Indexing and abstracting:
a current-awareness bibliography

Hans H. Wellisch

A&I SERVICES AND DATABASES

This section held over for technical reasons from part 6. See also additional entries on A&I services at 663–678. (See also 515, 521, 545, 546, 549–533, 556)

The director of INSPEC considers desirable qualities of databases, the expectations of their users, inhibiting factors, and quality improvement in his Miles Conrad Lecture at NFAIS' 30th annual conference in February 1988.

Analyzes a new A&I service for the social sciences, ASSIA (Applied Social Sciences Index and Abstracts), notes research into the information-using habits of social scientists, and reviews some other A&I services in that field.

An analysis of the coverage and indexing of 30 pharmacological papers by BIOSIS, CAssearch, Embase, and MEDLINE online and their printed versions showed that only MEDLINE, Index Medicus and Embase had adequate coverage of letters. Indexing of drug formulations and syntheses, clinical trials, pharmacokinetics, drug synergy and other topics was often inadequate and varied considerably. Online searching by natural language words proved to be valuable.

Describes three demographic databases: CENDATA, Donnelly Demographics, and Urban Decision System’s ONSITE. Includes a table comparing characteristics of 16 demographic databases.

Surveys World Patent Index, Chemical Abstracts, and the Claims files on Dialog, Inpadoc, and Patsearch regarding their coverage and scope, as well as the command languages used, graphics searching, and search aids. Results are displayed in a comparative table.

Discusses the uses and advantages of InfoTrac in a university library. The database covers more than 1000 periodicals, beginning in 1982. It is very expensive (more than $15,000/year), has serious problems of vocabulary control, and cannot be searched by either author or title.

Discusses the indexing policies of PROMT and Trade & Industry Index regarding the journals Online and Database.

A review of indexes to arts and letters, films, historical subjects, magazines, political and social science shows that coverage of Soviet materials in English is deficient.

The Beilstein Handbook, one of the world’s largest organic chemistry databases, uses an abstracting system based on paperless, decentralized organization for factual and numerical data.


Subject headings from MEDLINE and NAHL files are compared to determine coincidence and number of headings assigned. Suggestions for the best use of both databases for optimal retrieval results are given.

Part 1 deals with general and news databases; Part 2 examines 20 application databases and Part 3 covers five comprehensive computer product databases.

Although MDC’s primary product is full-text online service, abstracts and structured files in its new Reference Service complement and add value to the full-text material.


Online database structures do not match the structure of disciplines closely enough, as demonstrated by the discipline of tropical medicine. Better indexing of signs and symptoms mentioned in articles is needed.

The Indexer Vol. 16 No. 3 April 1989 181
GENERALITIES


The review covers bibliographies of A&I, relevant British Standards, the Wheatley Medal and Carey Award, some major indexes produced during the quinquennium, the impact of computers, the Data Protection Act, automatic indexing, NEPHIS and PRECIS, thesauruses, design of indexes, education and training, the indexing process, the profession, and A&I services.


Anaphora are abbreviated subsequent references used in discourse to avoid repetition. They are easily understood by humans but are not interpreted in current free-text search systems. This study tested the basic assumption that anaphora are used by authors to refer to concepts integral to the topic rather than to peripheral concepts. The results support the hypothesis and suggest that resolution of anaphora may be of greater importance in future free-text systems possessing fuller natural language processing facilities.


Discusses various aspects of optical information systems, including the contents of databases, retrieval software and indexing.


INDEXING LANGUAGES (see also 621)


Discusses the differences between classifications, category codes, subject headings, descriptors and post-controlled vocabularies and the use made of these controlled indexing languages in various major databases.


The study sought to evaluate the feasibility of transforming LCSH into an indexing language in which the thesaural and syntactic elements are separated, in order to create a rigorously controlled syntactic structure that would make subject searches in online catalogs more user-friendly. An analysis of a stratified sample of LCSH headings for films showed that about 70% could be retained unaltered; less than 50% of incompatible headings would require revision before benefits of an improved syntactic structure could be realized.


The French national standard for the construction of subject headings, the form of proper names, corporate names, and abbreviations. Contains also rules of punctuation and alphabetization.


If a document is indexed by the terms A and B, it represents an overlap between the set of postings for A and the set of postings for B. The patterns of overlap between terms which are closely related were observed in the MeSH thesaurus with respect to selected MEDLINE postings. The impact of the overlap patterns on the potential effectiveness of a proposed algorithm for handling negation is considered.


