Searchers' Selection of Search Keys: II. Controlled Vocabulary or Free-Text Searching

Raya Fidel
Graduate School of Library and Information Science, University of Washington, Seattle, WA 98195

Searching with descriptors from controlled vocabularies complements free-text searching with textwords. The case study method provided data about the manner in which the two types of search keys interact through: (1) observation of 47 professional searchers performing their job-related searches; and (2) analysis of verbal and search protocols, denoting reasons for the selection of each search key and for each search modification. Results show that searchers used thesauri and indexing when it was of satisfactory quality and available to them, and that these and other database-related reasons were the most influential in search-key selection. Further, having to perform a multidatabase search induced the use of textwords without consulting a thesaurus. There is a need for high quality thesauri which are easily available and for mechanisms, such as switching languages, to aid in multidatabase searches.

Introduction

The study of online searching behavior—whether of professional searchers or inexperienced users—is critical to the development of online retrieval systems. Understanding how searchers of all types look for information, and how they interact with existing systems, can provide guidelines for searchers' training and assistance. But more importantly, understanding searching behavior is essential to the design of advanced, user-oriented information systems.

The research project reported here investigated online searching behavior as manifested by *actual* searches of bibliographic databases. It focused on *professional,* experienced online searchers, and explored the selection of search keys, whether descriptors or textwords for free-text searching.

This article is based upon work supported by the National Science Foundation under Grant number IST 85-09719.

Received January 17, 1990; revised April 30, 1990, June 27, 1990; accepted September 7, 1990.

Controlled Vocabulary and Textword Keys

The issue of search key selection has been the focus of many research projects and publications. As Svenonius (1986) points out, the debate over whether controlled vocabulary is necessary for effective retrieval began in the last century, long before the introduction of computers. Although this debate originated from problems encountered when using controlled vocabularies with printed catalogs, the notion that controlled vocabularies are an unnecessary burden on information specialists, as well as on end users, has been the driving force behind much research in recent years.

The construction and use of controlled vocabularies involves a large number of variables, and some theoretical issues have not yet been resolved. For example, there is no agreed-upon measurement for the degree of control exercised in a given index language, nor any well-grounded theories about what constitutes useful indexing practice. Further, controlled vocabularies are expensive to construct and indexing is labor-intensive especially given the alternative of free-text searching where the text is already available and requires only the automated generation of indexes.

Despite the expense and difficulties in the construction of controlled vocabularies, they *are* built and used because they improve retrieval. It is not surprising, therefore, that studies to examine their necessity centered around retrieval performance. Starting with the Cranfield studies (Cleverdon, 1962), investigators have carried out tests to determine which types of search keys provide the best retrieval: textwords or descriptors (Parker, 1971; Keen, 1973; Blair & Maron, 1985). The methods used are sometimes questionable, and results are contradictory; the issue is still unresolved and is heavily debated in the literature (Cleverdon, 1984; Dubois, 1987; Lancaster, 1980).

While some may believe that persistent experimentation will eventually resolve the issue of which type of search keys is best for retrieval, there is increasing evidence that textword and descriptor searching actually *complement* one another, and no single type

 CCC 0002-8231/91/070501-14$04.00

outperforms the other. This relationship has been derived by Fugmann from his theory of indexing (Fugmann, 1982), tested in a set of experiments (Katzer et al., 1982), and substantiated by a series of independent case studies (Carrow & Nugent, 1981; Henzler, 1978; Markey et al., 1980).

The study reported here contributes to the resolution of this controversy by developing a model that represents the rules for the selection of search keys. The model, which is described in a previous article (Fidel, 1991), uncovers such rules and shows that each type of search key is selected for a reason. Thus it substantiates the belief that textword and descriptor searching indeed complement one another, but more importantly, the model shows *how* they complement one another. This article explores the role of free-text searching and of controlled vocabulary in online searching.

The Method

The data for this project were collected through observation of searchers performing their regular, job-related searches and through interviews with them (Fidel, 1984). Using the case study method with controlled comparison (Diesing, 1971), the research team analyzed search protocols, verbal protocols of thought processes while searching, and transcripts of interviews with searchers. The team examined each instance of search-key selection to identify the conditions and reasons that led to the specific selection, and each modification of search strategy to identify its purpose.

In a prior, exploratory study, the method was used to investigate the searching behavior of eight searchers who completed a total of 80 searches. This study resulted in a formal model which describes the selection of search keys and is called the selection routine. To expand and modify the model, the research project reported here investigated an additional group of 39 searchers who completed a total of 201 searches. In addition, searchers in this study were also asked to explain the *reasons* for their selection of each search key, a practice which was not followed in the exploratory study.

The combined studies (47 searchers, and a total of 281 searches) expanded the model to create the selection routine which is a decision tree for the selection of search keys (Fig. 1). The routine describes the conditions that searchers considered and the options that each condition generated. For example, the condition "a searcher does not know if a single-meaning term is mapped to a descriptor," generated the options: use textwords [*P*], use textwords to probe indexing [*Q*], or enter as descriptors terms that might be descriptors [*R*].

The reasons for the selection of each option are displayed in Table 1. In the table, the code of each reason begins with the letter of the option which the reason explains. The second letter designates whether, by its nature, the reason relates to *request* characteristics (*R*), to *database* attributes (*D*), or whether the reason stems from a *searcher's* general beliefs (*S*). For example, the reason [*LD2*] was given when option [*L*] was selected; it is related to a database attribute; and it is the second reason in this category to explain option [*L*]. This reason describes a situation where a textword key was entered for a term that could not be mapped to a descriptor, rather than probing the indexing, because the searcher did not trust the index language and/or the indexing. A detailed description of the selection routine is given in a previous article (Fidel, 1991).

In addition to the creation of the selection routine, data collected in the study were analyzed to (1) determine the frequency in which the searchers who participated in the study selected search keys, options, and reasons; and (2) look for associations among variables that measure these frequencies as well as other variables, as will be explained later.

The descriptive statistics on frequency are based on data collected from 47 searchers performing a total of 281 searches (participants in the first and second studies); the data on reasons for option selection were collected from 39 searchers performing a total of 201 searches (participants in the second study).

The reasons provided by searchers to explain their selection of search keys reflect their perceptions. These perceptions are highly relevant because they guide searchers in their selection of search keys. Because they are subjective, however, these perceptions cannot be used as the sole source of evidence to determine the factors that affect search-key selection; they need to be supported by objective measurements.

To substantiate searchers' perceptions, and to discover additional factors that were not observed by the study's searchers, statistical associations among seven variables were measured. Data for the associations were derived from 281 searches performed by 47 searchers. Most associations were analyzed on two levels: (1) the search level, where each search was considered a distinct instance (a total of 281 instances); and (2) the person level, where the data for each person were averaged so that each person was considered a distinct instance (a total of 47 instances). One should note, however, that the instances on the search level are not independent because several searches were performed by the same person.

