

2. DOCUMENT SEARCH

Introduction. How closely a collection of documents on a topic approaches a complete collection on that topic depends on how extensively and thoroughly the literature is searched. One can make an initial estimate of the comprehensiveness of a collection from how comprehensively the search which formed it was made. In this Section, for this purpose, a description of the search strategy used to form the present collection will be outlined. The Section concludes with an assessment of the completeness of the collection and of its possible deficiencies.

Search Systems; Narrow and Broad Searches. If the full text of all documents were in machine-searchable electronic form, the selection criterion could be applied directly to all documents, and selection and search would be identical. Since this is not the case, a separate search is required to bring forward, for manual application of the criterion, the most likely documents from the many tens of millions in existence. For such a search to be feasible, documents, or document surrogates, must be ordered by their attributes, for example by the 'values' of recognised fields. The search can then retrieve documents with values in specific ranges of specific attributes, selected as most likely to contain relatively high numbers of the required documents. Thus, only those documents that have been so classified can be searched; documents on the topic which are not accessioned by such a *system* can not be found, except by chance. A wide literature search requires the use of a wide variety of systems, and the description of the search which follows will detail the types of systems used. A further point made in the analysis in Chapter 2 needs restating here: the likely distribution of literature on a topic suggests that it is best to divide the search effort into an intensive search of the literature of Library and Information Science, a *narrow* search, and an extensive and more superficial search of all other literature, a *broad* search.

Time Limits. An inspection of the seed literature showed that the first paper on the topic is commonly taken to be that of S.C. Bradford (1934a). Accordingly, literature searching prior to this year was highly selective. As noted in Section 4, §1, p.149, a slightly earlier paper on the topic was discovered, S.C. Bradford (1933c), which might rightly be given priority. An inspection of the three selected A&I Services also shows that documents on the topic continue to be published into the 1990's. Initially, it was intended to close the collection at the end of 1984, that is to restrict the study to a fifty-year span. However, to allow time for documents published in the 1980's to be assimilated into the literature and cited by later documents, and for the latter to be

accessioned by A&I Services, the time span of collected documents was extended to include those published till the end of 1989. Less intensive searching has continued through to the present, but the collection referred to in this study was closed in mid-1993, and restricted to documents published till the end of 1989. The publication time-span of documents in the collection described is thus from 1933 to 1989, inclusive. The collection is judged to be 'reasonably complete' from c.1965 to c.1985, but is likely less so, both prior to this time, and, more importantly, in its last four or five years.

A. SEARCHING ON ABSTRACTING AND INDEXING (A&I) SERVICES

§1. TRADITIONAL A&I SERVICES

Use of Traditional Services. To locate an accessioned document, what might be called 'traditional' A&I Services are entered through document fields other than those associated with references. In the present case, fields most indicative of subject content are the title and the abstract, and the extra-documental descriptor field/s provided by the A&I Service indexers; the former are suitable for free text searching, and the latter are usually searched with controlled vocabulary searching. (Other fields may also suggest specific subject content, for example, the author field, the serial field, and even the publication year; here, we may consider these to be additional alerting devices). Searching on these fields for specific words or phrases is effectively a one-off operation, producing a single set of document surrogates, though continual updating is required. The more promising of these surrogates must then be located in libraries, either directly or through interlibrary loan services, and inspected. Documents judged to be 'most likely on' the present topic, and those judged to be 'possibly on' the topic, are stored, and details of them recorded in special files -- as described in the introductory section to this Chapter, under the headings *Outline of Procedure*, p.73, and *File of Documents*, p.74.

Initial seed. The formation of a small initial collection of documents on the topic from traditional A&I Services has been described in Section 1. This seed literature was obtained by first searching for the word-stem '-BRADFORD-', in the indexer-assigned fields only, of three A&I Services specialising in Library and Information Science, to the end of accession year 1986; and then carefully evaluating the records retrieved. The A&I Services were *Library Literature* (1921-), *LISA* (1969-), and *ISA* (1966-). In, total, 137 distinct analytical-level documents, and several monographic level documents, all judged to be on the topic by the A&I Services, were assembled.