Examines the conditions under which the two types of indexing languages are used when searching a bibliographic database online. The two types not only complement one another, but both are necessary for effective retrieval. The use of one type or the other (or both) also seems to depend on the conditions of the search request.


Subject headings assigned by the University of Michigan Library to almost 4000 bibliographic records were analyzed for a match with LCSH terms. 44% of the sample matched completely with LCSH's 10th edition, and for headings without subdivisions the match was 88.4%. 93.6% of subdivisions belonged to the 'free-floating' list of such subdivisions. A machine-readable file of such a list is recommended as a tool for automatic subject heading assignment.


Discusses subject heading lists and thesauri in the light of linguistic and bibliographic analysis, reviewing the major contributions of American, English, and German authors.


LCSh in machine-readable form are being integrated into online catalogs but users mostly fail to match LCSH headings; they do not know that there is such a list as a source of search terms; they want lists of related terms; and only part of a library catalog's subject headings match those in LCSH and their cross-references. Proposals are made to achieve better matches between users' search terms and LCSH terms.

A textbook on the formulation of subject headings, the use of the French general thesaurus RAMEAU (Répertoire d’autorité matière encyclopédique et alphabétique unifié), and the French national standard NF Z 47–200 (see 585).


There is little overlap between two clusters of documents retrieved from the Canadian OON database which were indexed by title keywords and by Library of Congress subject headings respectively. Both access modes must be used to maximize recall.

INDEXING SYSTEMS


The computerized version of the Indian POPSI indexing system known as Deep Structure Indexing System (DSIS) is currently being compared with PRECIS at Loughborough University, using a microcomputer. A sample of 600 documents from different social science fields are used for the test. Some faults of DSIS which were discovered during the study are being discussed.


Associations between scientific concepts in new interdisciplinary research can be accurately translated into word maps (graphic representations of the logical processes involved). This opens new possibilities for database searching, as opposed to conventional keyword searching, or the use of expert systems. The French original of this article appeared in Documentaliste 24 (1987): 11–16.


Latent semantic indexing tries to overcome the incompleteness and imprecision of keyword matching by considering the higher order, or latent, relations among terms and documents. The model of latent structure uses a reduced set of about 100 descriptive parameters derived by linearly decomposing a term/document matrix. Each term, document and query is represented by values on this smaller number of orthogonal dimensions. These dimensions may be thought of as artificial concepts which represent the common meaning component extracted from many different terms and documents.

PRECIS

(see also 594)


PRECIS is being introduced in the National Educational Information System of Hungary. Its adaptation for Hungarian libraries is being considered.


In a small-scale experiment comparing PRECIS in the British National Bibliography to a general thesaurus and several special thesauri it was found that PRECIS works best for searches in specialized science areas, while general and social science queries are less well handled by it. Simple subject searches for only one retrievable item are handled well by PRECIS, but the dispersal of related subjects in the Dewey Decimal Classification is not quite compensated for by the verbal approach of PRECIS.

CITATION INDEXES

599 ROSEN, T.H. Coming soon: the Science Citation Index on compact disc. INSPEL 22 (1988): 208–211.

The CD edition of SCI will be an enhanced micro-computer-based version of the printed index. It will provide access to 600,000 source articles from 3,300 scientific and technical journals.

INDEXING SOFTWARE


A brief history of the development of indexing software, an evaluation of top-of-the-line and low-cost programs, and updated information on Index Preparation System and Macrex.


The University of Miami Library has developed a computerized keyword index to U.S. government publications (KWICIE) which can be used with IBM personal computers. Copies of the program are offered for free.


Discusses the pros and cons of DMS Delta and Masterfile III, with examples of index entries formatted by the latter program.


WordCruncher consists of 2 parts: IndexETC and ViewETC. The first part can create an inverted index from a textfile, the second part retrieves words, phrases, list of words, and contextually defined groups of words.


A review of the software package Tracer which supports flexible filing and is designed for office file systems and research project data.

COMPUTER-ASSISTED INDEXING


NEPHIS uses the computer to formulate index entries, freeing the indexer to concentrate on the intellectual task of analyzing text and naming important features. Novice indexers can learn the system quickly and use it to produce acceptable indexes.


A computer process generates a set of indexing terms selected from the NASA thesaurus based on the text (titles and abstract) of technical reports. The computer-generated terms are reviewed by human indexers.

The Indexer Vol. 16 No. 3 April 1989 183
AUTOMATIC INDEXING
(see also 621)


AIR/PHYS, an automatic indexing system for a physics database, assigns descriptors from a prescribed vocabulary to English abstracts. Its fault-tolerating features, its construction and the development of its thesaurus, as well as some problems in practical application are discussed.