The variables that measured characteristics of search-key selection were

(1) *Textwords ratio.* The ratio of textwords selected. Search level: the number of textword keys, divided by the total number of search keys selected for a search. Person level: the total number of textword keys, divided by the total number of search keys selected by a searcher.

(2) *Thesaurus neglect ratio.* The ratio of textwords entered *without* consulting a thesaurus. Search level: the number of textword keys entered during a search without consulting a thesaurus, divided by

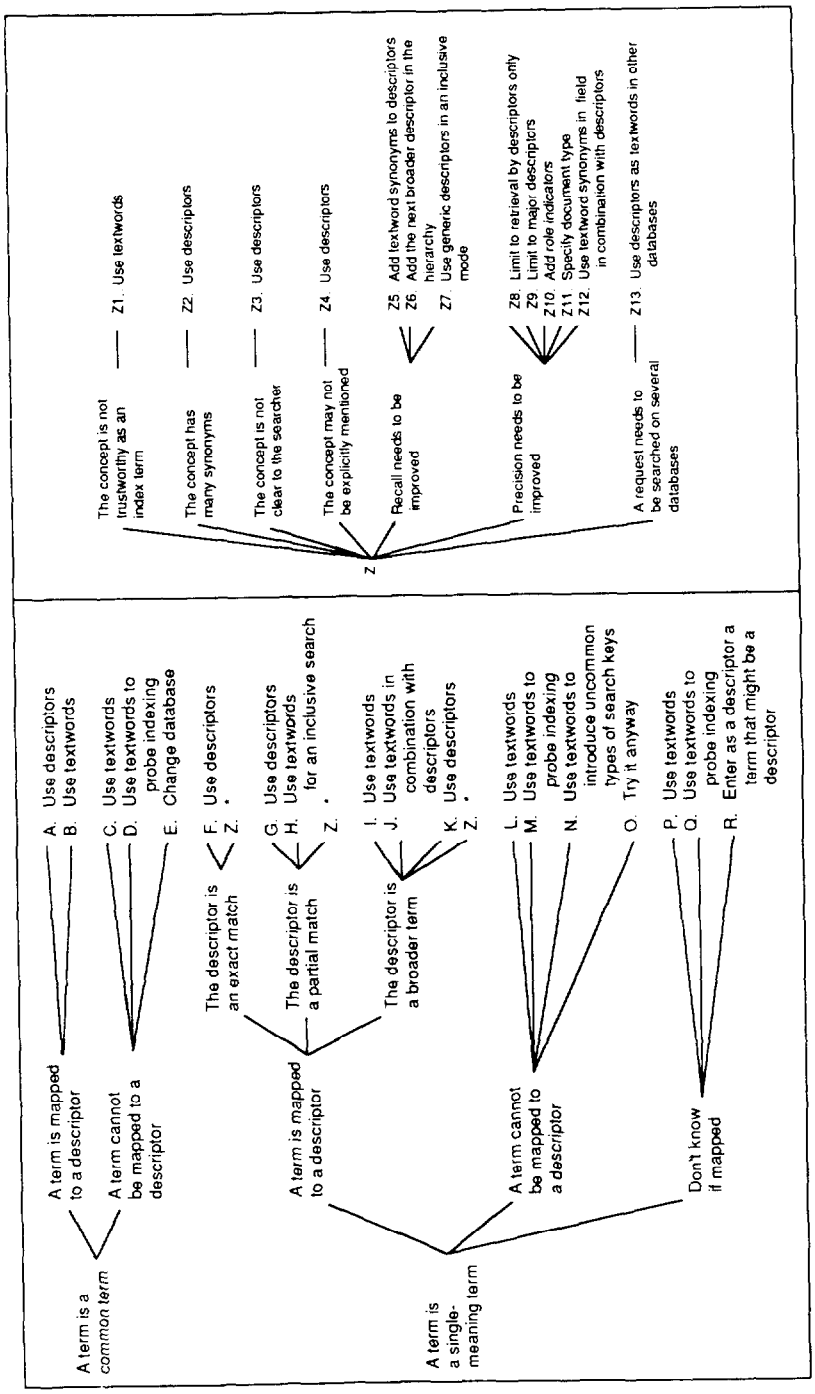

FIG. 1. The selection routine.

TABLE I. Reasons for selecting search keys (for descriptors in bold face).

[*CR*1]	The query formulation includes a relatively large number of components	[*PR*1]	The term was added while online
[*CD*1]	Needed to perform a multidatabase search	[*PR*2]	The searcher does not have time to look for descriptors—"just fishing"
[GR1]	**The term is only added to the formulation to increase recall**	[*PR*3]	The term is used to eliminate irrelevant citations
[GD1]	**The term was found as an index term in relevant articles**	[*PD*1]	The searcher does not trust the descriptors and/or the indexing
[GS1]	**The searcher prefers to use descriptors**	[*PD*2]	Needed to perform a multidatabase search
[*HR*1]	Recall needs to be improved	[*PD*3]	The thesaurus is not available
[*IR*1]	Precision needs to be improved	[*PD*4]	The term would not be in the thesaurus
[*IS*1]	The use of textwords increases recall	[*PS*1]	Terms suggested by users are the best for retrieval
[*JR*1]	Precision needs to be improved	[*PS*2]	The use of textwords increases recall
[*JD*1]	The searcher does not trust the descriptor and/or the indexing	[*QR*1]	The searcher is not sure what descriptors to use
[KR1]	**Recall needs to be improved**	[*QD*1]	The thesaurus is not available
[KR2]	**Precision needs to be improved**	[*QS*1]	The searcher prefers to start with textwords and then check descriptors
[KR3]	**The term is used as a limiting factor**	**[RR1]**	**The term was added while online**
[KD1]	**The term was found as an index term in relevant articles**	**[RR2]**	**Needed to perform a multidatabase search**
[KS1]	**The searcher prefers to use descriptors**	**[RD1]**	**The term is a descriptor in another database**
[*LR*1]	Most specific retrieval is desired	**[RD2]**	**The thesaurus is not available**
[*LR*2]	The term is specific and well-defined	**[RD3]**	**The searcher "knows" the terms are descriptors**
[*LR*3]	The term appeared in titles and abstracts of relevant articles	[*Z5R*1]	Recall needs to be improved
[*LR*4]	Recall needs to be improved	[*Z5R*2]	User insisted on using the terms
[*LR*5]	The searcher had gotten poor retrieval using related descriptors	[*Z5D*1]	Needed to perform a multidatabase search
[*LD*1]	The term would not be in the thesaurus	[*Z5D*2]	The searcher does not trust descriptors and/or the indexing
[*LD*2]	The searcher does not trust the descriptors and/or the indexing	**[Z6R1]**	**The user will be interested in the broader descriptor as well**
[*LS*1]	If a term represents a concept accurately and it is not mapped to a descriptor, there is no need to probe indexing	**[Z7R1]**	**The query formulation includes a relatively large number of components**
[*LS*2]	The use of textwords increases recall	**[Z7D1]**	**Recall needs to be improved**
[*LS*3]	Terms suggested by users are the best for retrieval	**[Z8R1]**	**To make sure that the concept is central to articles**
[OD1]	**The term might have been added to the thesaurus**	**[Z9R1]**	**The size of the set needs to be reduced**
[OD2]	**The term is a descriptor in another database**	**[Z9R2]**	**To make sure that the concept is central to articles**
		[*Z*12*R*1]	To extract a subset that includes citations that are highly relevant
		[*Z*12*D*1]	The searcher does not trust the descriptors and/or the indexing

