

## WRITING ABSTRACTS FOR FREE-TEXT SEARCHING\*

RAYA FIDEL

*Graduate School of Library and Information Science, University of Washington,  
FM-30, Seattle, Washington, 98195, USA*

A survey of abstracting policies used by producers of bibliographic databases examined abstracting guidelines which aim to enhance free-text retrieval. Of the 123 database policies examined, fifty-seven (46 per cent) included such instructions. Editors consider content of abstracts and their language as a primary factor in retrieval enhancement. Most recommend that once abstractors decide which concepts to include in abstracts and in which form to represent them, these terms should be co-ordinated with index terms assigned from a controlled vocabulary. Guidelines about the type of abstracts, i.e., informative or indicative, and about their length are not affected by the capability of free-text retrieval.

### INTRODUCTION

EVER SINCE ABSTRACTING AND INDEXING SERVICES provided bibliographic databases in a printed form, the functions assigned to indexing have been almost completely different from those assigned to abstracting. Indexing facilitates efficient retrieval of information, and abstracting facilitates judgements about the relevance of the retrieved information. Abstracts have even been viewed as substitutes for original documents.

The availability of bibliographic databases online changed the role of abstracts. The abstracts of most publicly available bibliographic databases can be searched in the free-text mode which allows users to search online for the occurrence of any terms they think appropriate. The system then retrieves references to all documents whose abstracts contain the requested terms or combinations of them. This process takes place without the use of a specific index language.

Plainly, abstracts have become an important *enhancement* for information retrieval.

Despite this new role, vital questions concerning the construction of abstracts to facilitate efficient retrieval are still unanswered. Active research on abstracting ceased in the mid-1960s, after the importance of abstracts for judging document relevance and for use as substitutes was established by a series of tests,<sup>1</sup> and a comprehensive survey of 130 abstracting policies produced criteria to be used in the evaluation of abstracts.<sup>2</sup> While these studies are milestones in abstracting research and their results are still highly valuable, they do not examine the appropriateness of abstracts for free-text retrieval.

Researchers in the design and evaluation of information systems, however, have been most interested in the performance of free-text searching. Indexing is a costly, labour-intensive process, and index languages are expensive to construct. Further, several theoretical issues related to their construction remain unresolved. The capability to retrieve documents without the aid of an artificially constructed

\* This study was supported by the Graduate School of the University of Washington.

index language seems, therefore, extremely attractive. Beginning with the Cranfield studies,<sup>3</sup> numerous experiments have tried to determine if free-text searching indeed outperforms searching with the aid of a controlled index language. Results are contradictory, and the debate continues as to whether or not future systems should dispose of controlled vocabularies and be restricted to the free-text searching capability.<sup>4-7</sup>

Among practitioners, designers of information systems are providing abstracts for free-text searching, and system users heavily employ this capability. Several abstracting and indexing services have investigated the use of free-text searching to determine how to design their index languages and abstracting policies.<sup>8-9</sup> In addition, online searchers attempt to identify efficient strategies for free-text searching of abstracts.<sup>10</sup>

In view of the extensive experience in free-text searching gained in recent years, the time seems ripe for a systematic investigation of characteristics of abstracts that are essential for successful retrieval.

To consider the criteria vital for acceptability of abstracts for free-text searching, this study surveyed abstracting policies used by database producers to determine the degree to which existing abstracting and indexing services already incorporate such criteria into their abstracting. Most importantly, the study synthesises the characteristics of abstracts important to free-text retrieval, as presented by abstracting policies, and touches upon their effects on the quality of abstracts.

#### THE STUDY METHOD

To survey abstracting policies, a list of producers of bibliographic databases who make abstracts available for free-text searching was compiled from various database directories. Information about individual databases, taken primarily from literature provided by search system vendors, eliminated producers who use only author abstracts.

Letters explaining the purpose of the study and asking for copies of abstracting and indexing guidelines were then sent to 159 database producers. Letters to American and Canadian producers were addressed to the editors responsible for abstracting in their organisation. If they were late in their response, they received a reminder and a follow-up telephone call. Letters to producers in other countries were addressed to 'the editor' and were *not* followed by any further enquiry.