Searching Traditional Services Online. Searches of online A&I Services were made using DIALOG Information Services (DIALOG, 1994). Initially, searches were restricted to databases believed to have some subject content in common with Library and Information Science, and a variety of terms were searched for in the title, the abstract, and the descriptor fields or its equivalents.¹⁸ From 1988 onwards, when DIALOG's (1991) OneSearch on all subjects became available, broad searching on all the c. 350 databases of the System was feasible. Two approaches were then adopted. Narrow searching was continued with a more charitable strategy, implemented on DIALOG's (1994) INFOSCI or Library and Information Science DIALINDEX/OneSearch category, using the single term 'BIBLIOMETRIC?'. Broad searching could be implemented on the Basic Indexes of all DIALOG's databases. This generally includes the following fields: the Title, the Abstract, Descriptors (or thesaurus terms), and Identifiers; some databases will have additional fields such as Section Headings, while others may lack some such as Identifiers. A typical search statement is:

```
'SELECT (BRADFORD? (5N) (LAW? ? OR SCATTER? OR DISTRIBUT? OR DISPERS? OR ZIPF? OR LOTKA?))'
```

The co-occurrence of the various truncated terms after the proximity operator, N (or NEAR), ensures that records with the first truncated term, 'BRADFORD', and at least one of the other terms appearing on either side adjacently or up to five words away, in the appropriate fields, will be retrieved. The DIALOG (1991) SearchSave feature was used on databases having one or more documents in both the narrow and broad search statements. Documents were obtained for the records retrieved, and these were examined and appropriately assigned and processed. The above strategy was conducted at the start of the project in the mid 1980's and periodical updates were made until the end of 1992.¹⁹

Searching Traditional Printed Services. Manual searching was undertaken of printed forms of A&I Services which specialised in Library and Information Science, when

¹⁸A&I Services searched were: *LISA*, *ISA*, *PASCAL*, *ERIC* and *INSPEC*; in some cases, *EiCompendex*, *MathSci*, and the 'source indexes' of ISI's *SciSearch* and *Social SciSearch* were added (see DIALOG, 1994). Terms searched include those later to emerge in the selection criterion, viz. '-BRADFORD-', '-LAW-', '-SCATTER-', '-DISTRIBUTION-', '-DISPERSION-', '-ZIPF-', and '-LOTKA-', as well as a variety of terms suggested by reading the text of documents, e.g. '-RANK-', '-CORE-', '-NUCLE-', '-ZONE-', '-BIBLIOGRAPH-', '-PRODUCTIV-', etc.

¹⁹Prior to closing the file in early 1993, the same seven-term search was re-executed on all databases in DIALOG. A few pre-1990 documents which were still not in the BRAD File were uncovered. It was not feasible to carry out a full analysis of these documents (e.g. obtaining them and tracing their references) at that late stage; they have been included in the BRAD File, but with a special code in Field (10) SUBFILE. (See Appendix). These documents are mostly non-English and were published in the late 1980's. Some entered the DIALOG databases much later than their English counterparts, hence they were missed in earlier search updates. Again, some omissions might be due to DIALOG's constant updating, maintenance and addition of a number of databases during the period of the study. For example, the *PASCAL* database (equivalent to the printed *Bulletin Signalétique*), which has a section on 'Information Science Documentation' came online in January 1989 and included documents accessioned from 1986 to 1988 only; later in the 1990's the files dating back to 1973 were added (DIALOG, 1989).

these Services existed before the inception of their online services on DIALOG, or where the Services were not available in machine-readable form through DIALOG. They were searched using the same or similar search strategies to that described earlier for narrow searching. The seven printed Services listed below were inspected for the years indicated. In the first four cases coverage is complete; for the Services marked with an asterisk, coverage was continued online.²⁰

Library Science Abstracts. (1950-1968) *
Documentation Abstracts. (1966-1968) *
Library Literature. (1921-1993)
Bulletin Signalétique: Section 101. Information Scientifique et Technique. (1970-1973) *
Referativnyi Zhurnal: Informatika. English Edition: *Abstracts Journal. Scientific and Technical Information.* (1968-1976), *Informatics Abstracts.* (1977-1992)
Hungarian Library and Information Science Abstracts. (1976-1985)
Indian Library Science Abstracts. (1974-1986).

§2. CITATION INDEXES.

Using Citation Indexes. A Citation Index lists for a document those later documents which are accessioned by the A&I Service, and which cite it in their bibliographic references; a document citing another document generally draws 'information' from it, and has a high likelihood of being on the same topic. Thus, any document assigned as 'most likely on' or 'possibly on' the present topic was searched as a cited reference in the three ISI citation databases: *Science Citation Index*, *Social Sciences Citation Index*, and *Arts & Humanities Citation Index*. Each citing document listed was then checked