the number of search keys selected for the search. Person level: the total number of textwords entered by a searcher without consulting a thesaurus, divided by the total number of search keys entered by the searcher.

The variables that measured factors that may affect the selection of search keys were

(3) *Number of search keys.* Search level: the number of keys selected for a search. Person level: the average number of search keys selected by a searcher per search.
(4) *Number of databases.* Search level: the total number of databases added during a search. Person level: the average number of databases added per search.
(5) *Number of moves.* Search level: the total number of search-strategy modifications made during a search. Person level: the average number of search-strategy modifications made by a searcher per search.
(6) *Subject area.* The subject area in which a searcher specializes (person level only). This variable had four values: medicine, sciences, social sciences (to cover both the social sciences and the humanities), and general (for searchers who habitually search requests in a variety of subjects, as is often the case in public libraries or for independent consultants).
(7) *Environment.* The environment in which a searcher works (person level only). This variable had three values: practical environments, theoretical environments, and general environments. A practical environment is a working place where searchers are usually called upon to search requests that result from immediate and practical problems, such as most small or medium-size consulting companies or industries. In contrast, a theoretical environment is an establishment whose users are often involved in research or investigation, for instance, universities, or regulatory agencies. Search environments that could not be assigned any of the first two categories were called general environments.

The results of these analyses clearly point to the significant role of controlled vocabularies in online searching.

Frequency of Search-Key Selection

The distribution of search-key selection for the 281 searches performed in the study is presented in the following table:

Number of search keys	All databases	Databases with thesauri
Descriptors	1,607 (44%)	1,607 (50%)
Textwords	2,028 (56%)	1,582 (50%)
Total	3,635 (100%)	3,189 (100%)

The results show that of the 3,200 search keys selected for searching databases that offer controlled vocabulary, 50 percent were descriptors and 50 percent were textwords. Thus: ***Searchers as a group did not show a preference for one type of search key: When they had a choice, they selected descriptors and textwords with the same frequency.***

This finding points to a general trend among searchers and does not characterize any individual searcher, some of whom may prefer textwords and others descriptors (cf. [*GS*1] or [*LS*3] in Table 1).

Data further show that: ***Searchers consulted a thesaurus for 75% of the search keys they selected for databases with controlled vocabulary.*** This finding attests to the significance of thesauri to online searching. It sends a clear message to database designers: Controlled vocabularies and indexing should not be eliminated from bibliographic databases because when searchers select search keys, they consult a thesaurus three-quarters of the time, and use descriptors at least half of the time.

Reasons for Selecting Textwords

The selection routine is a way of analyzing the terminological considerations that lead searchers to the selection of search keys. It shows, for example, that if a single-meaning term is mapped to a descriptor through a partial match, searchers may use the descriptor, or textwords in an inclusive search, but if the concept has many synonyms they may select the descriptor (Fig. 1).

Some reasons given by searchers for their search-key selection, however, are extraterminological (Table 1). An analysis of these reasons is useful in revealing additional conditions under which searchers think it "best" to use textwords or descriptors, at least as reflected by the searching behavior of the searchers who participated in the study.

To discover the conditions that are "best" for textword searching and those that are best for descriptors, we first isolated reasons that are common to both types of search keys. Examining these reasons, we discovered several instances of searching behavior that are contradictory to common knowledge about online searching. We then examined reasons that are exclusive to one or the other type of search key.

Request Characteristics. Searching lore has it that searching with textwords would result in high recall, while descriptor searching secures high precision. This notion is prevalent even though thesauri were first introduced to improve recall. The reasons given by the searchers in the study show that this rule does not govern their selection of search keys: both types of search keys were used to increase recall and precision. To increase recall, either textwords ([*HR*1], [*LR*4], [*Z5R*1]) or descriptors ([*GR*1], [*KR*1], [*Z7D*1]) can be used, depending on terminological conditions. Similarly, precision can be increased either with textwords ([*JR*1], [*JR*1],

[*LR*1],[*Z*12*R*1]) or with descriptors ([*KR*2],[*Z*8*R*1], [*Z*9*R*1],[*Z*9*R*2]).

To further test the belief that the use of textwords increases recall, we measured the association between two variables: recall tendency and textwords ratio. Recall tendency was defined as the ratio of moves to increase recall over all moves. On the search level this variable represents the degree to which a search required improved recall, and on the person level it represents the degree to which a searcher is habitually concerned with improving recall.

The working assumption was that if searchers act on the belief that the use of textwords always increases recall, then searches in which recall needed to be improved would show a relatively high use of textword keys. Similarly, searchers who habitually seek to improve recall would use a relatively high percent of textword keys. Under this assumption, the variables recall tendency and textwords ratio should correlate. The Pearson Product-Moment Correlation test, however, shows no such association (search level: $r(279) = .060$, NS; person level: $r(45) = .104$, NS). That is, textword keys are not used with significantly higher frequency to increase recall than descriptors. Nevertheless, the general *belief* that the use of textwords increases recall was mentioned by several searchers as a reason for the selection of a textword key ([*IS*1],[*LS*2],[*PS*2]).

A similarly common assumption is that terms that are entered during the session at the terminal are mostly textwords. Here again, the assumption is *not* supported by the results of this study. The reasons mentioned by searchers show that both types of search key might be added while online ([*PR*1],[*RR*1]). Searchers may add textwords while online because the terms appear in titles or abstracts of relevant citations ([*LR*3]), or they may choose to enter descriptors assigned to such citations ([*GD*1],[*KD*1]).