Seventy-five of the ninety-six North American database producers responded to the request, resulting in a 78 per cent response rate. Producers in other countries responded 59 per cent of the time, with thirty-seven of sixty-three producers answering the request. Respondents provided answers in various forms: those who have printed guidelines provided a complimentary copy, while others wrote letters or telephoned to describe their abstracting policy and practice.

Based on information provided by database producers, data were analysed on the level of individual databases. Some editors reported that their abstracts are no longer available for searching, and others claimed to use no abstracting policy. These databases were dropped from consideration. A total of 123 databases were identified which provide searchable abstracts that are written according to some sort of guidelines. These guidelines were analysed to extract the instructions given by editors of abstracting policies to enhance online retrieval.

This paper describes and discusses these instructions. While no mention of in-

dividual databases is made in the text, a list of database producers whose abstracting policies and practices directly contributed to the compilation of these instructions is given in the appendix.

Of the total number of databases considered for this study, eighty-nine (72 per cent) provide their abstractors with written guidelines while the other thirty-four instruct abstractors informally. Statistics show that editors who provide written guidelines are more likely to be concerned about instructions for free-text retrieval than their counterparts: while 50 per cent (forty-five) of printed guidelines include free-text instructions, only 35 per cent (twelve) of databases who provide informal guidelines address this issue.

It should be noted that individual database producers often use one set of abstracting guidelines for a number of their databases. Thirty-six producers (listed in the appendix) issued the guidelines for the fifty-seven (46 per cent) databases including free-text instructions; these databases are the basis for the free-text instructions described in this paper.

Examination of the relationship between descriptor indexing and free-text instructions reveals two trends as shown in Figure 1.

First, a large number of databases that use descriptor indexing still avoid instructions for free-text retrieval. Among the 103 databases that use descriptor in-

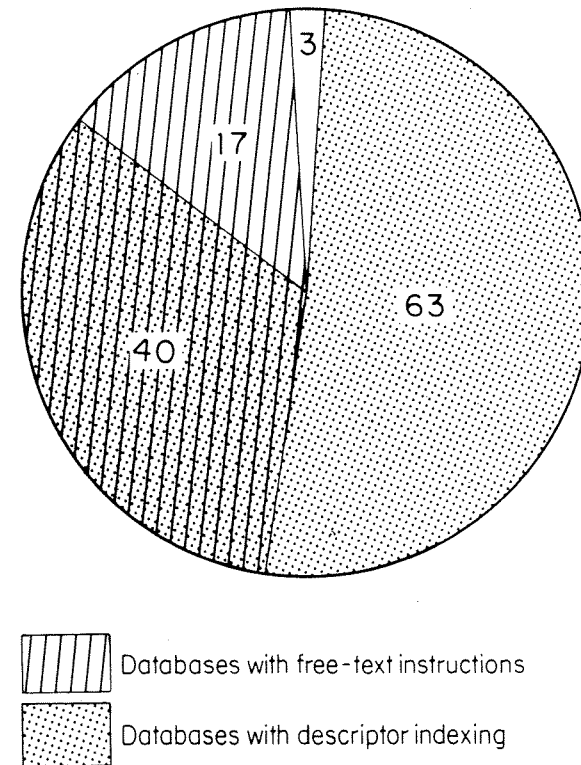


FIGURE 1. Breakdown of the 123 databases included in the study based on the presence of free-text instructions and descriptor indexing

dexing, only forty (39 per cent) provide instructions for free-text retrieval. In fact, sixty-three databases (51 per cent of the 123 databases considered in this study) use descriptor indexing but *do not* provide instructions for free-text retrieval, even though their abstracts can be searched online and they do have some sort of abstracting guidelines. In other words, free-text instructions are not highly common among editors who use descriptors for their indexing.

Some editors may not yet have seen the connection between abstracting guidelines and free-text retrieval; others realise the need to incorporate them in their guidelines but have not done so yet. Still other editors have decided to ignore the effect of abstracts on retrieval based on considerations of cost-effectiveness. Some of these editors believe that since abstracts and titles, which have already been composed, provide additional access points with no additional effort at the input stage, special instructions for writing such abstracts make free-text retrieval less cost-effective. Alternatively, others believe that descriptor searching is the best retrieval method for their databases; thus, investment of resources in enhancing free-text retrieval is not cost effective to them.

