 |
| 10 | 10 | 6 |
| 7 | 7 | 7 |
| 8 | 8 | 8 |
| 12 | STORE | 20 |
| STORE | - - - | STORE |
| - - - | EXPAND | - - - |
| SCOPE | SCOPE | SCOPE |
| SETPCT | VOCCNT | SETPCT |
| 50 | 4 | 66 |
| VOCCNT | | VOCCNT |
| 4 | | 4 |

20 Expanded synonyms
6 Phrase file switching
10 Narrower terms
7 Ward file
8 Stem file
12 Adjacency
3 Exact match
4 Limited synonym

TEST SITE 2 (Continued)

SEARCHER'S WORKSHEET

Search Number: _____
Search Type: ☐ VSS ☐ non-VSS
Pre-search Prep. Time: _____ min.

Search Number: _____
Search Type: ☐ VSS ☐ non-VSS
Pre-search Prep. Time: _____ min.

Search Number: _____
Search Type: ☐ VSS ☐ non-VSS
Pre-search Prep. Time: _____ min.

Search Number: _____
Search Type: ☐ VSS ☐ non-VSS
Pre-search Prep. Time: _____ min.

Search Number: _____
Search Type: ☐ VSS ☐ non-VSS
Pre-search Prep. Time: _____ min.

Search Number: _____
Search Type: ☐ VSS ☐ non-VSS
Pre-search Prep. Time: _____ min.

Discounting the separate logon and switching strategy set-up procedures, which admittedly are tedious in this experimental system, did VSS help reduce the mental burden of search strategy preparation?

What did you like about VSS?

What did you dislike about VSS?

TEST SITE 2 (Continued)

USER'S EVALUATION SHEET

(1) For search _____, do you feel that you received about the right amount of relevant or useful citations compared with your initial expectations? ☐ Yes ☐ No

(2) For the same search, was the amount of irrelevant citations:
☐ Tolerable ☐ Annoying ☐ Didn't matter
in your opinion?

(3) What was your overall degree of satisfaction with this search? Circle one.

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| Very Dissatisfied | Moderately Dissatisfied | Moderately Satisfied | Highly Satisfied |

---

(4) For search _____, were the additional relevant citations:

☐ Indispensable, i.e., worth the additional cost and effort to get them.

☐ Useful, i.e., happy to have but could have gotten by without them.

☐ Not too useful, i.e., of marginal or little value and probably not worth the cost and effort to get them.

(5) For the same search, was the amount of irrelevant citations:

☐ Tolerable ☐ Annoying ☐ Didn't matter

in your opinion?

(6) What was your overall degree of satisfaction with this search? Circle one.

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| Very Dissatisfied | Moderately Dissatisfied | Moderately Satisfied | Very Satisfied |

TEST SITE 3

## VSS Experimental Methodology.

Step 1 Determine if request is applicalbe to this experiment:

(a) The request must involve VSS databases:

| Social Science | Business |
|---|---|
| ERIC | ABI Inform |
| Psy. Abs. | Mgmt. Contents |

| Life Science | Physical Science |
|---|---|
| BIOSIS | CA |
| BIOSIS-C | DOE |
| CA | EI |
| MeSH | INSPEC |
| MeSH-R | IRON |
| MeSH-S | NASA |

(b) Request must involve a search of two or more databases.

Step 2 Determine end user's willingness to participate in the experiment. To participate, end users will be required to (a) negotiate their search request in the presence of two searchers (intermediaries), (b) review search terms and strategies prior to the online session, as needed, (c) be agreeable to forfeit any interaction during the online search itself, (d) evaluate printed citations for relevance (a maximum of 100 citations will be evaluated), and (e) complete a specially designed questionnaire.

Step 3 Both intermediaries conduct a joint pre-search interview with the end user. All three people must be present at the start (end user and both searchers). Record search negotiation time on intermediary worksheet.

Step 4 Both intermediaries prepare search strategies based on pre-search interview, one intermediatey will use VSS, the other will not. IMPORTANT: The intermediary who does not use VSS must not go out of his/her way to do anything extraordinary to try to look good or beat the VSS system. In fact, there may be times when the intermediary would not consult a printed thesaurus under certain circumstances. If this is the case, then a thesaurus should not be consulted for this experiment. This is an honor system. **DO NOT DO ANYTHING OUT OF THE ORDINARY FOR THE SEARCH** if you are the intermediary who is performing the non-VSS search.

TEST SITE 3 (Continued)

Step 5 — End user "fine tuning" of the search will be permitted only to the extent of saying "Yea" or "Nay" to terms produced by VSS or non-VSS methods. Under no circumstances may a term used in one strategy be added to the other if found missing there. However, in evaluating each term in each strategy, the end user may, in effect delete identical terms from each strategy. This is permissible.

Step 6 — The searcher using VSS must do the following: (a) select all the vocabularies in the module chosen, (b) select at least eight terms to be displayed per vocabulary, and (c) perform switching on two VSS options, browse and narrow/broader.

Step 7 — Intermediary using VSS must complete VSS Evaluation form for each VSS search performed.

Step 8 — Intermediaries must record time for using VSS or manual thesaurus and search preparation time on intermediary worksheet.

Step 9 — When both search strategies are ready for execution as online searches, both strategies will be entered in their entirety and stored as profiles. Each strategy will be executed against each database requested by the end user. Although duplicate citations cannot be eliminated between databases, within the same database the following procedure must be followed:

(a) Subtract (boolean NOT) the non-VSS profile from the VSS profile to produce Set A. Label the offline printout as SET A. Perform the reverse subtraction to produce Set C. Label this printout as SET C. In other words, SET A will always be citations unique to VSS and SET C will always be citations unique to non-VSS.

(b) If either SET A or SET C contains more than 50 citations, limit both sets by date to reduce the number of citations to 50 or less in each set before invoking offline print.

(c) Intersect (boolean AND) SETS A anc C to produce Set B. Label printout as SET B. Set B represents citations common to VSS and non-VSS profiles.

Step 10 — Present each set to the end user without reference to VSS or non-VSS and without reference to the searcher's name. Set B must always be the intersection of both profiles. Set A must always be VSS unique and Set C must always be non-VSS unique.