Requests with a relatively large number of components often require special consideration, but usually could usefully employ both types of search keys. For instance, searchers explained that they entered a common term not mapped to a descriptor as a textword key because the request contained a relatively large number of components ([*CR*1]). But, the same reason was used by searchers who attempted to increase recall by using a generic *descriptor* in an inclusive search ([*Z*7*R*1]), entering the descriptor as well as its narrower terms. The preceding examples show that: ***Depending on terminological conditions, the same reason may sometimes result in the use of textwords and at other times, the use of descriptors.***

A number of request-related reasons pertain *only* to the use of textwords. Given specific terminological conditions, searchers may enter textwords (but never descriptors) if:

- A term cannot be mapped to a descriptor, but it is specific and well-defined ([*LR*2]);
- A term cannot be mapped to a descriptor and poor results were obtained using related descriptors ([*LR*5]);
- It is not known if a term is mapped to a descriptor and the searcher is "just fishing" and did not consult a thesaurus ([*PR*2]);
- It is not known if a term is mapped to a descriptor and the term is used to eliminate irrelevant citations and the searcher did not consult a thesaurus ([*PR*3]);
- A term is mapped to a descriptor but textword synonyms are added to increase recall because the user insisted on using certain terms ([*Z*5*R*2]).

Database Characteristics. Factors characterizing databases, their thesauri and indexing also affected the selection of search keys. Two of the most important were availability and quality. Although searchers entered either textwords *or* descriptors when a thesaurus was not on hand ([*PD*3],[*QD*1],[*RD*2]), they used *only* textwords when they did not trust the descriptors and/or the indexing ([*JD*1],[*LD*2],[*PD*1],[*Z*5*D*2],[*Z*12*D*1]), or when they thought that a term would not be in the thesaurus ([*LD*1],[*PD*4]).

On the other hand, there were reasons that generated the exclusive use of descriptors. At times, searchers entered a term as a descriptor even when it was not listed in the thesaurus, assuming that it might have been added ([*OD*1]), or because it was a descriptor in another database ([*OD*2],[*RD*1]). Sometimes, searchers entered a term as a descriptor without consulting the relevant thesaurus because they "knew" that the term was a descriptor ([*RD*3]).

Searchers' Beliefs. At a more general level, searchers' individual tendencies and beliefs affected the selection of search keys as well. As expected, some searchers prefer to use descriptors ([*GS*1],[*KS*1]) and others prefer textwords. We solicited specific explanations about the choice of textwords. Though a general preference to starting with textwords before checking descriptors was one of the reasons ([*QS*1]), other reasons provided technical rationalizations. Searchers claimed that they used textwords because:

- The use of textwords increases recall ([*IS*1],[*LS*2],[*PS*2]).
- Terms suggested by the users are the best to use ([*LS*3],[*PS*1]).
- If a term represents a concept accurately and it is not mapped to a descriptor, there is no need to probe indexing ([*LS*1]).

In summary, the reasons provided by searchers for their selection of search keys show that while searchers may prefer one type of search key to the other, explanation of their behavior as represented by searching lore is too simplistic most of the time. Terminological conditions, request characteristics, and the availability and quality of databases and thesauri, all interact to affect the selection of search keys.

TABLE 2. Frequency of option selection.

Option	No.	(%)[a]	(%)[b]
[A]	2	.05	.06
[B]	6	.16	.18
[C]	13	.36	.40
[D]	1	.03	.03
[E]	1	.03	.03
[F]	1122	30.86	35.18
[G]	44	1.21	1.37
[H]	13	.35	.41
[I]	3	.08	.09
[J]	22	.60	.69
[K]	96	2.64	.10
[L]	972	27.62	16.49
[M]	16	.44	.50
[N]	1	.03	.03
[O]	8	.22	.25
[P]	631	17.36	19.79
[Q]	34	.93	1.07
[R]	141	3.88	4.42
[Z1]	10	.27	.31
[Z2]	1	.03	.03
[Z3]	1	.03	.03
[Z4]	1	.03	.03
[Z5]	302	8.31	9.47
[Z6]	6	.16	.19
[Z7]	146	4.01	4.58
[Z8]	1	.03	.03
[Z9]	31	.85	.97
[Z10]	1	.03	.03
[Z11]	1	.03	.03
[Z12]	2	.05	.06
[Z13]	14	.38	.44

[a]Percent of all search keys selected.
[b]Percent of search keys selected for databases with thesauri.

Frequency of Option Selection

Table 2 lists the frequency of option selection. The first four columns list: (1) the option; (2) the number of times it was selected; (3) the relative frequency with which it was selected in *all* databases; and (4) the relative frequency with which it was selected in only those databases that have controlled vocabularies.

Findings show that the most frequent options were:

- [F] use descriptors when a single-meaning term is mapped to a descriptor through an exact match (35% of all options);
- [P] use textwords when it is not known whether a single-meaning term is mapped to a descriptor (20%); and
- [L] use textwords when a single-meaning term cannot be mapped to a descriptor (16%).

A summary of these findings shows that: ***Over 70% of the time, searchers selected the most straightforward options, that is: If a term was mapped to a descriptor exactly, they entered a descriptor; and if it could not be mapped, or when they did not consult a thesaurus, they entered textwords.***

This finding suggests that 70% of the time that they were selecting search keys, searchers did not perceive that they had terminological difficulties, regardless of whether or not they consulted a thesaurus, and whether they entered a textword or a descriptor. While some readers may conclude that the study's searchers performed simple searches, this study took a different approach: It did not attempt to check whether the searchers' perceptions were "correct." The basic assumption was that experienced searchers are most knowledgeable about the art of online searching. Adopting this assumption, we can predict, therefore, that 70% of the terms to be represented in a query formulation are likely to present no terminological difficulties.

Findings also show that searchers were quite successful in locating descriptors that matched their terms. Only 20% of the times that they consulted a thesaurus, did they fail to find a match, and the frequency with which they were successful in finding an *exact* match is twice that (44%).

This does not mean, however, that 80% of request terms are likely to be matched to descriptors. We must first examine the use of thesauri. As is evident from the reasons given for choosing options, searchers were *selective* in the use of thesauri: At times they neglected to consult a thesaurus because they felt it would not be useful. (As we see later, the quality of thesauri had a significant effect on the selection of search keys.) It is plausible to assume, therefore, that searchers consulted only the thesauri they deemed to be useful. It would follow then: ***When thesauri of relatively good quality are consulted, one can expect to match 80% of the terms for which a thesaurus is consulted.***

From a different point of view, this finding suggests that among thesauri available today, a "good" one provides for at least an 80% match between request terms and descriptors.

Frequency of Reasons for Option Selection

Table 3 provides data about the reasons for selecting a certain option for those conditions that produce more than one option. These data were derived from 39 searchers. The first column lists the option and the second the category which represents the reason: whether the reason was related to a *request,* the *database,* or a *searcher.* The third column tallies the total number of times that reasons in a particular category were given for the option. The fourth column gives the percentage of each category within the option. The next column shows the code for each individual reason, followed, in the next column, by the number of times the reason was mentioned. The last column represents the percentage for each reason within its category.

