Indexing and abstracting:
a current-awareness bibliography

Part 5

Hans H. Wellisch

GENERALITIES


A general model, resulting in two linguistic formulae for the representation of relationships between subjects.


A theme is defined as compressed information on a subject; its verbal expression is a thematic phrase, which includes titles. Subjects are represented by objects, properties, processes, methods, values and psychological factors. Distinguishes 7 types of a theme.


On the nature of textual information and its representation in the framework of interactive retrieval systems. A typology of levels of representation is proposed. Various levels may render the interaction between user and database more flexible.


INDEXING LANGUAGES


The MeSH vocabulary is updated annually. A study of changes showed that (1) new terms cannot be predicted by the increase in postings of existing terms, and (2) searchers’ and indexers’ use of the terms’ tree structure does not correlate with the terms’ distribution in the MeSH trees.

386 BERTRAND-GASTALDY, S. De quelques éléments à considérer avant de choisir un niveau d’analyse ou un langage documentaire. [Factors to consider before choosing a level of analysis or an indexing language.] Documentation et Bibliothéques 32 (Jan.–June 1986): 3–23. (FRE)

Surveys recent US, UK and French literature on subsystems for document analysis (classification, indexing) and retrieval from databases.


Describes ongoing research and development in Poland on: general problems of indexing languages (linguistic problems, typology, construction, compatibility, and indexing languages used in Poland); structure of indexing languages; indexing and information retrieval.


Graph-theoretic techniques are used to solve the minimum vocabulary problem in dictionaries (a problem in set theory). Potential applications in indexing are discussed.


A collection of articles old and new about LCSH, its development and use in the past, and possible improvements for online use.


The index terms in the List, intended to provide subject access to the national database of newspapers created by OCLC in 1985, are critically examined.


Free text and controlled vocabulary searching can no longer be viewed as antagonistic techniques in information retrieval since they both display advantages and weaknesses dependent on a fairly wide range of context, with the option to use both increasingly favoured. An attempt is made to present a list of features associated with the two techniques and to suggest a methodology to assist in deciding on the optimal retrieval technique for a particular purpose.


Indexing is a two-step process: recognition of the essence of a text that is to be retrieved; and representation of that essence in a form that is sufficiently representative and well predictable. The quality of indexing can be enhanced if concepts from a small number of semantically important categories are employed. Pre-coordinated index terms should be avoided, except those that are very common.
The proposed technique uses the log normal distribution equation, see item 329.

Results of a survey showed that most information centers and libraries use 2 separate indexing languages (IL). 2 tables list the types and kinds of IL used and a breakdown by major subject categories covered by the information centers.

A study of 7,834 LC records showed that, as subject specificity increased, length of LC and DDC class marks increased. There was no statistically valid association between number of subject headings per record and class notation length. For Part 1 of this study, see item 329.

The proposed technique uses the log normal distribution equation and the average frequency of each term.


Exploring the relationship between similarly co-cited references and co-assigned MeSH terms in over 8500 documents published in cardiology journals between 1975-77.

Considers the choice of indexing language for an IR system, and reviews both controlled and text-derived index terms, as well as the design of large databases and their indexing problems.


Subject access in the Delft University of Technology library in an online catalog is provided by both controlled and free terms and by classification. The importance of the controlled vocabulary is stressed, and a concordance between subject headings and class marks is described.

The Australian Defence Force Academy Library, disappointed with Library of Congress subject headings, decided to employ the Subject Access Project system, developed at Syracuse Univ.

Explains the application of Taube's Uniterm system on cards to the indexing of Arabic documents and linguistic difficulties in assigning Arabic keywords. Includes a table of various calendars currently used in Muslim countries.

A database of detailed descriptions of monographs based on words and phrases taken from contents pages and indexes, along the lines of Syracuse University's SAP project has been produced at Bar Ilan University in Israel.

The proposed technique uses the log normal distribution equation and the average frequency of each term.

The proposed technique uses the log normal distribution equation and the average frequency of each term.

Explores the relationship between similarly co-cited references and co-assigned MeSH terms in over 8500 documents published in cardiology journals between 1975-77.

Examines the way an indexer selects an index entry by focusing on the final entry and its relation to the text. The hypothesis that term location within a paragraph will correlate with indexability, and that term frequency on cited pages will correlate, was not supported by the data studied.