The second trend that emerges from the data is that most databases that provide free-text instructions do not consider descriptor indexing obsolete. Among the fifty-seven databases that provide free-text instructions, forty (70 per cent) use descriptor indexing. In other words, a considerable percentage of the editors who judge free-text searching to be an important method of retrieval also believe that it cannot substitute for descriptor searching. Thus, the debate whether or not free-text retrieval should substitute for descriptor indexing is somewhat academic; databases that enhance free-text retrieval are likely to provide both types of access points.

#### ABSTRACTING GUIDELINES

In their study of general abstracting instructions, Borko and Chatman addressed three issues: the function, the content and the form of abstracts.<sup>2</sup> Interestingly, only the latter two are considered by database producers in their specific instructions for computerised databases.

The first issue, the function of abstracts, determines if abstracts should be informative or indicative: whether they should condense the information in documents or report on what documents are about. While editors of computerised databases usually instruct abstractors about which type of abstract to compose, none of them presume that it relates to free-text retrieval. In other words, the nature of abstracts, whether informative or indicative, is still determined only by the purpose and use of abstracts and not by any factors that relate to information retrieval.

When general abstracting policies address the second issue, the content of abstracts, they indicate the *information* that should be included in abstracts, such as study method or research results. Editors who are interested in retrieval with abstracts additionally designate which *concepts* should be included in abstracts. As we shall see later, while editors generally agree about the nature of concepts to include, they present a variety of opinions, some contradictory, about the specific criteria to use for concept selection.

The third issue, the form of abstracts, relates primarily to writing style and length of abstracts. Surprisingly, in all the abstracting manuals reviewed, no mention is made of possible relationships between the length of abstracts and efficiency of retrieval. Writing style, however, is related to retrieval: a number of editors issue policies about the way in which concepts should be represented to enhance

retrieval. Unfortunately, some instructions make it difficult to maintain an acceptable writing style, with brevity and clarity of writing most affected.

A short summary of the guidelines for writing abstracts that were identified in the study to be aimed at free-text searching are given in Figure 2.