²⁰(1) *Library Science Abstracts* (Vol.1-19, 1950-1968) was superseded by *Library and Information Science Abstracts* (Vol.20-, 1969-); this became available online in DIALOG in 1976, and has 1969 as its first year of accession. *Documentation Abstracts* (Vol.1-3, 1966-1968) was superseded by *Information Science Abstracts* (Vol.4-, 1969-); this became available online in DIALOG in 1983 and has 1966 as its first year of accession (DIALOG, 1983). *Library Literature* (1921-) is not available online in DIALOG. *Bulletin Signalétique, Section 101: Information Scientifique et Technique* (or variant title) exists in printed form 1970-1984 (Vol.31- 45); it became available online in DIALOG in 1989, and now has 1973 as its first year of accession (but see previous footnote). *Informatics Abstracts* (or variant title) exists from 1963; excluding some isolated 1966 issues, the first issue scanned was for 1968 (Vol.6); it is not available online in DIALOG. (The five previous A&I Services were available at UNSW library, with some assistance from the National Library of Australia (NLA) and the State Library of New South Wales). *Hungarian Library and Information Science Abstracts* (Vol.1-, 1972-) was scanned in the period 1976-1985 (Vol.5-14) at the CSIRO Library, Melbourne. *Indian Library Science Abstracts* (Vol.1-, 1967-) were scanned in the period 1974-1986 (Vol.8-20) at the library of the University of California at Los Angeles. Neither of the last two services are available online in DIALOG. (2) This list contains -- to quote Todorov (1982) -- the 'big four' traditional A&I Services for Information Science (i.e. *LISA*, *ISA*, *Informatics Abstracts*, and *Bulletin Signalétique 101*). Nevertheless, services such as *Informationsdienst Bibliothekswesen* (1971-) could not be consulted. Furthermore, the coverage of Library and Information Science serials in many former Third World Countries by these A&I Services is deficient vis-à-vis that of the developed nations; e.g. see Buckley (1982) for Third World coverage, as contrasted with Hepfer (1979) for North American coverage.

against records in the BRAD and Non-BRAD Files; if there was no match, a new record was created for the citing document in a holding file. All documents listed in the ISI Citation Indexes as citing a probable or possible topic document were recorded in the CICTS Field (45) of that document, and correspondingly had this cited file document recorded in their own PAREFS Field (35) (see Appendix, Section 1). An attempt was made to inspect all these ISI citing documents, either in libraries, or, when this was not possible -- and the cited text and citing document fields suggested that the document might well be on the present topic -- then through interlibrary loan services or by personal reprint requests. If a document proved not to be on the topic, its record was transferred to the Non-BRAD File. If it proved to be most likely on the topic, its record was transferred to the BRAD File, and subsequently the cited reference search was repeated for this new document. Citation searching is thus an iterative or cycling process, producing a growing set of document surrogates.

Searching Citation Indexes, Online and Printed. The Cited Reference (CR) field of all three of the ISI citation indexes available online through DIALOG, viz. *SCI*, *SSCI*, and *A&HCI*, was searched for all those documents judged to be either on or quite possibly on the present topic. These online services accession documents onwards from 1974, 1972, and 1980, respectively. For documents published prior to these years, the Citation Indexes of the appropriate printed services were consulted. These extend coverage back to 1945 for the *Science Citation Index*, to 1956 for the *Social Sciences Citation Index*, and to 1975 for the *Arts & Humanities Citation Index*.²¹ Citation searching commenced with the seed literature, and was periodically repeated both as new topic documents were discovered, and for an updating of established documents, through to mid-1993.²²

²¹Volumes consulted are: (i) For *Science Citation Index: 1945 -1954 Ten Year Cumulation* (1988); *1955-1964 Ten Year Cumulation* (1984); *1965-1969 Five Year Cumulation* (1971); *1970-1974 Five Year Cumulation* (1976); as well as (initially) annual volumes for 1961-1966, which were published between 1963 and 1973. (ii) For *Social Sciences Citation Index: 1956-1965 Ten Year Cumulation* (1989); *1966-1970 Five Year Cumulation* (1979); *1971-1975 Five Cumulation* (1979). (iii) for *Arts & Humanities Citation Index: 1975-1979 Five Year Cumulation* (1987); and (initially) annual volumes for 1977-1980, which were published between 1978 and 1981.