Examines two approaches to the calculation of discrimination values in selecting potential index terms in automatic indexing. Concludes that a centroid approach would yield the best results.


Term discrimination values must be calculated for use in automated retrieval systems. The algorithm described can be used when document terms are weighted according to a particular scheme. Its running time is proportional to the number of documents and the number of assigned terms.


A document component is defined as any text unit that can be judged as to its relevancy and is independent of other components in the same document. The theory derives approximately the same weighting schemes as those suggested by earlier researchers.


Document components (terms, sentences, the whole document) have been tested for indexing and retrieval by probabilistic similarity measures. Terms and sentences performed substantially equally, and both are superior to the document as a whole.


Describes frame-based knowledge representation methods of information retrieval, research on user interfaces and automatic document classification as performed in the FERRET project of the Carnegie-Mellon University, using a text-skimming parser.


Single terms in text are not always adequate for correct retrieval, and term relations or associations are therefore often used, but their complexity leads to difficulties in retrieval. Replacement of existing text analysis procedures by artificial intelligence approaches is likely to be even more difficult and will not lead to better retrieval.


Discusses a book on abstracting by V. P. Leonov regarding the means of automating the abstracting of scientific and technical texts, the construction of thesauri, and the automatic creation of indexes.


A formal computerized language is used to generate a transformation table for converting statements indexed according to the SNOMED system into the International Classification of Diseases (ICD).

NAME INDEXES


Explores differences in precision of terms used in subject access in the humanities. Proposes 5 categories of terms, differentiated by precision which is based on the exactness with which phenomena can be located in space and time. Samples from 7 A&I services in the humanities are analyzed as to their precision, and implications for indexing in the humanities are discussed.

SUBJECT INDEXING
(see also 584, 585, 588-593)


Discusses discrepancies between how metal alloys are indexed for CA, and how metallurgists describe them.


Speculates on the way in which subject indexing in libraries may be affected by online catalogs and the indexing practices of bibliographic databases.


Discusses the level of subject access in current online public access catalogs (OPACs), the need for enhanced subject access, research on enhancement methods, obstacles to improvement of subject access in OPACs, and suggestions for future improvement.


Proposes several ways for improvement of subject access and help for the searcher in query formulation: frequency ranking of terms combined with the Subject Access Project (SAP) principles developed at Syracuse University and subsequently applied in Scandinavia, as well as an automated analysis system, ZOOM, for enhanced indexing.

Rejects the use of the Social History and Industrial Classification (SHIC) and recommends the use of a subject heading list compiled by the Department of Urban Archaeology (DUA) for the indexing of archaeological documents.


Presents analyses that demonstrate the potential of bibliographic contents and summary notes for enhancing subject access in online catalogs. Conclusions: (1) summary and/or contents notes did not occur in more than 5% of MARC records; (2) these notes occurred less frequently in local libraries' MARC records than in LC records; (3) when summary and/or contents notes occurred in bibliographic records, they contributed a significant proportion of unique words; and (4) subject-rich words added to MARC records from the DDC contributed a large proportion of unique words regardless of the presence of summary and/or contents notes in the record.


A comprehensive, all-purpose theoretical code for Library of Congress subject headings may be developed in the 1990s. Though such a code is much needed, there are many obstacles to its development.


A future subject heading code should consist of two parts: a main volume explicating the basic theory underlying LCSH, stating specific rules for the structure of headings, and addressing special subject areas; and a loose-leaf section containing lists of headings, subdivisions and their application.


Contains a study on the restructuring of LCSH by M. Dewar (see 584), an annotated bibliography on 'Bibliographic instruction for online public access catalogues' by F.C. Nowakowski, and an introduction discussing problems of subject access in online catalogs by M. Dykstra.


An annual review of classification, indexing and subject authority control, citing about 150 references.


Part of a research project exploring the improvement of subject retrieval in online catalogs.

 THESAURIS
(see also 660)


Discusses LEXX, an indexing and retrieval system for medical information, used at University of Pittsburgh's Medical Archival System.


Lists 122 controlled vocabularies (not all of them thesauri, but also subject heading lists, keyword lists, etc.) used by a large number of online databases. Gives details of sponsoring organization, edition, date, update frequency, publisher, physical format, price, ordering information, and examples of vocabulary.


A classification of degrees and types of descriptor equivalences for the construction of multi-lingual thesauri, with a description of an algorithm for establishing equivalent descriptors in two or more languages. Possible applications are indexing of multilingual documents and searching in multilingual databases.