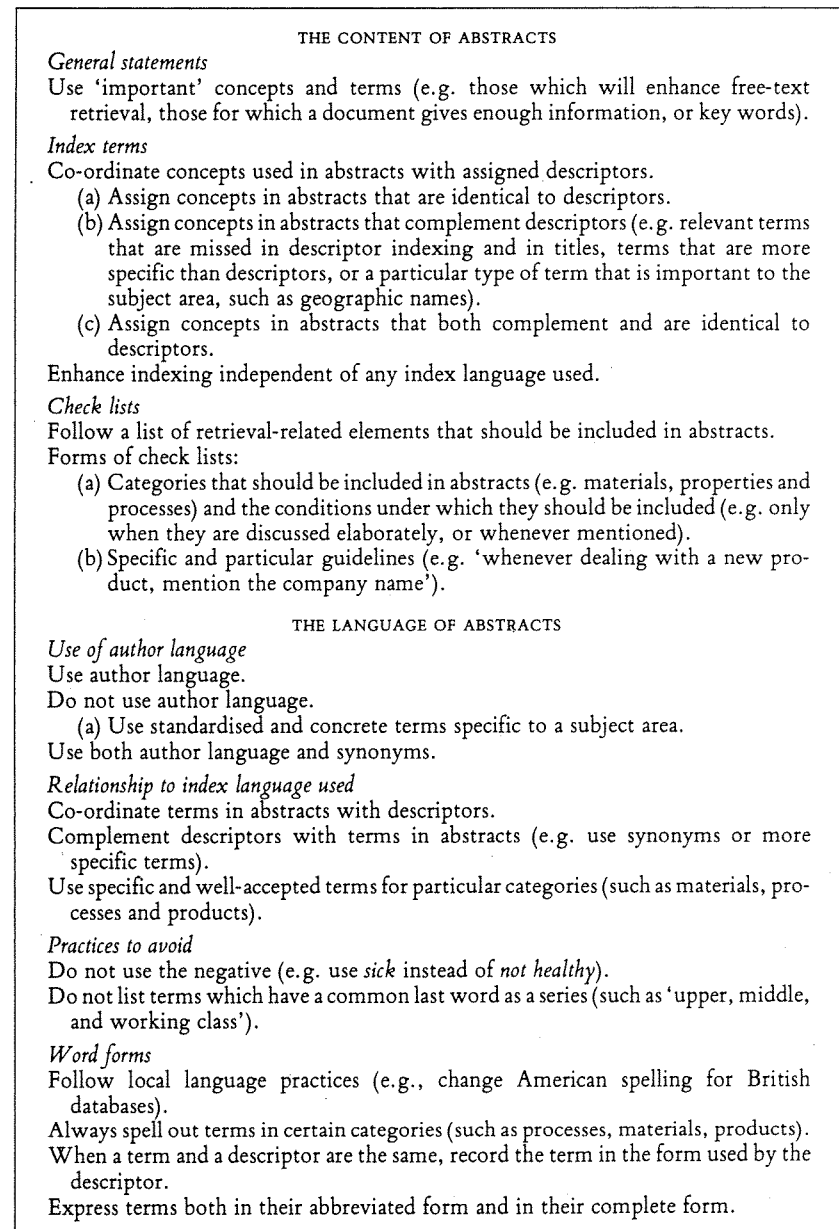


FIGURE 2. A summary presentation of abstracting guidelines for free-text searching

*The content of abstracts*

One maxim is agreed upon by all: concepts 'important' for retrieval should be included in abstracts. Notably, some database producers find it sufficient to include a general statement of that kind in their abstracting policy without providing further details. But when specific criteria for selecting 'important' concepts are devised, editors part company, in particular with regard to the relationship between these concepts and other indexing terms.

*General statements.* Because a statement of this sort reminds abstractors that their abstracts will be used for retrieval, it should include a mention of concepts central to the abstracted documents. The statement issued by the International Organization for Standardization, for instance, asks abstractors to 'use significant words from the text which will help computerised text searching',<sup>11</sup> While this statement can be accepted as an international standard, individual editors are in a position to define more specifically which topics are 'significant for computerised text searching.'

Editors may elect to be more or less rigorous in their definitions, but available guidelines appear inadequate when they do not give some additional guidance to abstractors. Without being too specific, an editor may ask abstractors to include concepts for which a document gives enough information. One may even specify the desired form of information and require the inclusion of only those topics for which, say, test procedures and results are given. Another approach is to suggest that key words be incorporated into abstracts, especially if they are often repeated in the text.

*Index terms.* To require that important concepts be explicitly represented in abstracts for computerised searching is actually a requirement that documents be indexed using natural language. While general statements about the contents of abstracts do not refer to indexing, numerous abstracting manuals clearly point out that abstracts should enhance indexing. In fact, some organisations require that abstractors thoroughly review the indexing of documents before they write or edit abstracts. Editors of abstracting manuals, though, vary in their notion about the way indexing should be enhanced by abstracts.

One approach claims that abstracts should be composed with indexing in mind but independently of any index language used. This approach makes it difficult to give specific instructions, and normally a general statement is given to the effect that abstracts should include terms under which the subject of a document would be searched. Editors may want to remind abstractors that this function is particularly important for documents whose titles are short or uninformative. Keeping in mind the notion that terms in abstracts are access points, abstractors can be provided with examples to sharpen their awareness. They can be guided to exclude general or inclusive concepts, such as 'various locations', and to use specific ones, such as names of locations.

The other approach co-ordinates concepts used in abstracts with index terms assigned from a thesaurus, i.e. with descriptors. Here, we confront the question: should concepts in abstracts be identical to descriptors or should they complement them? Although most services choose concepts in abstracts to complement descriptors, some editors specifically instruct their abstractors to write abstracts with embedded descriptors. Others see no contradiction between the two options. While abstracts are written to complement indexing, they think that

redundancy in index terms is actually desirable. In fact, they maintain that an abstract is deficient if it does not include the terms under which it was retrieved using descriptors.