²²(1) In fact, citation index searching began with a smaller literature still, of nine early and influential documents suggested by reading the seed literature. Five documents were contained in the seed itself, viz. S.C. Bradford (1934a, reprinted in 1977 and 1985), Brookes (1968,1969b), Fairthorne (1969), and Leimkuhler (1967). To these were added four documents published prior to the recognition of the field and the inception of most of the suitable A&I Services. These are S.C. Bradford's book *Documentation* (1948, 1950, 1953) which was published in three English versions and editions -- only Chapter 9: *The Documentary Chaos*, is on the topic, and this is inter alia effectively a reprint of S.C. Bradford (1934a); and B.C. Vickery (1948). These nine documents largely precede the literature explosion of the field in the late 1960's, and from their subsequent citing -- from 150 to 200 other documents each in the BRAD File, and over 100 each in ISI's three Citation Indexes -- are influential in its formation. (2) Documents checked in the Citation Indexes include all analytical-level documents in the BRAD File; these are either in the main collection defined by the graphical selection criterion, or in the four marginal collections admitted, i.e. those coded 'SEM', 'PRE', 'PAR', and 'DIF' (See Sections 4 and 7 of this Chapter). In addition,

B. OTHER SYSTEMATIC SEARCHING

§3. SEARCHING LIBRARY AND INFORMATION SCIENCE SERIALS IN LIBRARIES

The arrangement by subject classifications of monographs and serial publications in libraries, and the restricted subject scope of most serial publications, provides another and even more traditional avenue for the systematic narrow searching for documents. Thus, journals in the field of Library and Information Science thought at all likely to contain papers on the topic were scanned retrospectively, and new issues continually inspected. Even though most of these serials were accessioned by at least one on the searched A&I Services, the low inclusion level required of a document for it to be selected as on the topic made this additional searching both necessary and profitable.

The systematic searching of serials began with the more productive of the journals to emerge as the collection developed, but this was gradually widened and a special effort was made to overcome any bias to journals primarily in the English language. To these were added seemingly-suitable serials in Library and Information Science for many countries, from various compilations, the most important being *Ulrich's International Periodicals Directory 1987-88* (1987) and the companion *Irregular Serials & Annuals 1987-88* (1987).²³ Other possibly productive serials in related disciplines were

some fifty other marginal documents, which were not admitted to the collection, and a number of monographic level documents in the BRAD File (e.g. S.C. Bradford's *Documentation*, 1948, etc.) were also checked, but citation data were not recorded in the files for these documents.

²³(1) Earlier and later additions of Ulrichs' directory were also checked. The lists of serials accessioned by the A&I Services *Library Literature*, *LISA*, *ISA*, and *Informatics Abstracts* were also useful guides. A particularly valuable aid here was Janzing & Brown (1975), which has the added value for a printed source of organising serials by country rather than alphabetically. Fuller accounts of mainly English serials are given in (e.g.) Herson, et al. (1973), and Stevens & Stevens (1982). (2) It is sobering to note that over 900 regular serials are listed for Library and Information Science in *Ulrich's 1987-88 Directory* (1987), and nearly 600 irregular serials and annuals are listed in the companion directory; if serial publications in the discipline subclassified as 'abstracting, bibliographies and statistics' and 'computer applications' are added in both cases, then a total of over 1600 serial publications are listed for the discipline. Janzing & Brown (1975) list some 950 serials in print for the discipline, from 79 countries; they comment that the number of such serials has nearly doubled since 1968. Of course, it is reasonable to presume from their communication type and specific content scope that the large majority of these serials would never receive papers on Bradford's Law of Literature Scattering. Many appear to deal with local library matters. Only about two fifths of the regular subannual serials are indexed by even one recognised A&I Service, and only one fifth to one quarter are indexed by any of the five major Library and Information Services: *LISA*, *ISA*, *Informatics Abstracts*, *Bulletin Signalétique 101*, and *Library Literature*. But against this, papers on the topic are also published in serials outside of the discipline. (3) Further to the previous comment, even these various lists proved to have a serious omission -- suitable publications from the People's Republic of China. A revivification of research in (inter alia) Library and Information Science occurred in China in the late 1970's, and numerous serials, some with papers on the present topic, began publication. This literature was only discovered fortuitously by the present author in browsing at NLA; the Orientalia Room of this Library has collected a number of more important serials

occasionally added as well.²⁴ The locations of these serials were ascertained and steps taken to inspect their content. Those held in libraries in mainland south-eastern Australia were examined directly, in some cases on a regular basis: most important of these libraries were the University of New South Wales (UNSW) Library, with a reasonably good collection of English-language Library and Information Science serials, and the National Library of Australia (NLA) in Canberra, with the more important serials in the discipline from many countries.²⁵ Gaps in all collections meant that a careful accounting was needed, and full coverage sometimes required visits to several libraries. For more promising serials held exclusively elsewhere in Australia, or outside of Australia, the tables of contents for appropriate years were requested through the interlibrary loan service of the University of New South Wales. From these, a generous selection of articles judged from the titles as likely to contain information on the topic were then requested through the same service. Finally, through incidental trips to the United States and New Zealand, serial titles not otherwise seen but held in accessible libraries could also be scanned. The most important of these libraries was the Research Library of the University of California at Los Angeles (UCLA).²⁶

A sample of the serials 'inspection list', for titles commencing with A and B, is provided in Table 3-2; this is reasonably representative of all serials inspected.²⁷ As will be

from China (and Japan, Taiwan, etc.) Because of translation and transliteration difficulties, this literature required extra effort to interpret; see e.g. Su (1984), Tao, et al. (1989); and Zuo (1988).