A database of detailed descriptions of monographs based on words and phrases taken from contents pages and indexes, along the lines of Syracuse University's SAP project has been produced at Bar Ilan University in Israel.

Examines the way an indexer selects an index entry by focusing on the final entry and its relation to the text. The hypothesis that term location within a paragraph will correlate with indexability, and that term frequency on cited pages will correlate, was not supported by the data studied.

Examines the way an indexer selects an index entry by focusing on the final entry and its relation to the text. The hypothesis that term location within a paragraph will correlate with indexability, and that term frequency on cited pages will correlate, was not supported by the data studied.

Describes the indexing of International Pharmaceutical Abstracts by a combination of traditional methods and modifications required for full-text databases.

Describes the indexing of International Pharmaceutical Abstracts by a combination of traditional methods and modifications required for full-text databases.

Generates a framework for placing index commands in the text of a document and a general purpose indexing processor which transforms the 'raw index' into an alphabetized list.
INDEXING AND ABSTRACTING BIBLIOGRAPHY PART 5


A comparative and evaluative study of the pros and cons of several commercially available systems.

The full report on which item no. 315 is based.


Discusses the use of the MACREX indexing program for the production of a 5000+ entry index to archeological abstracts.


Offers suggestions for the right choice of hardware, including the type of PC, compatibility, quality of display and keyboard, discs, and printers.


Experiences in the use of an English automated dictionary for the indexing of texts, using a microcomputer.

AUTOMATIC INDEXING


A collection of papers on German research and practical uses of various methods of automatic indexing.


Considers logico-semantic connectors as a phenomenon of superphrase syntax. So-called 'false' connectors hamper automatic analysis of relations among phrases. An algorithm for the recognition of true connectors is proposed for effective automatic indexing and abstracting.


Indexing software that produces tables of contents and indexes from ASCII text files for short documents. It was found to be slow and needed manual postediting to eliminate control codes in printout.


421 VLEDUTS-STOKOLOV, N. Concept recognition in an automatic text-processing system for the life sciences.


Describes a natural-language text processing system designed as an automatic aid to subject indexing at BIOSIS. The system is intended to model deep indexing with a controlled vocabulary of biological concepts. The system encounters problems of lexical and situational ambiguities. The techniques used for disambiguation are discussed and demonstrated in examples.

SUBJECT INDEXING


Originally written as an introduction to the 11th ed. of the ERIC thesaurus (see no. 347), this review is intended to make users more aware of the system as a whole, explains the process of indexing educational literature, and discusses the role of the Thesaurus of ERIC descriptors in the information retrieval process.


Information on place names is widely scattered, and existing indexes are often inadequate.


Describes a detailed format for the systematic indexing of factors in interdisciplinary studies in the fields of criminology and mental health.


A listing of reference books under keywords representing often-asked questions, produced and updated on a PC.


On the problems and pitfalls of indexing taxonomic literature.


Experiences of the Bavarian Library Network with the new German rules for subject headings (RSWK).


A critique of the Local Historian’s index for its vol. 16 (1985/86) and the author’s model for a better index of places and subjects.


A comparison of 50 medical charts with MeSH vocabularies showed that about 50% of the words in charts were MeSH-related, and 40% were either MeSH terms or cross-reference terms used in MeSH.


Examines the Soviet standard GOST 7.45-84 ‘Subject analysis of documents: general requirements’ which has been in force since 1985. The standard covers all aspects of the topic, including the choice of an indexing language, syntagmatic and paradigmatic relations, arrangement of headings, and editing.

The role of classification in subject indexing is demonstrated by two annual indexes to the journal Het Boek (The Book). The first index, which subsumed specific topics under hierarchically higher ones, performed less well than the second index, which showed index terms in their context.

Investigates the possibilities of using various subject access systems in online catalogs. An experiment has shown that an expert system helping inexperienced users with search strategies can improve their success in finding relevant items, and reduce zero hits.

A comparative study found that more than 20% of records were thought to be in need of revision, about 11% needed added subject headings, and less than 5% had various errors that needed corrections. Public libraries made the most changes, academic ones the least. The cost-effectiveness of checking and changing practices is questioned.


An anthology of significant writings on subject analysis and indexing from Cutter and Dewey to recent researchers.

A complete revision of the first ed. of 1973, taking into account new national and international standards, computer aids, and the problems of multi-lingual thesauri.