Editors who decide that concepts in abstracts should *complement* descriptors, may give various instructions to secure indexing enhancement with these concepts. Examples of such instructions include: concepts not in the title or descriptor fields that the author may want to be retrieved; any relevant terms which enhance the thesaurus terms, which are appropriate synonyms, or which are not adequately presented by descriptors; or language equivalents to subject codes for databases where descriptors are in the form of subject codes.

These instructions are still quite general and rely on the subjective judgement of abstractors. Additional guidance can be provided by indicating the nature of the concepts to be selected for indexing enhancement. The prevailing notion is that concepts in abstracts should be more specific, or more detailed, than the descriptors they are intended to enhance. For individual databases, however, editors may have more specific requirements such as the inclusion of geographic names, terms that are unique to the field, catchwords, jargon, and phrases from documents, or proper names appearing with terms.

*Check lists.* The most rigorous approach to indexing with terms from abstracts involves developing a list of retrieval-related elements that should be included in abstracts, whether or not they are covered by descriptor indexing. Such check lists may take a variety of forms.

The most structured check list contains categories that should be included in abstracts. The actual categories usually vary from one database to another. In a technology area, for instance, one database always includes the *type* of welding process and steel *types*, and others suggest the inclusion of materials, forms, conditions, properties and processes. In another area; company names, product names and individual names are considered especially significant. One service lists eleven categories and requires that any relevant categories be included in abstracts.

Another use of check lists is to write down very specific guidelines such as: whenever dealing with a new product, process or technology, mention the company name; include the full standard name for all tests, questionnaires etc.; if the document is about a survey, state who conducted the survey; or, if a company is a subsidiary, mention the name of the parent organisation if it is not included in the name of the company.

In their goal to enhance indexing with concepts from abstracts, editors may overlook other considerations. For example, by insisting that abstracts include a detailed list of categories they may make brevity impossible. Editors should, therefore, inform abstractors under what conditions to include each category. Such conditions can be specified in a variety of ways.

For a selection of categories, e.g. tests or questionnaires, one may require that whenever they are mentioned in a document they should be reported in abstracts, even if they are mentioned in passing or if the author does not make them appear significant. For other categories, editors may require that they be relevant or significant to the document, or that they are discussed elaborately. Using a quantitative approach, one may devise an indirect measurement of significance and stipulate, for example, that company names be specified in abstracts unless more than ten companies are mentioned in a document.

*The language of abstracts*

Once the concepts to be included in abstracts are determined, instructions must be given about the use of language. More specifically, editors must indicate which terms should be selected to represent these concepts and which word forms to use. Here again, editors agree about the general nature of terms that *should not* be used; abstracting policies include statements that recommend avoiding the use of arcane or colourful language or of phrases which would result in 'false drops' when searching. It seems, however, that such loosely defined statements provide very little guidance to abstractors, and editors must be more concrete about term selection and word forms. At this point, one finds contradictory instructions, each presented with plausible, but untested, arguments. Moreover, some guidelines violate a number of rules for clear writing style.

One example of such instructions is the question whether abstracts should use author language. While this is an issue in abstracting policies for printed publications, it seems to be of special significance for free-text retrieval. The rationale for using author language is that users often remember key phrases used by authors and will incorporate them into their search strategies. The counter argument is that the use of direct quotations is undesirable because they usually do not carry enough information. In the middle ground, one editor requires that abstracts include both author language and synonyms. While synonyms are a safeguard against missing a concept in retrieval, one should remember that their use affects the quality of writing style. Readers may find it confusing when a concept, which is usually also a significant one, is represented by a number of terms within one paragraph.

When abstractors are not asked to adhere to author language, the most frequent recommendation is for the use of standardised, concrete terms and a language that is specific to a subject area. No argument against this approach was found in abstracting policies, but editors evidently disagree about which specific instructions achieve uniformity in term assignment across abstracts.

Once more, an important decision is whether or not to co-ordinate terms in abstracts with descriptors. Both approaches exist in abstracting policies, but the manner in which these terms should be co-ordinated is undecided.

Editors who wish to induce consistency in term usage across abstracts demand that thesaurus terms be used whenever applicable. In contrast, those who believe that retrieval can be improved if one provides a variety of access points recommend avoiding the use of the same terms in both indexing and abstracting. Abstractors are thus instructed to complement index terms truly and use synonyms or more specific terms. The use of synonyms, however, may degrade the value of writing style, as explained previously.