TEST SITE 3 (Continued)

Step 11 Have the end user evaluate each citation in each set (A, B, C) for relevance to his request. His/her decisions should be marked in the column on the printouts using R for relevant, I for irrelevant, and ? for undecided. **NO OTHER NOTATION SHOULD BE USED.**

Step 12 Upon completion of relevance judgements, end user must complete end user's worksheet for that search. End user must return marked-up printouts and worksheet to searcher for photocopying.

Step 13 Return all worksheets, printouts and copies of online sessions to Battelle for analysis.

Agreement Intermediaries should have similar capabilities, educational background and online experience. Intermediaries agree to perform a minimum of six searches over a 6-week period in dual or replicated fashion, as follows:

| | Intermediary 1 | Intermediary 2 |
|---|---|---|
| Search 1 | VSS | Non-VSS |
| Search 2 | Non-VSS | VSS |
| Search 3 | VSS | Non-VSS |
| Search 4 | Non-VSS | VSS |
| Search 5 | VSS | Non-VSS |
| Search 6 | Non-VSS | VSS |

TEST SITE 3 (Continued)

INTERMEDIARY'S WORKSHEET

(1) Institution I.D.: _____

(2) Intermediary I.D.: _____

(3) Search I.D.: _____

(4) VSS Module used:

- ☐ business
- ☐ life sciences
- ☐ social sciences
- ☐ physical sciences

(5) PERFORMANCE DATA FOR THIS SEARCH

| Pre-Search Preparation (in minutes) | | | Online Connect Time (in minutes) | | |
|---|---|---|---|---|---|
| VSS | | Non-VSS | VSS | | Non-VSS |
| _____ | Search Negotiation | _____ | _____ | DB 1 | _____ |
| _____ | VSS Online | xxxxxxx | _____ | 2 | _____ |
| xxxxxx | Thesaurus Look up | _____ | _____ | 3 | _____ |
| _____ | Strategy Preparation | _____ | _____ | 4 | _____ |
| _____ | TOTAL | _____ | _____ | 5 | _____ |
| | | | _____ | TOTAL | _____ |

(6) Comments about this search, including any difficulties encountered (VSS or Non USS): _____

_____

_____

TEST SITE 3 (Continued)
END USER'S WORKSHEET

(1) Institution I.D.: _____

(2) End user I.D.: _____

(3) Search I.D. _____

(4) State your search request in your own words <u>after</u> you have discussed it with the person who will do it.

_____

_____

_____

_____

_____

(5) In your past experience with online searches, are you generally:

- ☐ satisfied with the results
- ☐ dissatisfied with the results
- ☐ varies from search to search
- ☐ no past experience

---

(6) PERFORMANCE DATA FOR THIS SEARCH

| Group | Number of Relevant Citations in Group | Number of Irrelevant Citations in Group | Number of Citations Undecided About in Group | Total Citations in Group |
|---|---|---|---|---|
| A | | | | |
| B | | | | |
| C | | | | |

TEST SITE 3 (Continued)

END USER'S WORKSHEET
(Continued)

(7) Rank the following citation groupings according to your overall degree of satisfaction:

| Citation Groupings | Satisfaction Scale | | | | |
|---|---|---|---|---|---|
| Group A plus Group B | 1 | 2 | 3 | 4 | 5 |
| Group B plus Group C | 1 | 2 | 3 | 4 | 5 |
| Group A plus B plus C | 1 | 2 | 3 | 4 | 5 |

1 = very dissatisfied 5 = very satisfied

(8) If you could select only one combination shown below which combination would it be?

| Group Combination | Choice (check only one) |
|---|---|
| Group A plus Group B | _____ |
| Group B plus Group C | _____ |

Why did you choose this combination? _____

_____

_____

_____

(9) If you had to pay extra for the combination you selected in Question 8 compared to the combination you did not select, how much extra would you be willing to pay?

0 $5 $10 $15 >$15

# APPENDIX F

# VSS EVALUATION FORM

# VSS EVALUATION FORM

(1) Institution I.D.: _____

(2) User Name I.D.: _____

(3) Your institution can best be described by one of the following:

Private sector

- ☐ independent broker
- ☐ corporate library/information center
- ☐ non-profit institute library/information center
- ☐ database producer (primarily)
- ☐ online search service (primarily)
- ☐ database producer/online search service (about equally)

Government (Federal)

- ☐ library/information service
- ☐ database producer (primarily)
- ☐ online vendor (primarily)
- ☐ database producer/online vendor (about equally)

Academia

- ☐ library/information service
- ☐ library/information school or department

Other

- ☐ public library/information service
- ☐ _____

(4) Your Academic background:

| Degree | Field of Study for Each Degree |
|---|---|
| No Degree | _____ |
| Bachelor's | _____ |
| Master's | _____ |
| Doctorate | _____ |
| Post Doctorate | _____ |

(5) Years at present company: _____

(6) Years in library/information activity: _____

(7) Years experience with online systems: _____

# VSS EVALUATION FORM
(Continued)

(8) On a scale of 1 to 5, rank the databases listed below according to your current proficiency with and usage of them. Please rank all choices listed. If you currently do not use a particular database, so indicate this, and ignore the proficiency scale for that one.

| Data Bases | Your Proficiency | | | | | Frequency of Use | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ABI Inform | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Management Contents | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| ERIC | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Psychological Abstracts | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| MEDLINE | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| BIOSIS | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Chemical Abstracts (CA Search) | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Compendex | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| INSPEC | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| NASA Recon | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| DOE Recon | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |

| | |
|---|---|
| 1 = little or no proficiency | 1 = no usage |
| 2 = below average proficiency | 2 = less than once/month |
| 3 = average proficiency | 3 = one or more times/month |
| 4 = above average proficiency | 4 = one or more times/week |
| 5 = expert | 5 - daily usage |

(9) How often do you search with the following; (answer should total 100%)

a. controlled vocabulary terms only _____ %
b. controlled plus free text terms _____ %
c. free text terms only _____ %

(10) Indicate your preference below (Please rank "1" for most preferred, "2" for next most preferred, and "3" for least preferred).

a. controlled vocabulary terms only _____
b. controlled plus free text terms _____
c. free text terms only _____

(1) Tally your actual usage of VSS during the test period.

Search No.: 1 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

Search No.: 2 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

(11) Tally your actual usage of VSS during the test period.

Search No.: 3 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

Search No.: 4 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

(11) Tally your actual usage of VSS during the test period.