Paradigmatic descriptor relations of organic compounds (synonymous names, genus, species, associations) are automatically established by an algorithm. Redundant or faulty relations must be manually removed.


A trilingual thesaurus of almost 65,000 terms and 6400 synonyms to be used in the PASCAL system for indexing and searching. The base volume is French-English-Spanish, two others are indexes in English and Spanish referring back to the base volume.


Nine articles on thesaurus construction and application. Titles translated: New thesaurus approaches—fresh breeze in old sails?—Thesauri and knowledge-based systems; thesauri and knowledge bases.—Object-oriented modelling of a thesaurus based on a frame system with a graphic user interface.—Automatic indexing and links to thesauri.—Unconventional thesaurus relations as orientation aids for indexing and research; an analysis of selected examples.—The Berlin-Thesaurus as an example of a complex regional thesaurus.—PROTERM: a software package for construction, updating, and handling of thesauri and other vocabulary lists.—INDEX: a program for the construction of dictionaries and indexing languages on personal computers.—Looking and acting with the CICERO thesaurus.—All articles carry English summaries.

INDEXING SPECIFIC FORMATS


Discusses international computerized indexing systems for visual materials, and the technology available for reproducing images (videotex, videodisc, and digital storage). The interfacing of these technologies with online, post-coordinate retrieval systems could facilitate much better indexing of visual materials.


A sequel to an earlier article in the same journal (March 1983, p. 52–55) describing further developments in the computer-based indexing of vocal sheet music by title, composer, lyricist, first line and date.


Describes the design and compilation of the concordance to Gilbert and Sullivan’s Savoy operas.


‘Hard’ indexing of pictures is concerned with description (what the picture shows); ‘soft’ indexing relates to meaning (what the picture is about, personal feelings and reactions to it).


Analyzes the form subject headings for AV media used in Anglo-American and German cataloging practice, their use in various libraries with large AV collections, and proposes a scheme of 8 dimensions for such headings.


Experts in 5 different disciplines were asked to submit verbal characterizations of 120 early sound recordings on cylinders. Results indicated that the experts used substantially different vocabularies but there was also considerable overlap in the subject headings used for pairs of recordings. A faceted scheme for content representation of recordings is proposed.


The history of the *EC Index* and the reasons for its demise.


Describes the development of a coordinated indexing system for a collection of photographs, postcards and pictures. Entries are arranged by peoples, places and subjects, supported by a more specific subject classification.


A list of prescribed terms used by the Library of Congress for indexing the genres of movies.

SERIALS & NEWSPAPER INDEXING


Few users are trained to make optimal use of indexes. Idiosyncrasies of different kinds of indexes, and problems encountered by different user approaches.


A critique of periodical indexing which may result in less rather than more or better information for the user. The cost of conducting searches may define and restrict what is included in an index.


INDEX PRODUCTION


The index was prepared on an IBM PC-XT, using the IBM Professional Editor, custom programs written in IBM Pascal, and TeX, a typesetting program. Entries were chosen and keyed in by indexers, then alphabetized, formatted and typeset by computer.


Reports on preparatory work on a complete index to books published in the USSR since 1917 which will fill some 120 volumes.

USE OF INDEXES & ABSTRACTS

(see also 630)


Few users are trained to make optimal use of indexes. Idiosyncrasies of different kinds of indexes, and problems encountered by different user approaches.

ECONOMICS OF A&I


How to develop a marketing plan for an A&I service, based on the experiences of BIOSIS.

INDEXERS

655 How I became an indexer. Indexer 16 (Oct. 1988); 117-122.

Seven indexers describe how they joined the ranks of indexers.


A brief biographical sketch of the author of The technique of indexing (1904), followed by her article on 'Indexing as a profession for women', published in Good Housekeeping in 1923.


ABSTRACTS & ABSTRACTING


An investigation into the possibility of automatically detecting how concepts exist in relationship to each other in abstracts. The goal is to capture these relationships in structured representations of abstracts' contents so that users can require not only that the concepts of interest to them co-occur in the retrieved documents, but also that the roles they play in relation to each other are the ones of interest. A discourse-linguistic analysis of a sample of 276 abstracts identified the lexical-syntactic clues which could be used by a system to automatically instantiate the frame-like structure of individual abstracts.


Reference librarians who are well acquainted with A&I services are uniquely qualified to write abstracts. The basic rules for abstract writing are presented, and the importance of choosing the right terms for online searching of abstracts is stressed.