As a compromise between these two approaches, editors can list selected categories for which abstractors should use specific, standardised or well-accepted terms. Examples of such lists of categories are: (a) materials, forms, processes, conditions, properties, and products; (b) test names, program names, institutions and intellectual processes; or (c) names of subsidiary companies. Although no authority list especially developed for terms used in abstracts was found in abstracting policies, some database producers choose to develop detailed policies for the use of terms to designate relevant categories. A typical example of such a rule is the requirement that Latin names as well as common or vernacular names of certain entities be explicitly mentioned. Another example is a set of rules to determine how to record geographic locations.

Peculiar to language usage for free-text retrieval are two practices that are considered undesirable but which were overlooked by many editors. The first is: avoid using the negative in order to minimise or eliminate false retrieval. Suppose, for instance, that authors describe a test to evaluate vendors of online search systems. A phrase in an abstract such as: 'evaluates all search systems, except DIALOG, for response time' will retrieve the abstract when documents about DIALOG are desired. Moreover, the abstract will be rejected in a search about response time on BRS or ORBIT. The second practice is: avoid listing terms which have a common last word as a series, with the common term written only once. For instance: write 'primary care, secondary care and tertiary care' rather than 'primary, secondary and tertiary care.' Some caution is, however, desired: while the first requirement may affect the accuracy of abstracts, the second should be weighed against brevity and clarity.

Lastly, editors can make a choice of word forms to be used. While the importance of consistency in word forms is generally agreed upon, the ways to guarantee such consistency are diverse. A common requirement is to follow local language practices, such as to alter British spelling to American spelling for databases that are primarily used by Americans, or vice versa. Another common concern is the use of abbreviations and acronyms. Editors can require that certain categories, such as geographical names, metals, elements, processes, are always spelled out. Or, they may require that initials of personal names are recorded differently from acronyms for organisations. Descriptors can be used to induce consistency in word forms when terms which are also used as descriptors appear in the abstract in the same form in which they are listed in the controlled vocabulary.

Consistency in word forms, however, is not always considered desirable by every editor. Some maintain that it is preferable to express a name both in a complete and abbreviated form, and even that nicknames be included. A document is then likely to be retrieved no matter what form is used by a searcher.

## SUMMARY

Two issues are addressed by editors of abstracting policies with regard to free-text searching: first, which concepts to include in abstracts, and second, which language to use to represent these concepts.

Concepts to be included in abstracts for retrieval enhancement may be selected in co-ordination with assigned descriptors or may be added independently of any other indexing. When editors elect to co-ordinate concepts in abstracts with descriptor assignment, they can give abstractors clear instructions about whether these concepts should complement descriptors or duplicate them. If both trends are desired, specific guidance should be given as to when each approach should be applied. Regardless of whether the selection of concepts for abstracts is co-ordinated with indexing, editors may find it useful to supply abstractors with a check list of categories or specific elements to be included in abstracts. It would also be advantageous to designate the conditions under which each category should be included.

The language of abstracts may or may not closely follow the language of authors. If it is not tied to the original document, editors must decide whether or not to co-ordinate terms in abstracts with index language. When co-ordination is preferred, editors should decide whether terms selected for abstracts should duplicate or complement index terms. For some databases, editors may select

categories for which a particular form is desirable. While certain language usages can be ruled out, only future research will tell whether consistency or variety in word forms should be required.

While research in abstracting has not addressed issues relating to the acceptability of abstracts for free-text retrieval, editors of abstracting policies issue special instructions which are intuitively based on experience with online searching gained by information specialists during recent years. The disagreement among editors about a number of issues presents the need for research that evaluates the effect of abstracting guidelines on retrieval performance.

## ACKNOWLEDGEMENTS

I wish to thank Kathy Waldera for her assistance with data collection and contacts with database producers, and Oren Sreebny for analysing abstracting guidelines and handling the statistics. I would also like to thank Editor R. T. Kimber for his insightful comments on an earlier version of this paper.