Search No.: 5 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

Search No.: 6 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS-switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

(11) Tally your actual usage of VSS during the test period.

Search No.: 7 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

Search No.: 8 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use?_____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

11) Tally your actual usage of VSS during the test period.

Search No.: 9 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use? _____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

Search No.: 10 Date:

Search Statement:

- Circle VSS vocabularies selected for this search:
  (refer to VSS online menu)
  (1) A B (2) A B (3) A B C D E F (4) A B C D E F
- Circle VSS switching features selected for this search:
  (refer to VSS online menu)
  1 2 3 4 5 6
- Were you able to use any of the VSS results for this search? ☐ Yes ☐ No
- How many VSS search terms or concepts were you able to use? _____
- Was VSS helpful in suggesting any other ☐ terms or ☐ avenues of approach for improving this search, beyond what the system actually produced? ☐ Yes ☐ No
- In your best judgement, how would VSS output compare to your own normal effort for this search?
  - ☐ VSS output would probably not be as good as my own normal effort
  - ☐ VSS output would probably be better than my own normal effort
  - ☐ VSS output would probably be about the same as my own normal effort
- How would you rate VSS for this particular search?
  1 2 3 4 5 (1 = of no help 5 = very helpful)
- Comments:

VSS EVALUATION FORM
(Continued)

(12) Was VSS easy to:

| | | | | | |
|---|---|---|---|---|---|
| (a) learn | 1 | 2 | 3 | 4 | 5 |
| (b) use | 1 | 2 | 3 | 4 | 5 |
| (c) understand | 1 | 2 | 3 | 4 | 5 |

1 = ease ⟶ 5 = hard

(13) Do you feel confident with VSS's:

| | | | | | |
|---|---|---|---|---|---|
| (a) capabilities | 1 | 2 | 3 | 4 | 5 |
| (b) output | 1 | 2 | 3 | 4 | 5 |

1 = confident ⟶ 5 = no confidence

(14) If VSS were available as another feature of present online search services, would it make your job of searching multiple databases:

☐ easier ☐ harder ☐ no difference

compared to your present method(s) of online searching.

Why?

(15) What was your overall reeaction to VSS?

(16) What do you think about subject switching in general?

(17) Comments:

# APPENDIX G

## SUMMARY OF OPEN-ENDED QUESTIONS

14. IF VSS WERE AVAILABLE AS ANOTHER FEATURE OF PRESENT ONLINE SEARCH SERVICES, WOULD IT MAKE YOUR JOB OF SEARCHING MULTIPLE DATABASES:

EASIER HARDER NO DIFFERENCE

Easier. There would no longer be a need to look through several data base thesauri. My search area would be neater, and thus I could think clearer!

Easier. I would not need to keep paper thesauri for every database I use.

No difference. Your system doesn't really save time for an experienced searcher who knows his or her thesauri.

No difference. I use controlled terms and free text now.

Easier. It would make my job easier only if seldom used, but expensive thesauri were a part of VSS, and if it were inexpensive itself. My job would be made easier in that I would have to justify fewer expenditures.

Easier. It enables me to determine the entries to search on another data base if I cannot get the info I need on the present data base.

Easier. Faster. Would not have to look up preferred terms.

Easier. No Comment.

No answer. No comment.

No difference. With free text searching, the majority of my hits come from these terms, not so much the controlled vocabulary.

No difference. I search in life sciences and need the NT/BT feature so VSS would not help and so I couldn't use it.

No difference. No comment.

Easier. Saves time.

No difference. No comment.

Easier. It can't hurt, only help.

No difference. BRS's CROS and DIALOG's File 411 have provided sufficient guidance to potential files -- at a reasonable cost and quickly. ERIC's print thesaurus is used frequently before going on-line.

No difference. No comment.

Harder. Yet another protocol to learn and deal with.

Easier. Print thesauri for databases are often not commonly available.

No difference. It is very time consuming and relatively unresponsive even with terms one would expect to be in the vocabulary. Some of the terms were just plain garbage!

Somewhat easier. Useful in areas in which I am aware of only one or two terms and VSS is able to come up with related terms that I was unaware of.

Easier. Did suggest more terms in a couple of cases.

No difference. Not clear enough how switching works. Too cumbersome to use in present form.

Easier. I would not always need to refer to thesaurus -- could even get by without having so many. It is a quicker more efficient way of checking terminology across files than using hard copy thesauri.

No difference. I don't think we'd use it as is. We'd still have to go to printed thesauri for notations (esp. mesh) and we'd still prefer to pull up citations and look at descriptors.

No difference. Approximately 65% of my searching is free text. Though VSS would contribute some additional thesaurus terms on the whole -- it's still free text searching that I require.

No difference. May make it somewhat easier but I think I more or less know the terms I need. If I didn't, it's easier to print out the descriptors in the data base I'm already in.

Easier. Assist in locating controller terms for file we don't yet have thesauri for.

No difference. Because I usually use both free text and controlled vocabularies.

No difference. No comment.

No answer. Application of thesaurus terms varies from database to database, e.g., EI assigns only a few broad terms, INSPEC is more in depth, BHRA only when concept not retrieveable in TI or AB. To only search the DE field often misses relevant hits. Strategy must vary from DB to DB.

Easier. I can search several data bases at once; it gives me a "lead" into data bases I am not familiar with.

Easier. Not all data bases have an online Thesaurus.

Harder. Present methods: (1) produce more relevant references, (2) consume less time in formulating a search strategy, and (3) incur no additional costs for online connect time.

No difference. Text searching of title and abstract for terms or phrases one wants to see in title or abstract will produce citations of interest. A display of these will provide indexing terms and identifiers for that data base.

No difference. The type of format used is time consuming.

Easier. Extends the EXPAND capabilities by merging files. Does a very good job of relating terms if the database producer has built a BT/NT thesaurus. Outside the sci/tech files, VSS's usefulness is decreased.

Harder. I'd still check manually. Your system doesn't really save time for an experienced searcher who knows his or her thesauri.

Harder, until it was reloaded to exhibit all thesaural characteristics. BT-NT first in browse rework adj. and stem match.

Easier. Only if it were available for databases that lack good thesauri. Nearly all the databases indexed have excellent thesauri, which I think are more easily used one at a time than combined. Also, the thesauri have certain features that are not included in the VSS. If you put such a system together for databases with no thesarus, or very costly thesauri or multiple thesauri (NTLS, for example), then it would be extremely useful.