It should be noted that the total number of reasons associated with a particular option was frequently dif-

TABLE 3. Frequency of reasons for option selection.

Option	Category	No.	(%)	Reason	No.	(%)	Option	Category	No.	(%)	Reason	No.	(%)
[C]	Request	3	23	CR1	3	100						101	100
	Database	10	77	CD1	10	100		Database	461	57	PD1	129	28
		13	100								PD2	117	25
[G]	Request	19	68	GR1	19	100					PD3	108	24
	Database	3	11	GD1	3	100					PD4	107	23
	Searcher	6	21	GS1	6	100						461	100
		28	100					Searcher	241	30	PS1	179	74
[H]	Request	13	100	HR1	13	100					PS2	62	26
[I]	Request	3	75	IR1	3	100						241	100
	Searcher	1	25	IS1	1	100			803	100			
		4	100				[Q]	Request	29	76	QR1	29	100
[J]	Request	1	50	JR1	1	100		Database	4	11	QD1	4	100
	Database	1	50	JD1	1	100		Searcher	5	13	QS1	5	100
		2	100						38	100			
[K]	Request	36	77	KR1	28	78	[R]	Request	14	9	RR1	13	93
				KR2	6	16					RR2	1	7
				KR3	2	6						14	100
					36	100		Database	140	91	RD1	57	40
	Database	2	4	KD1	2	100					RD2	42	30
	Searcher	9	19	KS1	9	100					RD3	41	30
		47	100									140	100
[L]	Request	187	49	LR1	84	45			154	100			
				LR2	66	35	[Z5]	Request	119	72	Z5R1	117	98
				LR3	27	15					Z5R2	2	2
				LR4	8	4						119	100
				LR5	2	1		Database	46	28	Z5D1	35	76
					187	100					Z5D2	11	24
	Database	105	28	LD1	99	94						46	100
				LD2	6	6			165	100			
					105	100	[Z6]	Request	5	100	Z6R1	5	100
	Searcher	89	23	LS1	56	63	[Z7]	Request	6	10	Z7R1	6	100
				LS2	17	19		Database	52	90	Z7D1	52	100
				LS3	16	18			58	100			
					89	100	[Z8]	Request	1	100	Z8R1	1	100
		381	100				[Z9]	Request	11	100	Z9R1	6	55
[O]	Database	4	100	OD1	3	75					Z9R2	5	45
				OD2	1	25						11	100
		4	100		4	100	[Z12]	Request	2	67	Z12R1	2	100
[P]	Request	101	13	PR1	62	61		Database	1	33	Z12D1	1	100
				PR2	29	29			3	100			
				PR3	10	10							

ferent from the number of times that option was selected. There are two sources for this discrepancy. First, the data for the number of times an option was selected were derived from observing a group of 47 searchers, while only 39 searchers from this group contributed the data for the reasons a given option was selected. Thus, for example, while option [*G*] was selected 44 times, the total number of reasons for selecting this option is only 28. Second, a selection of an option may be caused by more than one reason. A searcher may decide, for example, to select a textword key because she believes that textwords increase recall (a searcher-related reason), but also because she does not trust the indexing (a database-related reason). For instance, option [*I*] has only three instances, but four reasons.

A summary of the reasons used for the selection of search keys shows the following distribution:

Category of reasons	Number (percentage)
Database-related reasons	829 (48%)
Request-related reasons	553 (32%)
Searcher-related reasons	351 (20%)
Total	1,733 (100%)

That is: ***When searchers had options in the selection of search keys, their choice was most frequently (48% of the time) determined by the databases they were searching and least frequently (20%) by their habitual searching behavior.***

This result suggests that while searchers develop habits, these habits do not dominate their selection of search keys; their primary considerations are the characteristics of the requests and of the databases they search most frequently. It also shows that database characteristics is the factor having the largest impact on the selection of nonstraightforward search keys.

At first glance, this finding is not surprising. After all, as mediators between users and databases, searchers *should* examine the database they search before they select a search key, and then select keys that are most useful for the database. However, the fact that the percentage of request-related reasons was so much lower than that of database-related reasons is disturbing. Ideally, the selection of search keys should be determined primarily by request characteristics, and databases should be designed so flexibly that they free searchers to consider first the requirements of each request. Clearly, current databases do not approximate this ideal, and the need for advanced, more usefully designed databases is apparent.

To shed more light on the reasons for selecting textwords, the association between the variable textwords ratio and the other variables was measured.

Textwords Ratio

Textwords ratio (the percentage of textwords selected) measures the degree to which textwords were used in a search, and the general preferences of searchers in the selection of search keys. This variable is primarily associated with three variables: number of databases, subject area, and environment for science searchers. Further, textwords ratio *does not* correlate either with number of search keys or with number of moves.

The Number of Databases. The variables textwords ratio and number of databases are directly related (search level: $r(279) = .277, p < .01$; person level: $r(45) = .414, p < .01$). That is: ***Searches which require several databases, and searchers who habitually search several databases for a request, are likely to use more textwords than searches or searchers using a single database.***

This correlation was expected: A search that spans a number of databases is likely to include more textwords than descriptors because it is time consuming to look for descriptors for each database. For the same reason, searchers who usually search a number of databases for each request are likely to develop a habit of using more textwords than descriptors.

The association between textwords ratio and number of databases warrants an examination of possible causal relationships. While searchers are free to choose whether to enter textwords or descriptors, the number of databases to search for a request is determined by the distribution of information among the databases; it is a given. Undoubtedly, a searcher's preference for a search-key type (textword or descriptor) can determine the number of databases he searched because it is plausible to assume that a searcher who preferred searching with textwords would move from one database to another more easily than one who preferred to use descriptors. But even so, a free-text searcher would change databases only when it was required for the success of a search. In addition, 20% of the time searchers explained their selection of textwords with reasons that related to attributes of databases, they mentioned the need to perform a multidatabase search as the reason for their search-key selection. Moreover, having to search a number of databases even caused some searchers to enter textwords that were inadequate for free-text searching, i.e., common terms (reason [*CD*1] in Table 1). The causal relationship is, therefore, clear: ***Having to search several databases for a request induces the use of textwords.***

The Subject Area of a Searcher. Analysis of variance shows that the variable subject area correlates with the textwords ratio ($F(3, 43) = 13.16, p < .01$). On the average, the percentage of textwords used in each subject's literature is as follows:

Subject area	Textwords ratio for searcher (%)
Medicine	34
Social sciences and humanities	39
General	57
Science and technology	76

A post-hoc test shows that the difference lies between science and both medicine and the social sciences searchers. Textwords ratio for general literature did not differ significantly either from social sciences and humanities or from science. That is: ***Science searchers are more likely to use textwords than their colleagues who specialize in other subject areas.***

At first glance, this finding seems to support common belief. It has been long assumed that searches in the scientific literature do not require the use of controlled vocabulary because the scientific terminology itself is already controlled. Note that this argument is not completely valid because it ignores the process of indexing, which is performed mostly with controlled vocabulary but which accomplishes additional functions such as assigning explicit terms to represent concepts which are only implicit in the text.