²⁴Representative of these serials, from the more to the less obvious, are: *Journal of Chemical Information and Computer Sciences*, *Methods of Information in Medicine*, *IEEE Transactions on Professional Communication*, *Scientometrics*, *Social Studies of Science*, *Science*, *Voprosy Filosofii*, *Priroda*, and *Vestnik AN SSSR.*, (or variant titles).

²⁵Apart from the UNSW Library and the NLA, the following Australian libraries were visited, sometimes repeatedly, and all visits were productive: Fisher Library at the University of Sydney, the State Library of NSW, Michael Birt Library at University of Wollongong, Chifley and Hancock Libraries at Australian National University, the former Canberra CAE library (Belconnen, A.C.T.), the former Queensland Institute of Technology library (Brisbane), Monash University Main Library (Clayton, Victoria), and the former CSIRO Information Services Unit Library (East Melbourne). Regrettably, the library at the University of Melbourne was not visited, but the table of contents of several serials were provided as interlibrary loans.

²⁶Apart from the UCLA Research Library, the following libraries outside of Australia were visited, again all visits being productive: Hamilton Library at the University of Hawaii (Honolulu), Pattee Library at Pennsylvania State University (State College, Penn.), and the University of Auckland Library (New Zealand).

²⁷The serials are listed in the Table approximately as they are titled in *Ulrich's Directory* (1987) and its *Irregulars* (1987) companion for 1987-88. We may compare the regular subannual serials listed here with those listed in *Ulrich's Directory* (1987), after some adjustments (e.g. conflating *American Documentation* and *Journal of the American Society for Information Science*, moving *Bulletin of the American Medical Library Association* to 'M', and admitting *American Archivist* from the category 'History'). *Ulrich's* lists c. 200 serials, of which c. 80 are indexed and c.50 are indexed by one or more of the top five Library and Information Science A&I Services. I list 24 at least partially scanned serials, of which 22 were indexed. Thus, for Library and Information Science serials with titles beginning with the letters 'A' and 'B', I scanned about one eighth of those existing at the time, a little under one third of those indexed, and somewhat under one half of those indexed by the top five Services; these figures are representative of the full set.

apparent, this searching technique was very labour intensive, but it uncovered many documents, especially non-English documents, which would otherwise have not been located. Despite these efforts, it must be noted that coverage is not as thorough for the period 1985-1989 as for 1965-1985. Many of the journal title runs scanned outside the UNSW Library did not go beyond 1985, and some of the more 'esoteric' journal titles were discontinued in the 1980's when library budgets started decreasing; one library was even disbanded in 1989.

§4. TRACING BIBLIOGRAPHIC REFERENCES

Analysis of References. The bibliographic references of all those documents judged to be on the present topic were examined. Each document surrogate was checked against records in the BRAD and Non-BRAD Files, and if there was no match, a new record was created for it in a holding file. All documents listed in the references of a probable or possible topic document were recorded in the REFS Field (30) of that document, and correspondingly had that record's number listed in their own BCITS Field (40) (see Appendix, Section 1). An attempt was made to inspect all referenced documents, particularly those at the analytical level and more so those which were articles in journals; even when there was no likelihood of these documents being on the topic, a verification of the accuracy of the references was considered necessary. Documents were located in libraries, or, when this was not possible -- and the citing text and cited fields suggested that the document might well be on the present topic -- then through interlibrary loan services or by personal reprint requests. If a document proved not to be on the topic, its record was transferred to the Non-BRAD File. If it proved to be most likely on the topic, its record was transferred to the BRAD File, and subsequently its bibliographic references were examined and searched for. Reference tracing is thus an iterative or cycling process, producing a growing set of document surrogates.