No difference. We have on-line search aids that basically serve our needs.

Easier. Easier to locate terms in other vocabularies.

No difference. If there was no match of terms among thesauri, there's no place to go from there. If I'm going to rely on exact matches, why not just free text across databases?

No answer. No comment.

No difference. I spend a good deal of time in the interview with the patron examining thesauri, developing synonym lists and plain figuring out backward variations to sneak up on topic.

Easier. Because is brings together the terms used by different thesauri, it helps in the selection of appropriate terms for effective searches of particular databases.

Harder. Introduces an additional "head" in the way. Finding valid "hits" is a matter of understanding the need -- the database, the source of the database. Having one more "help" doesn't really help.

Harder. Use of this system represents a substantial amount of time that could be better spent searching traditional printed sources when composing an on-line search strategy.

No difference. No comments.

Harder. There is very little information received by the user which is valuable for switching purposes.

No difference. I probably wouldn't use it very much, as free-text searching works almost as well.

Easier. My brief experience with VSS did provide more possible search terms.

No difference. Its data base is <u>too</u> limited.

No difference. I would not use it in databases for which the vocabulary was familiar to me, and I would need a far more detailed explanation of what to expect from the system, and how to use it.

Easier, if you don't predetermine subsets. I want to chose the databases to cross-search.

No difference. I need more information on <u>how</u> it matches and retrieves terms.

No difference. When only 2 or 3 databases are to be used in doing the same search, it is just as easy to consult their printed thesauri. Moreover, the scope notes provided in the printed thesauri can give a better idea of the appropriateness of using a particular term.

No difference. I can already use Dialindex which not only covers controlled, but also free-text vocabulary and also gives me an indication of the postings.

Easier. When you are switching databases, it is helpful to know in advance the differences in the terminology.

Easier. Would not have to look up terms in thesauri.

No difference. No comment.

Easier. When switching databases and vocabulary, it would take the guess work and/or checking printed thesauri out of the searching process. I often switch from ERIC to PsychInfo with "end/savetemp" and this VSS system could eliminate the less than successful switch by giving the right vocabulary. Also in search strategy, we could use the VSS as prompts for alternative search strategies.

Easier. Good selection of terms is suggested.

No difference. Most of the searching done is on scientific data bases with which I am thoroughly familiar. The system would be of value when I have to search business or social science data bases.

No difference. VSS gave me little, if any, additional information.

15. WHAT WAS YOUR OVERAL REACTION TO VSS?

Even though my use was limited, I was impressed with the VSS. The system is relatively easy to learn, quick, and reliable. Logging on isn't easy.

Since it is only a subset of terms a real commercial VSS could be, it did not score highly in the Biological area. Its strength must be in the hard sciences.

Increasingly negative.

Interesting.

Interest. It appears to be a first step. By the third or fourth generation, it should really be worthwhile. But it must be very inexpensive to use. Unless you keep the thesauri updated -- without fail --VSS will make more enemies than friends.

Can be useful if you access a number of data bases frequently for different topics.

Very favorable. It appeared to work as advertised.

From what we have seen, VSS looks like it could be a valuable searching aid. A better evaluation might be made if we knew more about how it would integrate with an active on-line system (e.g., DIALOG).

No comment.

No comment.

It was handy, but the message to choose an option after each result --became tedious. It would be helpful if you could stack commands to get over the repetitive message.

I doubt if it is worth what it costs.

Indifferent

Very positive. It seems as if it is a straightforward, relatively uncomplicated (in execution, that is) system that can aid searchers by saving time and being comprehensive.

I didn't feel it was very useful for everything I needed it for.

Some of the options didn't work -- As in more info available.

Interested at first but the results from the "quality circles", search began a negative trend.

No comment.

There were not enough synonyms to be of real assistance; nor were there enough hierarchical relationships represented to offer assistance.

There seemed to be some programming difficulties and the manual was not easy to use, either as a reference or a learning tool.

I got very tired of the menu displays and the reams of paper output generated. ???????? would be helpful; also it appeared that you had to logoff and back on again to reach a different file group -- rather inefficient.

Interesting, but mildly disappointing. Perhaps greater familiarity with the system would have led to better results, but too many of the switches led to nothing useful.

Too user friendly -- I got so bored waiting for it to list the few menus I could choose from. Needs an "expert" mode (or "semi-expert").

Good idea -- needs work.

Positive - for reasons above.
Negative - switching features part not necessary. Can't all terms be listed for a term at once (broader, narrower, synonyms, etc.) instead of having to select option each time?

Seemed a good idea; primitive stage; seems overwhelming to contemplate how you could ever make it really useful by entering any free text term--have to have broad descriptors only -- some preknowledge required.

I liked it. If it was set up so that one could save terms and then go onto those databases and run the search on those terms.

Seems cumbersome.

VSS would be much more valuable with vocabularies having broader and narrower terms -- I continued my searching to the Business Vocabularies.

Need more experience. As with all systems, it takes practice.

No comment.

Should be continued as a project. Have you considered using REPORT DISPLAY feature of BASIS to provide tabular display?

I thought it was a good system; I don't like the "synonym" feature, but "browse" was excellent.

There is a great deal of information but it is poorly presented and out of date.

VSS is too time consuming for productive online searching.

My overall reaction to VSS is that the end result is not worth the effort and cost to reach higher correlation of indexing terms between data bases. In a majority of instances the switch is direct or can be by truncating. The added step required by VSS equates to time and cost increases far in excess of benefits.

VSS would be more useful if the vocabularies were continuously updated. Many of the current subject terms are missing.

Browsing did not help significantly while doing a subject search; might be more useful with author's, company names, etc. Synonyms, broader/narrower terms good to the degree of the vocabularies input. Option 6 needs to be explained more completely before I could judge.

There is some potential here.

No comment.

We were not able to use VSS to its full capacity because of difficulty with tymnet. We also had to borrow a 300 baud terminal because our terminal (TI OMNI-820 KSR) cannot be set at half Duplex.

Instructions were difficult to understand. Consequently, I don't think I was able to explore VSS capabilities adequately.