This study seeks to determine whether a significant difference exists between the natural language text of a relevant abstract and a non-relevant abstract when judged using some similarity measure derived from a comparison with the user's natural language expression of an information need. Results seem to indicate that there is some power in the user's vocabulary to discriminate between the text of relevant abstracts and those which are non-relevant. The thesaurus enhancement proved less promising, but may serve as a technique for normalizing the variability of vocabulary so that other characteristics of the user's statement, namely word associations, will emerge as important discriminators.


Compares the abstracting and indexing of patents in 2 Soviet, 1 British and 1 Japanese A&I service, and suggests methods for improvement.


A reply to Moravcsik's article on abstract writing (see 563), reiterating the need for concise and informative abstracts, with examples of 3 abstracts of the same article, written in different styles.

A&I SERVICES AND DATABASES

(see also 618, 653)


An alternative measure to White & Griffith's term discriminating index (see 497) is proposed. The measure takes into consideration the size of the database, and its value can never exceed 1.


Describes the features of a new A&I service, Applied Social Sciences Index & Abstracts (ASSIA) which started in 1987. It covers issues affecting the lives of people, and thus differs from other A&I services which stress sociological phenomena.


Discusses the difficulties in accessing online information on women's issues across several databases. A project for a single database covering literature on women's employment, equality and gender is described.


Reviews InfoTrac regarding its coverage of information on Africa south of the Sahara, the list of periodicals covered, and examines some specific searches. Concludes that the database has many shortcomings as far as it relates to Africa.


The Canadian Women's Indexing Group is developing a new A&I service covering Canadian women's studies in response to a need expressed by Canadian librarians and researchers.


The Chemical Abstracts Service Chemical Registry System, operating since 1965, uniquely identifies chemical substances on the basis of molecular structure. Chemical Abstracts Service is now registering chemical substances cited in indexes to Chemical Abstracts prior to 1965. This effort will result in several hundred thousand additional chemical structures, along with their names, being available for online searching in the Registry File.


The front-end system, designed for end-user access to more than 20 medical and related databases was tested for a broad and for a specific medical topic. Comparisons are made with DIALOG-LINK, other front-end systems, their costs and specialist training requirements.


A brief description of this service which has expanded its scope and is now available online.

Compares retrievals on 11 topics in 5 databases: MEDLINE, Excerpta Medica, PSYCINFO, SCISEARCH and Social SCISEARCH. Searches were made with descriptors, cited references, and natural language terms common to all five databases. MEDLINE had the highest recall; all searches produced high precision; novelty ratios varied widely.


An investigation into the effectiveness of Australian A&I services covering librarianship and information science, and suggestions for improvement.


While many A&I services provide good patent coverage and indexing, the selection criteria for patents are sometimes questionable and bibliographic data may be inaccurate or misleading. Standards for identification of patent documents have been developed by the World Intellectual Property Organization and should be adopted by all A&I services.


Databases contain valuable linguistic information on terms, and can serve as substitutes or supplements to dictionaries and sources of linguistic statistical information.


Contains also a list of 50 core journals in immunology.


The availability of two full-text medical databases, MEDIS from Mead Data Central and CCML from BRS, provided an opportunity to compare the efficacy of the full-text to the traditional, indexed system, MEDLINE, for retrieval effectiveness. One hundred searches were solicited and the requests were searched on all three databases. The results were compared and preliminary analysis suggests that the full-text databases retrieve a greater number of relevant citations and MEDLINE retrieves a greater proportion of relevant citations.


On DIALOG’s reload of the MEDLINE database in 1987 which included a new indexing format for subheadings.


Discusses the various formats in which MEDLINE is available from U.S. search services and gateways. Provides lists of such services, CD-ROM availability, software packages, and foreign databases.

---

**Society of Indexers**

**training in indexing: open learning units**

**now available**

1. Documents, authors, users, indexers by Pat F Booth
2. Choice and form of entries by Pat F Booth and Mary Piggott
3. Information sources and reference tools by K G B Bakewell
4. Arrangement and presentation of indexes by Pat F Booth

**forthcoming**

5. The business of indexing by Pat F Booth and Elizabeth Wallis
6. Computer assisted indexing
7. Indexing of images


[Tuition and formal testing leading to the status of Accredited Indexer is available at extra cost to members of the Society of Indexers only.]

**price per unit £15.00 [SI members £10.00]**

available from Society of Indexers Subscriptions Officer
Mrs Sarah Mitchell-Cameron, 16 Coleridge Close, Hitchin, Herts SG4 0QX

---

*xlviii INDEXING AND ABSTRACTING BIBLIOGRAPHY PART 7*