§5. SEMI-SYSTEMATIC SEARCHING

Several other forms of searching were employed. First, an attempt was made to locate and inspect all documents written by the more prolific authors on the topic; this was largely completed for S.C. Bradford (see Section 4, §1, p.142) and B.C. Brookes, but in general proved to be unproductive for the effort involved. Second, as noted, the arrangement of monographs by subject classifications in libraries allowed the author to browse through Library and Information Science and related publications in the course of visits to libraries to inspect serials. Also, the checking of references (see §4 above) which were more widely dispersed in libraries occasionally led the author to chance

upon documents outside the discipline of interest here. Finding works with discussion of the present topic, and/or citing other works possibly also on the topic, was quite fortuitous. Third, in the course of assembling the collection, documents which were (principally) bibliographies on subjects related to the present topic were located by various means. The document surrogates listed were examined and those judged likely to have material on the present topic were searched for as described earlier. A number of more obscure non-English documents on the topic were obtained in this way through interlibrary loan services.²⁸

C. SUMMARY OF DOCUMENT SEARCH

Search Procedure. The different techniques listed above for searching for documents on the present topic were carried out in parallel. New documents, found through searching traditional A&I Services or through the searching of consecutive issues of serials, and which proved to be possibly on the topic, were then checked for later citing documents in the Citation Indexes, and had their earlier referenced documents traced. This checking of the so-called 'referentions' of documents possibly on the topic lead to further documents being found which were possibly on the topic, and to further referentions-checking. Initially, new words, which were possibly indicative of the topic, were found in documents, leading to additional traditional searches; and even later, new serials emerged which warranted systematic searching. And except for reference tracing, the techniques required periodic updating. Yet even with this rather intricate protocol, new documents on the topic were periodically uncovered by chance; it is doubtful if some of these could ever have been uncovered by realistic systematic searching.

²⁸(1) Bradford wrote at least 60 papers, letters or biographical notes related to documentation, of which 57 have been examined; 11 are judged to be on the topic (see Section 4, §1). Prior to the publication of Shaw (1990), which lists 123 documents written by Brookes in the discipline of Library and Information Science, I had examined 99 articles, letters, etc. by this author; in all, 55 suitable works by Brookes are judged to be on the topic. (2) Most useful bibliographies are the following. (i) Closely related to the topic are: Simon (1986), c.200 items on Bradford's Law; Subramanyam (1979a), c. 200 items on scattering laws in Library and Information Science; and Zunde (1987), c.220 items on all laws and regularities in Information Science. (ii) On Bibliometrics are: Pritchard (1969b), c.700 items; Pritchard & Wittig (1981), c 600 items; Hjerpe (1980), c. 2000 items; Hjerpe (1982), c.500 items; and Hertz (1987), c.200 items. (iii) On Scientometrics are: a series in *Scientometrics* compiled by Schubert (1983-1991, bibliographies no.1-17, and continuing); and lengthy bibliographies on various aspects in articles by Vlasy commencing in 1970 principally in the *Czechoslovak Journal of Physics Series B*. (iv) Country-specific bibliographies deserving mention here are: for the former U.S.S.R., Doroshenko & Haitun (1989),c. 250 items, and Afremov, et al. (1984), c. 580 items; for China, Yang & Lu (1984), c. 150 items; for India, Sen & Kumar (1986), c. 200 items; and for the former Yugoslavia, Filipi-Matutinovic (1986), c. 80 items. Simon (1986), c. 200 items, and Gerdel, et al. (1976), c.300 items, were helpful for the German literature, and Quemel, et al (1980), c. 150 items, for the Brazilian literature.

Search Results. With the stabilisation of the selection criterion, a number of documents 'possibly on' the topic were finally rejected (see Section 4 of this Chapter, p.141), and records of their citing and referenced documents were removed. A total of 1691 documents published prior to or in 1989 were judged to be on the topic, and were recorded in the BRAD File; of these, 1362 were analytical-level scholarly documents of most interest to this study, and of these 1338 were inspected before the closure of the collection.²⁹ For these 1338 documents, their citing documents (as provided by the Citation Index) and their bibliographic references, if not also on the topic, were recorded in the Non-BRAD File. A total of 2607 Citation Index-provided citing documents were recorded, of which 569 (21.8%) proved to be on the topic, and 2604 (99.9%) were analytical-level documents, predominantly in regular serial publications. Fully 2536 (97.3%) of these 2607 documents were located and inspected; the remaining 71 documents did not seem likely to be on the present topic, as judged from the titles, serial titles, and abbreviated references provided in the source indexes of the Citation Index. A total of 12,216 bibliographic references were recorded, of which 995 (8.1%) proved to be on the topic, and 8285 (67.8%) were analytical-level documents. Some 6240 (75.3%) of the 8285 analytical-level documents were located and inspected, as were 625 of the 3931 non-analytical-level documents; of the remaining 2045 analytical-level documents which could not be directly verified, none seemed likely from the citing text or document titles to be on the topic.