I found the menu tiring. I'd like to be able to combine databases in a different way (e.g., Psyc, MESH, and BIOSIS or Psyc, Mgt., ABI). I'd like to be able to add narrower terms to the browse feature. I found the "switch not successful" response uninformative--why was it unsuccessful?

No comment.

Not at all impressed.

I was impressed with its switching capability. Matches were found for most terms and most of them were relevant. It would be expecially useful for an end user or new searcher unfamiliar with the terminology used by particular data bases.

Negative.

I was impressed with the technical accomplishment that this system represents but not with its practicality or utility in the on-line database searching context.

Menu characteristics frustrating. Not particularly user friendly.

It is a beginning but more development of the software is mandatory if it is to function in a reference capacity.

It was not worth the effort, although it would improve over time, as it became easer to use.

I would like to try it more.

Very limited value.

Very frustrated--the repetition of the menu is frustrating and time-consuming and there are too many term types to keep straight with an initial use. Feature 6 might be more useful if the commands and options were explained more fully.

Frustration.

It was cumbersome. Menu driven method got me bogged down.

1) Found the menu selection to be helpful in the beginning but very tedious after this.
2) It would be easier to use if all the terms from a particular vocabulary were placed together.

The menus need to be improved. There should be an option to move directly to another set. There should also be more features for experienced searchers so that they could skip over the repetition of menus.

I think there is a lot of potential in this kind of switching system. however, from my results I think it either needs some more work or it needs clear instructions, with examples, of its capabilities.

I think it is a very good system and something that is needed in on-line searching.

It was laborious and added very little value due to its limitations.

Favorable, except when system and/or phone lines went haywire and I could not get back on the right track by logging off and on again.

Easy to use; good design; very helpful.

The system as currently implemented is clumsy and time-consuming. For acceptance in an on-line mode for help in structuring searches, it would need to be streamlined and rely less heavily on its menu-driven features. The capability to update thesauri would have to be implemented. Most of the adj.-lead produced only nonsense. The system has potential of being a useful tool.

VSS gave me little, if any, additional information.

16. WHAT DO YOU THINK ABOUT SUBJECT SWITCHING IN GENERAL?

I think that it's a very useful mechanism. Very often searchers forget to adapt their searches to other data base indexing principles, etc. Subject switching allows searchers to constantly keep aware of the fact that data bases are unique and that every term must be checked.

I would prefer subject switching to file 411 (DIALOG), and would gladly pay for it. It is a much more systematic approach to database searching than plunging in with lists of free text terms. (If there is leftover time, let me get back to it!)

In theory, great -- but this is not it!

Useful

I think it is another unnecessary jargon term. You've built a gizmo that looks through several thesauri in a very heavy handed manner. Under the rubric "on-line thesauri," I like the idea.

No comment.

Needed. But the whole operation should be transparent to users. They should be able to select a term and search a data base without having to re-key the "selected" term. Because I was not familiar with searching the data bases that are part of the Vocabulary Switching System, I prepared search questions, constructed profiles, and then used the Vocabulary Switching System to augment or refine the terms I had selected. While this was not a true test of the Vocabulary Switching System, it did allow me to see where the System could improve on the kinds and number of terms I had selected.
Overall the Vocabulary Switching System was simple to use. I was not sure, however, of the meaning and use of the Vocabulary Switching System commands. These were displayed during searches, but I could not figure out exactly what each one meant. Had I had more time, rather than just two hours, I could have experimented more with the procedures to see what all the Vocabulary Switching System could do. I felt pressed for time and perhaps this was the reason some of my searches were not too successful.
The greatest asset of the Vocabulary Switching System is the amount of time it saves since the correct search terms are selected for the searcher. When working with several vocabularies, it is difficult, if not impossible, to remember the correct terms for each system. For data bases with both controlled and free vocabularies, the Vocabulary Switching System can improve search effectiveness since alternative pathways are suggested by some of the retrieved terms.
One feature that would have been very useful, even during this test, is a command that displayed which VSS vocabulary set and switching feature had been selected. Several times during the test I lost track of switching feature I had selected.

How about statistics on term searched (number of times used in each particular job).

No comment.

Good idea, would be nice to use updated thesauri.

It has to be done thoroughly or else the searcher is worse off than left to his own devices and printed thesauri.

No comment.

A very helpful tool for cross-disciplinary searching (e.g., education and psychology).

Might be useful in the future.

No comment.

Not too terrific in vocabularies 1 and 2 but it might prove more useful in the others.

No comment

No comment.

Good theoretical idea but development requires a thorough knowledge of thesauri being manipulated in order to truly exploit the subjects.

As long as on-line databases use a controlled vocabulary, subject switching is both useful and almost necessary for quick, easy on-line searches.

The idea is a good one since we cannot always think of all the relevant subject headings for a given topic. It is probably more useful in social science where searching must rely more heavily on descriptors/controlled vocabulary.

Theoretically an excellent idea, but I am left with the general impression that implementation, especially for a non-expert user, is extremely difficult.

I think its a great idea, but either I wasn't using it correctly, it has limited applications, or it needs more work.

Good idea -- needs work.

The definitions in blue sheets of workbook were not particularly easy to understand ex.-"rel phrase", "wd match", etc.

Synonyms, broader/narrower great idea -- even antonyms would be good.

I like it when it can be used in the databases directly without re-keying terms.

I think it could be useful if it can be done cheaply and quickly.

Generally a good idea -- still feel that obtaining thesauri and consulting them will be necessary.

It is necessary.

No comments.

Valid concept -- should continue to be researched.

Good idea, saves time, aids in searching.

Subject Switching is very valuable not only in doing searches on another data base, but also for lexicographical purposes.

Subject switching is a useful tool when properly utilized, but computerized subject switching is of questionable value when used to search multiple data bases.

In general, I think subject switching may be of assistance to the novice or to the unimaginative operator of machine retrieval equipment. I think the knowledgeable operator would find VSS of little assistance.

Subject switching, in general, is a useful tool for online searching, and could be very cost effective. The vocabularies must be kept updated to provide current material for high technology.

When an "intelligent" computer is available to handle subject switching, life will be easy for all of us. I feel that waiting for vocabularies to be updated, printed, etc., makes the information outdated, so it will be a greater benefit for the computer to be the impetus for the update than the databases.

We need it desparately but we need it better done.

No comment.

Subject Switching is of little use for the type of on-line searching we do. It probably would be extremely useful to on-line users in highly technical operations.