Discusses the scope, structure and mechanics of building a multi-disciplinary thesaurus.

Lists performance criteria for an ideal thesaurus display. A study of bibliographies listing thesauri revealed that less than 6% contain graphic displays, despite their many potential advantages for users.

A simple and practical approach to the construction of a thesaurus for medical informatics.


A subjective study of ideas and attitudes relating to information retrieval and the thesaurus, c. 1940–1967.

As libraries bring specialized catalogs online, searching begins to replicate that of multiple databases. Problems of online catalog use by patrons as well by professionals, and various strategies to cope with those problems are outlined.

A conference on thesauri, convened by the European Economic Communities in 1986, concluded after an analysis of almost 1000 thesauri that 'too many thesauri are hampered by inadequate design, inappropriate exploitation, sometimes even non-use, insufficient maintenance, and inadequate development and updating of software.' The conference recommended ways to solve such problems and avoid their recurrence.

A conference on thesauri, published in...
The creation and design of an index to cartographic materials in microform.


Describes the indexing of a local illustrations collection with the aid of a microcomputer.


The special indexing scheme devised for the picture collection at the Dickens House Museum in London. The scheme is based on the Dickens House Classification, used for the classification of textual material.


Describes the activities of the indexing department of Research Publications, Inc, who produce a range of microform publications of academic collections, newspapers and periodicals. Indexing is based on a machine-held thesaurus and on Library of Congress subject headings.


Urges that the Library of Congress change its policy regarding headings for geographic areas that are important for subject indexing of maps.

NAME INDEXES


Reviews the practices of various Japanese sci-tech periodicals regarding translation of titles, rendering of authors' names (including different Romanization and Cyrilization, and indication of forenames) and possible means of standardization.

SERIALS AND NEWSPAPER INDEXING


Account of a seminar held by the British Library Newspaper Library.


INDEX PRODUCTION


The production of a current-awareness subject index in the Renewable Energy Resources Information Center in Bangkok, Thailand.


Beginning in 1986, the Deutsche Bibliothek changed its subject indexing system to the formation of keyword chains. The syntactic sequence of keywords follows only the fixed categories of the new German subject heading code (RSWK). Each chain can be automatically permuted. The necessary indicators, operators and other markers are displayed in a summary.


Finding aids to the papers of Emma Goldman at the University of California at Berkeley are produced by means of a database management system on a network of microcomputers. (Goldman was a prominent American anarchist and feminist.)

USE OF INDEXES AND ABSTRACTS


Discusses the many different forms in which research results may be published and how A&I services (printed or online) can be used to find them.


Not what the title seems to promise but a report on a survey among academics at 5 British universities and their personal files. Only about 10% use computers, the rest keep card files. 14 software packages for personal files are examined for their suitability.


Aimed primarily at medical users, but also very useful for anyone searching the chemical literature. Contains many examples.


Asserts that indexing limited to the representation of aboutness serves novices in a discipline well but does not serve the researcher or scholar concerned with specific aspects or points of view.

ECONOMICS OF A&I


INDEXERS

EDUCATION AND TRAINING


HISTORY OF A&I


Traces the development of modern indexing and retrieval methods from the first (unpublished) paper on coordinate indexing to full text searching by computers.

ABSTRACTS AND ABSTRACTING


The CONSER A&I Project began in 1983 and ended in 1986. It provided A&I services with standardized information on journal titles.


The Soviet standard GOST 7.9-77 stipulates 500–2500 characters as the maximal length of an abstract. The 1,200,000 abstracts published annually by the services of VINITI conform to this standard. The rules of 7 other large databases from the US, UK, and France are compared with the Soviet standard. Only the French PASCAL has a low average of 370 characters, while most of the others have an average of between 700 and 1000. INIS-Atomindex has the most extensive abstracts with an average of 2000 characters.


Brief account of the standard’s history, rationale and content.


Gives a classification of automatic abstracting methods: statistical (absolute or relative frequency of words or phrases); positional (position of most meaningful part of text); linguistic (syntactic or semantic methods based on meaning in context); and hybrid methods using various combinations of the first three methods.


Anaphoric references in 600 abstracts randomly taken from PsycINFO and INSPEC had a mean frequency of 3.67 per abstract. Testing of rules for automatic recognition of actual anaphora showed that future algorithmic recognition of anaphoric terms is possible.