D. EVALUATION OF SEARCH: COMPLETENESS AND BIAS

Completeness and Bias of a Search. The level of *completeness* of a search may be defined as that percentage of *all* documents meeting some specification which is actually obtained by the search. The measure may be applied globally, or locally with respect to certain values of some document attribute(s), for example to only those documents on a topic which are in English. The *bias* of a search measures the deviation in the level of search completeness with a variation in the value of some document attribute(s), either between local values or with respect to an average value. To measure these parameters, then, requires that all documents meeting the specification (or selection criterion) must be fully known. Of course, if all documents were easily obtained, then the matters of completeness and bias would be of no interest; how one could be certain that they truly had all documents is not clear. Generally, it is only by omissions subsequently discovered that one can crudely estimate values for these

²⁹These 1338 documents include 1187 documents in the (main) topic collection, which is defined by the (final) graphical selection criterion, and 151 documents in four supplementary collections. Further details are given in Section 1, §2 of the Appendix.

parameters for any collection. More immediate estimates may be *suggested* by comparisons with other searches for other collections, or from known characteristics of the search procedures employed, for example if it is restricted to one database, its limitations will be at best those of that database. The following evaluations are necessarily subjective.

Bias in Publication Year, Language and Country. It has already been remarked that the collection is likely to be closer to completion from the mid-to-late 1960's to the mid 1980's, than before or after this time interval. Prior to this period, the capability for machine-searching is rare, and there are fewer A&I Services in existence and fewer still available to the present author; furthermore, the topic does not appear to have been recognised and documents indexed as such. Any bias against the literature of the late 1980's is in part due to delays in the actual publication and circulation of the literature, and its assimilation into later documents and their accession by A&I Services; also, later issues of some suitable secondary and primary serials were not available to the author, due to cancellations by libraries. It is felt that any bias against the earlier years is of little consequence since it predates the expansion of research on the topic; the bias against the later years is of more concern, and must be taken into account in the presentation and analysis of time curves for the literature. The collection probably also has some bias towards documents in the English language, and more so to those published in the U.S.A. and the U.K. -- especially, as will be discussed below, at the lower levels of topic aboutness. In part, this is due to the geographic location of the present author and the resulting accessible libraries and A&I Services, and in part, it is due to her inability to read non-English languages as fluently as English. However, it will be recalled that the topic itself grows from an English discovery, and citation patterns suggest that it is most studied in the U.K. and U.S.A. Further, as is shown later, the non-English literature is quite well represented in this collection, certainly at comparable levels to that reported in other studies, and better so for Chinese and Japanese.³⁰

Inclusion Level and Completeness. As finally determined, for a document to be regarded as on the present topic, it need only possess in the body of its text *one* occurrence of the word-stem '-BRADFORD-' (or an equivalent string of symbol(s) in the Cyrillic, Hebrew, Chinese or Japanese scripts) in close proximity to one of six other

³⁰The author *suspects* that some countries are under-represented in the collection. For example, documents were uncovered on the topic by Israeli, Cuban and Ukrainian authors, but runs of the appropriate serials could not be located for systematic searching. An absence of documents on the topic from certain nations in Europe (e.g. Portugal), the Middle-East (e.g. Turkey, Iran, Egypt, and Pakistan) and East-Asia (e.g. Malaysia, Indonesia and Korea) is also suspicious.

word-stems. This is a very low level of topic-aboutness for inclusion, requiring the present author to scan hundreds of thousands of pages, in many languages and several scripts, for occurrences of one short character string. The importance of additional alerting signs -- in the title, the abstract, the author field, headings, captions, figures, and bibliographic references -- can not be underestimated. The level for inclusion is generally below that required for traditional A&I Services to suitably index a document, and this puts an added burden onto searching in serials, searching in Citation Indexes, and reference tracing. The difficulty of finding documents with only limited mention of the topic in serials outside of the discipline of Library and Information Science further increases the burden on the latter two techniques. And as noted, a bias against the less important Library and Information Science serials in non-English language countries, both in accessible libraries and in the ISI Citation Indexes, still further increases the importance of reference tracing from documents on the topic, and from subject-related bibliographies produced for the literature of these countries. It is likely, then, that the collection will show a bias against documents at the lower levels of topic-aboutness, and more so against these documents written in non-English languages and published in serials not specialising in Library and Information Science. But at the higher levels of topic-aboutness, it is likely that this language and subject-scope bias will be quite small, and that the collection will be nearer completion.³¹

Inclusion and Importance Levels and Completeness. If the importance of a document on a topic to the implementation of research published on that topic is measured by the number of citations the document receives from the later topic literature (possibly corrected for document age), then the analysis of the references in §4, p.92, has an added dimension. For it is difficult to see how documents could be influential in research on a topic and not be detected in the technique of reference tracing used here. If these documents were also quite strongly on the topic, for example, in the present case, if they were generous in their usage of '-BRADFORD-' in the appropriate context, then