Good idea if easy to use.

I think it's difficult to do well. I think it requires more than simply "matching" terms. In other words, I think it takes some human thought to match up concepts expressed differently in various thesauri. I also think that a major use for a VSS would be for new concepts not necessarily in a controlled vocabulary. Free-text suggestions are useful (e.g., in a file such as BRS' TERM database. I found the TERM database concept more useful and more flexible, mainly because of free text suggestions and because term matches were determined by database creator.

My questions in using the VSS were always:

- If I already know the controlled vocabulary term, why search it here?
- What am I learning that I couldn't set as easily in a print thesaurus?
- How does this help me for new terminology or in searching fields I'm not familiar with?
- Do I really understand the purpose of this system?

No comment.

Battelle asked me 3 ± years ago -- I feel this is not one of the computer's strong points.

Subject switching makes searching across multiple databases much easier. The selection of the correct terms has a significant impact on the search results. Even free text searches would benefit from this type of subject switching.

Too cumbersome.

The utility of such a system cannot be denied especially if it could be incorporated into database searching as a kind of automatic feature if and when the time comes when simultaneous cross-database searching is commonplace. (1) Given the current capabilities of VSS, I don't believe I would have much occasion to use the system. (2) The menu-based command system is tood awkward and restrictive. (3) The current VSS Instruction Booklet is incomplete and unclear, failing to explain the full potential, application, and capabilities of the system.

Wonderful idea -- not done well w/VSS.

I think it is a useful aid to anyone who needs to work with numerous sets of controlled vocabularies.

Interesting concept, but of doubtful utility.

Good idea.

It could be a valuable source of information.

It has potential for searchers who search multiple databases but do not have access to printed vocabulary. However, it would have to be quicker to use.

Good idea.

Great idea. Needs more ability to work with phrases and scope notes--character matching is still too primitive.

1) It is a useful aid to have when searching on more than 2 databases or when printed thesaurus is unavailable.
2) Might be more useful if it was somehow incorporated into existing features like Dialindex in Dialog (or Cross in BRS) where postings of the terms available are also given.
3) Doubt I will utilize it at all if access to it was based on the same prices charged for the databases whose vocab are included.
4) In summary, I would, at best, use it occasionally as a pre-search tool if I could access it at a minimal charge. This is because I tend to mostly use a combination of controlled and uncontrolled vocab.

Unless it is broadened to include free-text vocabulary, it is not particularly valuable.

I like the concept. I think this system needs more work.

In view of the differences in vocabularies from one database to another, I feel it would be helpful to provide some aid to the searcher in automatically switching between the various vocabularies.

No comment.

In general, I think the idea has great potential, especially for searchers who routinely search numerous databases. The value of time saved alone would make subject switching worthwhile. I think system needs refining, and I would have liked to have tested it in use with the databases (i.e., searching a database, then logging on subject switching, then going on to next database.)

I like it.

With the proliferation of on-line data bases, subject switching will become a necessity. This is especially true for multi-discipline libraries.

Good idea, but needs much more work.

17. COMMENTS

No comment.

No comment.

No comment.

No comment.

I suspect that if I had to pay for VSS, I wouldn't use it. As currently set up, it produces entirely too much random noise. It may be greeted more lovingly by a librarian who has little-to-no idea what a topic is really about -- but then some of the false results may make her (or him) look like a real twit to the client.

No comment.

We would be interested in receiving any reports generated as a result of this evaluation. Thanks for inviting us to participate.

No comment.

No comment.

Since you can't use BT/NT for MESH, you have to use the printed thesaurus so use of VSS is limited.

Lack of entries for subheadings is a severe drawback.
Inclusion of non MESH terms is not useful if NT feature is not available -- except to send user back to printed Thesaurus.
VSS would be easier to use if it gave one the opportunity to increase the number of terms without having to review the whole menu.
Why are terms duplicated with and without dates?
Absence of some fairly obvious terms indicates that the vocabulary was not built from the 1980 MESH as was announced.
If you can't get NT/BT capability for the Life Sciences, at least terms on the same level should be retrieved.

No comment.

No comment.

No comment.

No comment.

CROS and 411 usually produce relevant files. Most of the files which we use are easily manipulated using free text and descriptions. As VSS is now I would not use it or recommend it.

No comment.

Interesting experience as a test but one soon tires of menu-driven systems; it seems it would be useful for foreign language vocabularies.

I personally often find menus extremely tiring, particularly if one uses the program often. For me, a two week trial would have been more productive since I do relatively little searching. I need more time to "play" with the system. I was unable to do any searching my last 2 days since I was involved with an unexpected rush job (not searching).

I was glad to have an opportunity to try VSS. It was an interesting notion, but I found it somewhat clumsy to use -- and the menus were ??????????! I use DIALOG a lot and by comparison, I found VSS to be very time-consuming for the amount of response I obtained. In addition the capability on VSS of recording the preceding and followup terms, i.e., (adj-lead and adj-word) is interesting but generally irrelevant since it applies only to descriptors (as opposed to an on-line index display where all the terms from the file that precede and follow will be shown).

Mainly though, I was surprised by the poor results, especially in the science files. And, because this was a stand-alone capability, I had no way of knowing if results would have improved and the whole file been searched, particularly in the case of NASA. —
I am convinced that there is a need for thesaurus search capabilities on-line -- however, they cannot stand alone as on VSS. Perhaps VSS in modified form along with other search system capabilities can be made into a useful tool. But, I have come away with the feeling, once again, that thesauruses in what ever form, are not the best search tools because they do not stand alone well in today's searching!

Some of my selections, e.g., electronic mail, may have been too current for the experimental vocabularies, which I assume are somewhat dated.

My rather quick study of the instructions and inflexible use of the browse option may have led to the limited success I had with VSS, but any sort of practical application of this sort of thing would have to cope with naive use.

You're truncating too far -- I got Hopi for homosexuality!

No comment.

I would like to see a system like this cover even more files - also, I was very busy that week and not able to use it as much as I wanted. I'm sure I would use this system regularly if it were not too expensive.

Logon takes too long, fast response time, menu format takes a lot of paper!

See letter.

No comment.

No comment.

The documentation was quite skimpy, both on-line and offline. Much had to be learned trial and error fashion.

No comment.

No comment.