However, even if accepted, this argument would not be a valid explanation for this finding because of the difference between science and medical searchers. Medical terminology *is* scientific terminology, and there is no evidence to assume that it is less controlled than other scientific terminologies, no matter how one defines or measures the degree of terminological control. Yet, medical searchers used the smallest proportion of textwords while science searchers used the largest proportion.

This discrepancy could be partially explained by the typical number of databases that is required for a search in each subject area: Searches of the medical literature required an average of 1.33 databases for a search, while those of the science literature 2.64. The results of this study also indicate that this difference in the use of textword keys relate to the quality of databases in the two subject areas, as will be explained later.

Environment. The nature of the environment, across all subject areas, has no significant effect on textwords ratio ($F(2, 44) = .69$, NS). However, analysis of variance shows that for those who search the scientific literature, the searcher's environment has a significant effect on this variable ($F(1, 21) = 7.43, p < .05$). Science searchers who typically answer requests that address practical problems used textwords 87% of the time; those who typically search for theoretical requests used textwords 67% of the time. That is: ***Science searchers who typically answer practical questions are more likely to use textwords than science searchers who usually address theoretical problems.***

The finding that environment in general does not affect the textwords ratio, but has an effect within science searching, may imply that the subject area has a larger effect on the selection of search keys than whether the requests searched are of practical or theoretical nature.

However, it is plausible to speculate that within each subject area, practical questions encourage the use of textwords because they are likely to include concrete and well-defined terms that are adequate for free-text searching. The failure of this study to find such an association for subject areas other than the sciences may be due to deficient sampling: the samples of searchers within other subject areas were small and therefore possibly not representative enough. This result suggests a hypothesis to be tested: ***Within a subject area, the nature of a request, whether it is practical or theoretical, may affect the percent of textwords selected.***

Number of Search Keys. The percent of textwords selected does not significantly correlate with the number of search keys (search level: $r(279) = -.016$, NS; person level: $r(45) = -.166$, NS). This association leads to the conclusion that: ***Searchers who prefer to use textwords and those who prefer descriptors use, on the average, the same number of search keys.***

This finding contradicts a wide-spread assumption that when searchers use textwords they are likely to use more terms than when they use descriptors, because with textwords they can choose any term that seems relevant to them. While this is a sound assumption, it is not supported by the data collected in this study. This result shows, then, that one of the assumed advantages of free-text searching does not hold in real-life searching.

Further, to account for synonyms, searchers *should* enter more terms when they use textword keys than when they use descriptors. The finding that searchers who prefer to use textwords do not enter, on the average, more search keys than their counterparts highlights the essential role of controlled vocabularies and of indexing. One of the central purposes of vocabulary control is to control for synonyms. Thus, instead of searchers having to exercise terminological control while searching by thinking up all relevant synonyms for a concept, control is conducted at the design stage and each concept is represented by only one term. This finding suggests, therefore, that searchers who prefer textwords *do not* exercise vocabulary control in searching because if they did, the average number of search keys they use would have increased. This conclusion is further supported by the observation that the searchers rarely used a thesaurus as a source for synonyms for a free-text search.

While it is easy to conclude that searchers should perform their searches more thoroughly, this notion warrants the attention of designers of database and of expert systems. If, in using existing systems, searchers do not exercise vocabulary control in searching (and whether they shy away from it because they feel inhibited or because it takes a special talent to do so while searching under cost constraints, is immaterial), database designers should encourage the use of thesauri by designing easy-to-use thesauri that are reliable sources for synonyms. Further, intermediary expert systems could be designed to help searchers in terminological control.

Interaction During a Search. The level of interaction during a search can be measured by the number of times a search strategy is modified; that is, by the number of moves. The variable number of moves does not significantly relate to textwords ratio (search level: $r(279) = .104$, NS; person level: $r(45) = -.030$, NS). Therefore: ***Interaction during a search does not increase the proportion of textwords.***

Coupled with the finding that textwords ratio is not associated with number of search keys, this result might seem somewhat surprising. It is sound to assume that the mechanics of the search process itself would determine the ratio of textwords used. Interactive searches, as well as the need to use a relatively large number of search keys, theoretically would require a large proportion of textwords because searchers presumably do not have time to look for descriptors in a thesaurus during online interaction. In contrast, our results show that neither number of moves per search nor number of search keys correlates with textwords ratio. That is, during their interactions and when they add search keys, searchers are likely to select textwords with the same frequency that they are likely to select descriptors.

This finding is not totally surprising because searchers gave the reason of being online when entering a search key to explain both the selection of textwords and of descriptors, as explained earlier. If we consider the mechanics of a search more closely, we can see that during their interaction, searchers use displayed records as a source for additional search keys. This finding suggests, then, that when examining these records, searchers select textwords from titles or abstracts and descriptors from the displayed index terms with the same frequency.

Reasons for Neglecting to Consult a Thesaurus

Of particular interest in the study's finding are instances where searchers decided to enter textwords without even checking a thesaurus (option [*P*] in Figure 1). In such instances searchers decide *not to have a choice* in the selection of search-key type. Further, this is not an obscure phenomenon: 37% of the textwords selected to search databases with indexing were picked without thesaurus consultation. It is useful, therefore, to spell out the reasons searchers cited to explain their decision to avoid consulting a thesaurus.

There were 803 instances in which searchers cited reasons for the option of not consulting a thesaurus. Of these, 179 times (22%) it was because they held a general belief that entering the user's terms directly would give more relevant citations. While this was the most cited reason for this option, over half of the reasons given for the option (57%) were related to the databases searched:

Reason avoiding a thesaurus	Times
Don't trust the thesaurus or the indexing	129 (16%)
The term would not be in the thesaurus	107 (13%)
Had to perform a multidatabase search	117 (15%)
Had no access to the relevant thesaurus	108 (13%)
Total database-related reasons	461 (57%)

That is: ***Over half the time searchers neglected to consult a thesaurus, they did so either because they did not trust the quality of the thesaurus, because the thesaurus was not available, or because they had to search several databases for a request.***

In other words, thesaurus availability and quality, as well as the need to search several databases simultaneously are important factors in the selection of search keys: Thesauri of poor quality or limited availability, as well as the need for multidatabase searches encourage searchers to enter textwords without checking first whether or not appropriate descriptors exist.

To further substantiate this conclusion, and to discover reasons for neglecting to consult a thesaurus that were not mentioned by searchers, associations between the variable thesaurus neglect ratio and the other variables were measured.