³¹Even documents at higher levels of topic aboutness can still escape detection. A near-example of this is provided by Novakova (1971), a 65-page review on the use of citation counting in the analysis of information sources in the pharmaceutical sciences. This article contains a headed section of text on Bradford's Law, a figure displaying a bibliograph, six occurrences of '-BRADFORD-' in the text and one each in its three foreign-language abstracts, and 18 of its 122 bibliographic references to documents known to be on the topic. Yet this paper, in Czech, is published in *Knihovna [Library]*, an irregular serial published for the State Library of the Czech Republic, with articles on scientific and theoretical aspects of libraries (see Podzimek & Mackova, 1980). At the time, according to Janzing & Brown (1975), this serial was indexed only by *Informationsdienst Bibliothekswesen* (1971-), an East German specialist A&I Service unavailable to the author. Further, the paper was not referenced by any document on the topic to hand, though there are 23 such documents in Czech and Slovak with references analysed. This paper was found fortuitously by browsing issues of the serial (from No.3 in 1962 to No.14 in 1988), which again was found fortuitously in Fisher Library, University of Sydney, only in the week before the collection was closed.

it is even more difficult to see how they could escape detection by the above search.³² It is likely, then, that the collection will show a bias against documents of low importance to research on the topic (as so defined), and more so against these documents when written in non-English languages. But for the more important documents, it is likely that any language bias will be negligible, and that the collection will be nearer completion. Combining this conclusion with that of the previous paragraph, we might afford high confidence to the statement that the (sub)collection of documents both strongly on the present topic and influential in its later research is very nearly complete.

Conclusions. More generally, we can assert that statements about the probable levels of completion attained for a subject collection will depend not only on the search strategy employed and the search effort made, but also on the selection levels chosen for inclusion and for importance. Even if these notions are not defined, they would influence selection decisions when a subject collection is compiled. For example, one suspects that documents subjectively evaluated as not being strongly on a topic are more likely to be selected for a literature collection on that topic if the compiler subjectively evaluates them as of importance to his/her understanding of the topic. Where these notions remain unspecified, or only broadly specified, it seems inappropriate to attach even broad numerical estimates to the level of completeness of a subject collection. For the present topic, given the multiplicity and thoroughness of the search procedures used, the author feels reasonably confident that most analytical-level documents strongly on the topic (as defined), which were published in the specified time period, have been gathered in the collection. This confidence is increased if we also require these documents to be of some importance to research on the topic (as defined). As documents become more peripheral to the topic, both in terms of amount of discussion in their text and in their subsequent influence as expressed through citation, then the level of completion most probably falls, and a bias against documents published in less well-known and non-English language serials in Library and Information Science, and serials with other subject scopes, most surely develops. These probable biases will have to be taken into account in later analyses.

³²Nevertheless, the situation may arise, e.g. as follows. We can imagine a set of documents, each with only a small segment on the topic, and perhaps published in obscure serials, and/or serials remote from Library and Information Science, to persistently cite as a source for this text an otherwise ignored document on the topic. While it is highly improbable that such 'citing islands' would occur for the main subject focus of these papers, they might do so for a persistent peripheral theme, especially if the papers were written by one author or a small local author group. Whether the persistently cited but otherwise ignored paper on the topic could be regarded as truly influential on the research on the topic in this case is debatable. And it does strain credibility that even this situation should remain undetected in the citation-conscious 1970's and 1980's, especially if the source document was strongly on the present topic.

Postscript: Is the Collection Formed Comprehensive? A stated goal of the search was to form a comprehensive collection of analytical-level and scholarly documents on the topic of Bradford's Law of Scattering. Can anything be said about the achievement of this goal? As noted, collections of documents are assembled for some purpose; whether it is at all necessary for a subject collection, however defined, to approach completion depends on that purpose. An introductory bibliography of a topic would perhaps only require a partial listing of the most influential documents written in the language of the bibliography; a bibliometric investigation of a collection parameter considered to have a normal distribution (if any there be) would only require small random samples of documents from different topics, taken perhaps from a broad database. The purpose of this collection is the quantitative analysis of the production processes underlying the generation of a subject literature at the analytical-level, with the present topic providing one case. For this purpose, I believe that the collection formed must be sufficiently comprehensive so as to display a diversity of sub-collections, each possibly showing a pattern of concentration and dispersion that is broadly characteristic of the whole literature. Despite the likely biases mentioned, it does appear from preliminary analysis that this goal has been adequately met.