Lead terms which are not posting terms should be identified to permit retrieval of the posting term (X-references). Some 'stem terms seemed incorrect. I was disappointed.

No comment.

The VSS system has a lot of information but would take up a lot of time and be too expensive in its present mode.

The user - defined switching feature is poorly explained and illustrated.

Machine retrieval of information as a judgment matter. A creative mind is a great help. Search of title and abstract for unique words or phrases combined with truncated terms in the indexing fields is a direct approach which will result in pertinent citations. These provide a ready display of indexing fields as well as identifiers and special numerical designators. This approach does pre suppose all data bases will provide text search capability and that emphasis is placed on good title, title extensions or supplements, and on a well written abstract.

The subjects selected for the four tests provided few specific subject terms. The Broader/Narrower Terms feature produced little specific output. The most helpful subject switching feature was Browse.

ESA-QUEST's ZOOM feature is comparable to this program in places.

Whatever you'll charge will be more than I'd want to pay to perhaps save a few moments of manual checking.

Nevertheless I must conclude that VSS in its present form leaves a lot to be desired if we are to consider its use as an attachment to DIALOG or to an on-line catalog which has access to more than one library's catalog. My reasons for this are as follows:

1. VSS in its present form is mainly menu driven, not allowing much flexibility other than those pre-ordained; the change from one set of thesauri to another is laborious; the options do not match the characteristics of the thesauri nor those of the searching terms.
2. The browse option, the easiest to access, has an ordered output which prevents a thesaurus display (related terms, BT-NT) until all the stemming lists are given; oftentimes these are way off target and the searcher must cycle back, increase the listing length, go through all the stems again, etc., before getting to the thesaurally linked terms. When the other options are used and there is not an identical match, "search is not successful".
3. I was very surprised to see what of the various thesauri you had NOT loaded, e.g., the MESH tree structure. This knocked out in one fell swoop all the BT-NT relationships. Surely you could have done something like a table look-up to preserve those relationships. The same thing is true of other thesauri that use category numbers to exhibit hierarchies in outline fashion rather than only one level at a time. The next version of VSS would have to remedy this weakness. I think the BSI ROOT program does better.

4. You never show scope notes, so valuable as "definitions".

No comment.

No comment.

No comment.

No comment.

No comment.

No comment.

The present tutorial approach used in VSS should be supplemented with a more advanced version in which several terms and combinations of options could be searched. This would provide faster switching and make VSS use more cost effective. This would also enhance its use for thesaurus construction.

No comment.

Although the "term type" relationship is given for retrieved terms, there is no scope note, and when searching for vocabulary on a topic about which I had very little personal knowledge, it was difficult to evaluate whether a particular retrieved term was closely related to my original term or whether it was only a remote, and therefore useless, relationship that would only confound the results of a bibliographic on-line search. Therefore, I would have had to spend a certain amount of time researching these unfamiliar retrieved terms in order to determine their potential contribution to an on-line search strategy. Since I would have spent time researching database thesaurus and other references as a matter of routine, before compiling a search strategy, this additional searching on VSS would only have represented redundant effort. Another specific criticism I have of VSS is the fact that one is severely restricted in retrieving pertinent terminology since one must initially enter a term or word that is represented in at least one of the vocabularies. Unless you happen to select a word that is part of an actual controlled vocabulary you are out of luck. The system does not appear to have any provision for directing the user to terms that would be more profitable. From my personal experience in thesaurus construction we attempt to keep USED FOR references at a minimum both to save space in our thesaurus and so as not to clutter the thesaurus with what is potentially an infinite number of synonymous expressions for terms. I imagine that other thesaurus constructors have a similar policy. If so, this sets a built-in limitation for VSS. In short, VSS output is only as good and extensive as the vocabularies that compose it, and it cannot serve as a substitute for research in other sources for terminology.

No comment.

Please seek more funding and work more on making the system responsive to someone who does not know how to do preterminal research (using thesauri).

No comment.

No comment.

No comment.

This form (#11) led me to believe that after locating terms I would then do the search and find out whether or not the search was successful. It is difficult to evaluate the usefulness of the terms without doing the search.

I didn't realize you would go down at 5 pm, which is about the time I signed on.

No comments.

No comment.

No comment.

For something like C.A. to be useful, you would need to have the synonyms of the various chemicals and trade names available.

1) In the login instructions, it would be helpful to underline what the user must enter (the actual words and symbols) rather than the phrase "you enter".
2) Since the options available for each vocabulary differ, it would be helpful if one was presented only with those options that applied to the vocabulary chosen, rather than the entire menu of options, some of which do not apply.
3) Since I had a tendancy to want to search all options (browsing, then broader and narrower terms, etc.) for a particular vocabulary term before going on to another term, it might be helpful to consider that sequence in searching rather than the one presently used, requiring that one reenter the vocabulary term each time a new option is chosen.
4) In my evaluation of the search results, I assumed certain terms were applicable to my search though, in fact, I was not sure of some of them, since no definitions were available on-line. It might be helpful, though, to provide the possibility of seeing the scope notes on-line, for those terms for which the meanings are unclear.
5) In the sample searches, in the booklet, it seems a step was left out, after one chooses the maximum number. One then must enter "1 or 2 or 3 or term or command" and it was confusing at first what the #s 1, 2, and 3 referred to until I had done an actual search and found there was a step before this which gave a menu of options from which to select.
6) A better introduction to VSS would be helpful, to explain its function.

No comment.

If a searcher makes a typo or selects wrong menu # there should be an easier way to go back and correct it. I got frustrated trying to get #5 recognized so I could start over. The problem may have been poor phone lines and not the system.

Suggestions for improvement:
a. Verify the spelling of the user by returning the search statement
b. Permit direct change of switching feature on any given term.

c. Switch feature (6) was very helpful; the user can combine the features that are most appropriate for his/her needs and the vocabulary.
d. Would it be possible to move up and down in the MESH tree-structure? That feature would be more helpful than adjective terms.

No comment.

Thanks for including BIOSIS in the study. Looks like we will have to make our index more thesaurus-like! I think the best "searching" happened on the ERIC, PSYCHABS and MEDLINE vocabularies.