Thesaurus Neglect Ratio

Thesaurus neglect ratio measures the percentage of textwords entered without consulting a thesaurus. The data show that this variable correlates with four other variables: Number of databases, subject area, number of search keys, and number of moves.

Number of Databases. Thesaurus neglect ratio and number of databases required for a search are directly related (search level: $r(279) = .294, p < .01$; person level: $r(45) = .397, p < .01$). This association was expected in that a multidatabase search was cited as a reason for not consulting a thesaurus over a quarter of the times when database-related reasons were given for this option. This association shows that: ***The larger the number of databases to be searched per request, the more likely is a searcher to neglect consulting a thesaurus.***

Since searchers used the reason of having to search multidatabases to explain their decision to avoid thesaurus consultation 13% of the times they elected this option, the effect of number of databases on thesaurus consultation deserves special attention. While some searchers may feel comfortable using several databases for a search *because* they habitually refrain from consulting a thesaurus, for others: ***Having to search several databases for a request induces entering textwords without consulting a thesaurus.***

Multidatabase searches, then, have adverse effects on searching behavior. These effects could be eliminated to a certain degree by increased standardization

in database design or by the construction of switching languages, as discussed later.

Subject Area. The subject area of searching has a significant effect on the frequency of neglecting to consult a thesaurus ($F(3, 43) = 3.51, p < 0.05$). The average frequencies for entering search keys without consulting a thesaurus for each subject area are revealing:

Subject area	Thesaurus neglect ratio (%)
Medicine	0
Social sciences and humanities	13
General	29
Science and technology	32

That is: ***Science searchers are more likely to enter textwords without consulting a thesaurus than searchers who specialize in other subject areas.***

This conclusion concurs with a previous finding: science searchers are more likely to use textwords than their colleagues.

Further, generalist searchers (who habitually search several subject areas) entered a significantly larger number of search keys without consulting a thesaurus than did their peers in the social sciences and medicine. This phenomenon can be explained by the fact that generalists search a relatively large number of distinct databases throughout their searching practice. It is plausible to assume, then, that generalists search the largest number of distinct databases, even though this factor was not measured in this study. They cannot familiarize themselves with the thesauri of the many databases they search and they are, therefore, more likely to refrain from using a thesaurus.

Number of Search Keys. The variable thesaurus neglect ratio directly relates to number of search keys *only* on the search level ($r(279) = .359, p < .01$), and not on the person level ($r(45) = -.164$, NS). This means that if searchers decide to increase the number of search keys for a particular request, they are likely to add terms without consulting a thesaurus, but searchers who habitually use a large number of search keys consult a thesaurus in the same frequency that other searchers do.

Since thesaurus consultation does not relate to personal inclination in the number of search keys used, the association between thesaurus neglect ratio and number of search keys is induced by the nature of specific requests and by situational searching conditions. While no causal relationships emerge clearly, it is plausible to suggest that: ***Requests that require a relatively large number of search keys may lead searchers to enter search keys without consulting a thesaurus.***

Number of Moves. The number of search-strategy modifications relates directly to thesaurus neglect ratio *only* on the search level ($r(279) = .318, p < .01$), and not on the person level ($r(45) = .003$, NS). That is, when an individual search requires a relatively large number of modifications, searchers are likely to avoid consulting a thesaurus; however, interactive searchers, those who habitually make more modifications than their peers, do not neglect to use a thesaurus more frequently than their colleagues. Since the association between thesaur neglect ratio and number of moves is induced by specific requests and by situational searching conditions, it is plausible to suggest that: ***Interactive searches cause searchers to avoid consulting a thesaurus.***

Discussion

The results of this study point to the significant role that high quality and easily available thesauri play in online searching. In addition, the results reveal the need for assistance in performing multidatabase searches.

Thesaurus Quality and Availability

The most important contribution to the debate about the need for controlled vocabulary is the finding that searchers used thesauri and indexing when it was of satisfactory quality and easily available to them. This finding is supported by various results.

First, searchers relied heavily on thesauri: They consulted them for 75% of the search keys they selected, and 80% of the times they consulted a thesaurus they selected a descriptor. Moreover, when searchers avoided consulting a thesaurus, they most often did so for a reason specific to the database they were searching, its thesaurus or its indexing.

Second, although some searchers preferred to use descriptors, and others textwords, generally speaking, searchers did not prefer one type of search key to another. When they searched databases that had thesauri, they selected descriptors and textwords in the same frequency. That is, eliminating controlled vocabulary would have prevented the searchers from entering their preferred choice of search key half of the time.

Third, the study results suggest that often searchers who prefer textwords do not compensate for the lack of vocabulary control by using synonyms; they use, on the average, the same number of search keys used by those who prefer descriptors. This implies that while theoretically possible, terminological control in searching is not practical. It would seem then that the main vehicles for terminological control are controlled vocabularies used for indexing or thesauri used for term expansion in searching.

Fourth, particularly revealing are the findings that relate to *un*availability of thesauri. Statistical associations among the study variables indicate that unavailability of a thesaurus may increase number of search keys and number of moves in a search. That is, unavailability of thesauri may increase the effort necessary to perform a search.

Fifth, when searchers had more than one option in the selection of search keys, their choice of a term was most frequently determined by the databases they were searching. Moreover, when they cited a database attribute as a reason for the selection of a certain search key, 62% of the time they referred to *deficiencies* in the quality of thesauri and indexing and in their availability. In addition, 16% of the instances in which searchers decided to enter a term without consulting a thesaurus at all, they did so because they believed the indexing and/or the descriptors would not be useful.

The tendency among searchers to avoid consulting thesauri that are not useful was further inferred from the finding that some specific thesauri were heavily consulted, while others were ignored most of the time. This finding, together with the finding that science searchers enter textwords and neglect to consult a thesaurus more frequently than their peers who specialize in other subjects, show that trusting a thesaurus and the indexing of a database is not an idiosyncratic perception of an individual searcher, but rather a common perception among searchers.

The most influential factor in the selection of a search key was the database, and databases achieved this top position partly because they *did not* provide what was expected of them. In other words, the lack in the quality of thesauri and indexing, and in their availability, are significant hindrances in searching.

At present, it is not clear what specific problems searchers have with descriptors and indexing of individual databases. It is clear, however, that higher quality standards in thesauri and indexing, as well as better availability of thesauri, are badly needed. It is important, therefore, to begin studying specific deficiencies searchers find in existing thesauri and indexing. Such studies would facilitate the establishment of recognized and agreed-upon standards to determine the quality of thesauri and indexing, which in turn would guide the creation of improved controlled vocabularies and indexing operations.

In summary, there is enough evidence that controlled vocabulary cannot be dismissed as "not cost-effective," or archaic. Even though it is still unclear whether searching with descriptors outperforms free-text searching, this study clearly shows that searchers rely heavily on controlled vocabularies and require that they be of high quality.