Explores logico-semantic aspects of the preparation of abstracts of scientific documents. Proposes a theoretical model which measures the influence of subject, object, goal and recipient on the structure of an abstract.


Identifies 7 types of summaries: sequential (abstracts, precis, secretarial minutes, digests) and synthesizing (locational, restructuring and abridging digests).


Distinguishes between objective and subjective types of information to be considered in the writing of abstracts, and establishes a logical scheme of their interaction, as well as a technique for the detection of possible errors of judgement by an abstracter.

479 WEBER, F. Zu inhaltlichen Anforderungen an automatisierbarkeit Patentreferate. [On requirements for the content of patents suitable for automatic processing.] Dokumentation/Information (68) (1986): 140–156. (GER)

Considers the structure, content and quality of patent abstracts in the German Democratic Republic and why they do not meet the requirements for automatic information retrieval.


Describes VINITI’s practices of abstracting patents for its A&I services. VINITI receives patents in 17 languages from 23 countries, constituting 20% of its total coverage (in 1985).

A&I SERVICES AND DATABASES


FINIS (Financial Industry Information Service) serves the information needs of the financial services industry, and is available on Dialog and on Meda Data Control’s Reference Service. Scope, sources and coverage are reviewed and special searching features are displayed.


Discusses the criteria for the design, implementation and marketing of online databases.

483 HAWKINS, D.T. Online information retrieval bibliography. 10th update. Online Review 11 (Supplement to no. 4) (Aug. 1987): S1–S89.


485 KOCHETOVA, E.K.; POTANOV, V.M. [Some comparative data on the abstract journals Khimiya and...
The two A&I services are compared regarding their structure, number of primary sources, total number of abstracts, and time lag. CA reports US sources very quickly but is slower than Khimiya in abstracting French, Dutch and other foreign material (based on a comparison of 100 organic chemistry abstracts).


Discusses and explains a British family of business-oriented databases, including examples of searches.


Combined Health Information Database (CHID) is devoted to professional and patient education, general health education, and community risk reduction. It is accessible through BRS.


Compares coverage, publication delay, probable causes of indexing and retrieval failure in the two indexes. While 98% of IM citations appeared within one year of publication, 79% of the BB citations appeared 2–3 years after publication. 34% of the sample appeared in both indexes, 17% were missed by both. 28% were unique to IM and 21% were found in BB only.


Compares the two databases and the online version of BPI with its printed form. BPI may be more cost-effective but ABI/INFORM is more widely available. Neither database is comprehensive, but they tend to complement each other.


Analysis of 60 popular health journals covered by the Consumer Health & Nutrition Index. Most of the journals are not well known because they are not well indexed, and they are not well indexed because they are so little known. The CH&NI now provides better subject access to these periodicals.


Describes the H.W. Wilson Co.’s indexing services now available both online and on CD-ROM.


Compares 6 A&I services covering periodicals on marriage and family issues, especially indexing terms used and extent of coverage. The most specialized service is Inventory of Marriage and Family Literature but it has no abstracts. Sociological Abstracts is deemed to be cumbersome to use, though its indexing is thorough. Psychological Abstracts, Sage Family Studies Abstracts, Social Sciences Citation Index and Social Sciences Index are also reviewed and the Thesaurus of Psychological Index Terms and Thesaurus of Sociological Indexing Terms are evaluated.


Articles on online searching are published mainly in Online Review and Database. LISA assigns more terms/article, while ERIC indexes more concepts/article. A small number of terms is used frequently by both systems, and both are consistent in assigning unique index terms.


A comparison of index terms assigned to clusters of articles on the same topic in MEDLINE, PsycINFO, BIOSIS and Excerpta Medica showed that BIOSIS assigned terms more indiscriminately than the others and that they were less like natural language terms used by MEDLINE. Excerpta Medica applied fewer but highly expressive terms. PsycINFO gave the fewest terms, while MEDLINE assigned the most and its quality of indexing was deemed to be the highest.


IRESIE is an index of journals in higher education and educational research, founded in 1979 by the Mexican University Center of Educational Technology, covering about 500 Mexican and foreign sources. 68% of which are in English. About 800 articles are indexed in each bimonthly issue.


Compares the Canadian Periodical Index and the Canadian Magazine Index for both manual and online searches. Libraries should buy and use both, since the two indexes complement each other and neither covers Canadian periodicals exhaustively. Both are reasonably priced.

Published as a supplement to The Indexer