# APPENDIX H

# SAMPLE QUESTIONNAIRE FOR ONLINE USERS SURVEY

16. What percentage of your present single data base searches might you perform across multiple data bases if you had a subject switching capability? ____% (71-73)

17. What single factor would most influence your decision to use subject switching in a given search. (Check only one) (74)

- Cost of using subject switching . . . 1 ____
- Number of data bases needed for the search . . . 2 ____
- The differences in data base vocabularies . . . 3 ____
- Other (Please Specify) _____ . . . 4 ____

18. Approximately how much would you be willing to pay for a subject switching capability, and the added cost of more output? (Check only one) (75)

- Nothing at all . . . 1 ____
- Up to $ 5 per search . . . 2 ____
- $ 6 - $10 per search . . . 3 ____
- $11 - $20 per search . . . 4 ____
- $21 - $30 per search . . . 5 ____
- Over $31 per search . . . 6 ____

19. How many data bases would a single query require before you would use the subject switching capability? (Check only one) (76)

- Two . . . 1 ____
- Three . . . 2 ____
- Four or more . . . 3 ____

Comments:





ONLINE USER SURVEY

Zip Code

1. Indicate the type of organization for which you work: (12)

- University or college . . . 1 ____
- For-profit industry . . . 2 ____
- Non-profit industry, association, or research institute . . . 3 ____
- Government agency . . . 4 ____
- Public library . . . 5 ____
- Other (Please Specify) _____ . . . 6 ____

2. Indicate whether you are an end user of information products or an intermediary performing searches for end users: (13)

- End user . . . 1 ____
- Intermediary . . . 2 ____

3. Of the end users for whom you perform searches, approximately what percentage are:

- a. Managerial staff . . . ____% (14-16)
- b. Scientists . . . ____% (17-19)
- c. Faculty . . . ____% (20-22)
- d. Students . . . ____% (23-25)
- e. Librarians/bibliographers . . . ____% (26-28)
- f. Other (Please Specify) _____ . . . ____% (29-31)

| Please estimate: | Number of Searches/mo. | Connect Hrs/mo. | Online Costs/mo. |
|---|---|---|---|
| 4. Your individual search activity | a. ____ (32-35) | b. ____ (36-38) | c. $____ (39-43) |
| 5. Your organizations's search activity | a. ____ (44-47) | ____ (48-50) | $____ (51-55) |

6. Approximately what percentage of the searches you perform are paid for by: (Answers should total 100%)

- a. End users' budget, department or project dollars . . . ____ (56-58)
- b. Library or information center budget . . . ____ (59-61)
- c. Combination of the above . . . ____ (62-64)
- d. Other (Please Specify) _____ . . . ____ (65-67)

7. Approximately what percentage of the searches you perform have the following specific stipulations placed on them:

   a. Limit the search by time coverage, quantity of citations or other output variables . . . . . ___% (68-70)
   b. Stay within a given cost figure . . . . . ___% (71-73)
   c. Use only a given data base or data bases . . . . . ___% (74-76)
   d. Stay within a given connect time . . . . . ___% (77-79)

8. Indicate the frequency with which you use each retrieval system:

| | 1. Frequently | 2. Sometimes | 3. Infrequently |
|---|---|---|---|
| a. Lockheed DIALOG | ___ (12) | ___ (12) | ___ (12) |
| b. SDC ORBIT | ___ (13) | ___ (13) | ___ (13) |
| c. BRS | ___ (14) | ___ (14) | ___ (14) |
| d. NY TIMES | ___ (15) | ___ (15) | ___ (15) |
| e. DDC | ___ (16) | ___ (16) | ___ (16) |
| f. RECON (DOE, NASA) | ___ (17) | ___ (17) | ___ (17) |
| g. KLHILL (NLM) | ___ | ___ | ___ |
| Others: | | | |
| h. _____ | ___ (18) | ___ (18) | ___ (18) |
| i. _____ | ___ (19) | ___ (19) | ___ (19) |

9. Approximately what percentage of your searches are (Answers should total 100%)

   a. Single data base searches . . . . . ___% (20-22)
   b. Multiple data base searches involving 2 data bases . . . . ___% (23-25)
   c. Multiple data base searches involving 3 data bases . . . . ___% (26-28)
   d. Multiple data base searches involving 4 data bases . . . . ___% (29-31)
   e. Don't know . . . . . ___% (32-34)

10. Under what conditions are your searches limited to one data base (Number your choices 1, 2, etc. in order of importance).

   a. Query doesn't require more exhaustive search . . . . . ___ (35)
   b. Too difficult to use other, less familiar data bases . . . ___ (36)
   c. Too time consuming to do multiple data base search . . . . ___ (37)
   d. Too costly to do multiple data base search . . . . . ___ (38)
   e. Too many data bases available to know which ones to use . ___ (39)
   f. Other data bases not available to my organization . . . . ___ (40)
   g. Other (Please Specify) _____ ___ (41)





11. If you are searching on two or more data bases, in which general subject areas is multiple data base searching most useful? (Number your first three choices 1, 2, and 3).

| | |
|---|---|
| ___ (42) Chemistry | ___ (43) Education |
| ___ (44) Physics | ___ (45) Agriculture |
| ___ (46) Mathematics | ___ (47) Energy |
| ___ (48) Engineering | ___ (49) Environment |
| ___ (50) Life Sciences/Medicine | ___ (51) Psychology |
| ___ (52) Business/Economics | ___ (53) Other (Please Specify) _____ |
| | ___ (54) Not known |

12. If you are searching two or more data bases, which data bases do you most frequently combine?

   a. Two data base search (12-17)

   _____
   Name of Data Base (e.g. COMPENDEX) Not Retrieval System

   b. Three data base search (18-26)

   _____
   _____
   _____

   c. Four or more data base search (27-80)

   _____ _____
   _____ _____
   _____ _____
   _____ _____

13. How often do you search with the following:

   a. Controlled vocabulary terms only . . . . . ___% (56-58)
   b. Controlled plus free text terms . . . . . ___% (59-61)
   c. Free text terms only . . . . . ___% (62-64)

14. Indicate your preference below (Please rank "1" for "most preferred", "2" for next most preferred, and "3" for "least preferred").

   a. Controlled vocabulary terms only . . . . . ___ (65)
   b. Controlled plus free text terms . . . . . ___ (66)
   c. Free text terms only . . . . . ___ (67)

15. Given a subject switching capability such as that described in the covering letter, in what percentage of your present searches do you think you might use it? . . . . . ___% (68-70)