## REFERENCES

1. RESNICK, A. Relative effectiveness of document titles and abstracts for determining relevance of documents. *Science*, 134(3484), 1961, 1004-1005.
2. BORKO, H. and CHATMAN, S. Criteria for acceptable abstracts: a survey of abstracters' instructions. *American Documentation*, 14(2), 1963, 149-160.
3. CLEVERDON, C. W. *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems*. Cranfield: College of Aeronautics, 1962.
4. CLEVERDON, C. W. Optimising convenient online access to bibliographic databases. *Information Services & Use*, 4(1/2), 1984, 37-47.
5. LANCASTER, F. W. Trends in subject indexing from 1957 to 2000. In: *New trends in documentation and information*. London: Aslib, 1980.
6. HENZLER, R. G. Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information systems. *International Classification*, 5(1), 1978, 21-26.
7. DUBOIS, C. P. R. The use of thesauri in online retrieval. *Journal of Information Science*, 8(2), 1984, 63-66.
8. CARROW, D. and NUGENT, J. Comparison of free-text and index search abilities in an operating information system. *Proceedings of the American Society for Information Science Annual Meeting*, 14(2), 1977, 2-e8.
9. MARKEY, K. and others. An analysis of controlled vocabulary and free text search statements in on-line searches. *Online Review*, 4(3), 1980, 225-236.
10. WAGERS, R. Effective searching in database abstracts. *Online*, 7(5), 1983, 60-77.
11. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Information transfer*. Geneva: ISO, 1982, p. 23.

## APPENDIX

Below is a list of database producers who are responsible for the fifty-seven databases that provide free-text instructions in their abstracting guidelines.

ABC-Clio Information Services, Santa Barbara, CA, USA.  
 American Mathematical Society, Providence, RI, USA.  
 American Psychological Association, Arlington, VA, USA.  
 American Society for Metals, Metals Park, OH, USA.

American Theological Library Association Religion Indexes, Chicago, IL, USA.  
 Arctic Institute of North America, Calgary, Alberta, Canada.  
 BioSciences Information Service, Philadelphia, PA, USA.  
 Boreal Institute for Northern Studies, Edmonton, Alberta, Canada.  
 Cambridge Scientific Abstracts, Bethesda, MD, USA.  
 Data Courier Inc., Louisville, KY, USA.  
 Dow Jones & Company, Inc., Princeton, NJ, USA.  
 EIC/Intelligence Inc., New York, NY, USA.  
 Federal Institute for Geosciences and Natural Resources, Hannover, Germany.  
 Finsbury Data Services Ltd, London, England.  
 FIZ Technik, Frankfurt-am-Main, Germany.  
 Greater London Council Research Library, London, England.  
 Health and Safety Executive, Library and Information Services, Sheffield, England.  
 Institute of Paper Chemistry, Appleton, WI, USA.  
 Johns Hopkins University, Population Information Program, Baltimore, MD, USA.  
 Management Contents, Northbrook, IL, USA.  
 National Clearinghouse for Bilingual Education, Rosslyn, VA, USA.  
 National Council on Family Relations, Family Resources & Referral Center, Minneapolis, MN, USA.  
 National Information Center for Educational Media, Albuquerque, NM, USA.  
 National Institute of Education, ERIC Processing and Reference Facility, Bethesda, MD, USA.  
 National Rehabilitation Information Center, Washington, DC, USA.  
 Ontario Ministry of Education, Toronto, Ontario, Canada.  
 The Research Association for the Paper and Board, Printing and Packaging Industries, Leatherhead, Surrey, England.  
 Royal Society of Chemistry, Nottingham, England.  
 Transportation Research Board, Washington, DC, USA.  
 Trier University, Zentralstelle für psychologische Information und Dokumentation, Trier, Germany.  
 United Nations Department of International Economic and Social Affairs, Information Systems Unit, New York, NY, USA.  
 University Microfilms International, Inc., Ann Arbor, MI, USA.  
 University of Sherbrooke, Programme de Recherche et de Développement sur l'Amiante, Sherbrooke, Quebec, Canada.  
 Volkswagenwerk AG, Wolfsburg, Germany.  
 Water Research Centre, Marlow, Buckinghamshire, England.  
 The Welding Institute, Cambridge, England.