Multidatabase Searches

This study shows that in addition to deficiencies in the quality and availability of thesauri, searchers neglected to consult a thesaurus when they had to perform a multidatabase search. Statistical tests reveal that having to search several databases for a request induces the use of textwords, in particular the use of textwords without consulting a thesaurus. Further, having to search a number of databases was cited as a reason for entering textwords and for not consulting a thesaurus.

This is understandable because searchers usually do not have the time to develop a separate search strategy for each database they plan to search. Because most thesauri are not available online, and searching those that can be accessed online is very costly, libraries that routinely perform multidatabase searches have to acquire a number of printed thesauri. Most small and medium-size libraries cannot afford such a purchase. Usually, the number of thesauri a library acquires is limited, and the larger the number of databases to be searched, the higher is the likelihood that the thesauri for some databases would not be available.

The fact that multidatabase searches provide strong incentive for avoiding thesaurus consultation has important implications. While it may seem that searchers are free to choose the number of databases they search, their decision is determined by the distribution of information among databases rather than by their desire to try new databases. That is, the distribution of information among databases within a subject area determined the number of databases that were used per search. Thus, the number of databases that need to be used is a given.

These findings provide evidence for the conclusion that the need to use several databases causes searchers to enter textwords and to avoid consulting a thesaurus. This effect is obviously an *impediment* to searching because it limits the choices in the selection of search keys that searchers can have.

It is naive to think that databases could be created so that a request would always require a search in one or two databases. Databases vary by the literature they cover and by the access they provide, and this variability is often productive. The fact is, however, that searches of the medical literature, for example, required an average of 1.33 databases for a search, while those of the science literature, 2.64. It is plausible to assume, therefore, that the variability among databases that now exists is determined not by a desire to optimize searching, but rather by economic considerations. To improve online searching then it is important to achieve higher standardization and cooperation among database producers and among search systems vendors, in order to make it possible for searchers to make informed decisions about search-key selection when they switch databases.

The main effect of multidatabase searching on search-key selection, however, is introduced by the fact that the thesauri of databases are not coordinated. Clearly, if searchers did not have to consult a different thesaurus each time they switch a database, they would be less likely to enter textwords without consulting a thesaurus.

An approach to minimize the effect of multidatabase searching on the selection of search keys is to introduce

a switching language that facilitates the "translation" of the vocabulary of one thesaurus into another, and the vocabulary of a user into the vocabulary of a designated thesaurus (Soergel, 1974). Such languages are suitable for intermediary expert systems that mediate between a searcher and a set of databases. With a switching language, descriptors and textwords can be selected by a system for each request and for every database that is to be searched without user assistance.

While there are still unsolved problems in the construction of switching languages, the use of such a language has already proven to be useful (Chamis, 1988). Further, a large scale project to construct a switching language for medical terminology, the Unified Medical Language System (UMLS), is under way at the National Library of Medicine (Schuyler, 1989).

The conclusion that the need to search multiple databases has an effect on the selection of search keys only emphasizes the importance of the role of intermediary expert systems in masking the differences between databases. One should remember, however, that most of the existing differences between databases are not necessary; they are introduced often because of commercial considerations that may or may not satisfy searching needs. It is more useful to avoid unnecessary inconsistency in database design, and to mask the necessary variability. To do that would require cooperation among database producers, something not common in the current highly competitive environment.

Thus, research should be carried out to discover which features of databases and their thesauri can be standardized without affecting retrieval quality. The role of intermediary expert systems will then be to bridge the necessary differences, employing switching languages and other terminological and semantic networks.

While the construction of working switching languages and intermediary expert systems still requires much research, other solutions to the problems generated by multidatabase searches are more realistic at present. The most immediate solution is for database producers and search-system vendors to provide, easy, flexible, and inexpensive online access to thesauri.

Acknowledgments

The study team on this project consisted of Nancy Phelps, Michael Crandall, Cynthia Altick Cunningham, and Kathleen McCrory. Their contribution to the study was substantial and highly appreciated. The online searchers who participated in this study were extremely cooperative spirits, their collaboration made the study possible. My special thanks to Dagobert Soergel and Elaine Svenonius who reviewed earlier versions of this report. Their insightful comments and suggestions provided an inestimable contribution.

References

Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieving system. *Communications of the ACM, 28,* 289–299.

Chamis, A. Y. (1988). Selection of online databases using switching vocabularies. *Journal of the American Society for Information Science, 39,* 217–218.

Carrow, D., & Nugent, J. (1981). Comparison of free-text and index search abilities in an operating information system. In *Information management in the 1980s: proceedings of the American Society for Information Science 40th annual meeting, September 26-October 1, 1977* (pp. 131–138). White Plains, NY: Knowledge Industry Publications.

Cleverdon, C. W. (1962). *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems.* Cranfield, England: College of Aeronautics, Aslib Cranfield Research Project.

Cleverdon, C. W. (1984). Optimizing convenient online access to bibliographic databases. *Information Services & Use, 4,* 37–47.

Diesing, P. (1971). *Patterns of discovery in the social sciences.* Chicago, IL: Aldine-Atherton.

Dubois, C. P. R. (1987). Free text vs. controlled vocabulary: a reassessment. *Online Review, 11,* 243–253.

Fidel, R. (1984). The case study method: a case study. *Library and Information Science Research, 6,* 273–283.

Fidel, R. (1991). Searchers' selection of search keys: I. The Selection Routine. *Journal of the American Society for Information Science, 7,* 490–500.

Fugmann, R. (1982). The complementarity of natural language and indexing languages. *International Classification, 9,* 140–144.

Henzler, R. G. (1978). Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information systems. *International Classification, 5,* 21–26.

Katzer, J. (1982). A study of the overlap among document representations. *Information Technology: Research and Development, 1,* 261–274.

Keen, E. M. (1973). The Aberystwyth Index Languages Test. *Journal of Documentation, 29,* 1–35.

Lancaster, F. W. (1980). Trends in subject indexing from 1957 to 2000. In P. J. Taylor (Ed.), *New trends in documentation and information.* London: Aslib.

Markey, K., Atherton, P., & Newton, C. (1980). An analysis of controlled vocabulary and free text search statements in online searches. *Online Review, 4,* 225–236.

Parker, J. E. (1971). Preliminary assessment of the comparative efficiencies of an SDI system using controlled or natural language for retrieval. *Program, 5,* 26–34.

Schuyler, P. (1989, October). *Improving access to biomedical information systems: the Unified Medical Language System and the role of MeSH.* Paper presented at 52nd annual meeting of the American Society for Information Science, Washington, DC.

Soergel, D. (1974). *Indexing languages and thesauri: construction and maintenance.* Los Angeles, CA: Melville.

Svenonius, E. (1986). Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science, 37,* 331–340.