Indexing and abstracting: a current-awareness bibliography

Part 6

Hans H. Wellisch

GENERALITIES


A review of the main features of the 1988 revision of the British indexing standard (see 501), by the chairman of the revision committee.


A revision of the standard, the previous edition of which was published in 1976.


Deals, among other topics, with the development of PRECIS, research into indexing language performance, and automatic indexing.


INDEXING LANGUAGES


Part 2 of item 387. Of 101 references, all but two are in Polish.


Among the three components of the semantic triangle, the concept is the central one for information retrieval. Concrete concepts have few expressions in natural language, but general and abstract concepts can be expressed in a variety of ways in natural language. Indexing languages are tools for increasing the predictability of a query expression. As in any translation, indexing requires good knowledge of the vocabulary and subject knowledge. Natural languages can be used in retrieval where high precision and recall are not required; otherwise, controlled indexing languages are better suited, especially for concrete subject fields.


A study of users' searching behavior in a university library showed that the importance of subject searching has been under-estimated in the past. It is concluded that the lack of interaction between different approaches in the searching process (the indexing language, the classification and titles) is the major obstacle to searching effectiveness.


Entry terms or near-synonyms used to access an indexed information store were evaluated by documents already indexed by MeSH and by an automatic indexer. For selection of new entry terms, terms from the Systematized Nomenclature of Medicine (SNOMED) were mapped onto MeSH and evaluated with the automatic indexer. The new terms did not help indexing, but a combination of MeSH with another thesaurus (on information science) did result in better indexing than MeSH alone.


INDEXING SYSTEMS


Explores the idea of using a conventional string indexing source description with a special phrase generator to generate multiple descriptor phrases. With proximity operators, these phrases should attempt to bring together groups of syntactically related words. Software for such a system is described.

PRECIS as a tool for content analysis is examined with a view to shift its rules from human to machine intelligence. It is shown that the classificatory aspects of PRECIS are of more potential relevance for future machine manipulation than the linguistic aspects (case grammar), although the latter will also play an important role. Extra-linguistic factors, e.g., whole/part and whole/dependent relationships have not yet been reduced to a set of rules. Most difficult are the 'to be' relationships. The present PRECIS classificatory rules cover all of these relationships, which shows that classification will have an important role to play in future automated content analysis systems.


Summary of item 513.

CITATION INDEXES


Claims that it is difficult to get coverage by ISI's Current Contents sections for new journals, and that the so-called 'impact factor' used as a criterion for inclusion by ISI depends also on various factors other than number and frequency of citations.

INDEXING TECHNIQUES


Analyzes three features of online public access catalogs: the varieties of bibliographic data processed; the kinds of indexes constructed; and the ways indexes are searched. No one configuration applies to every library project with equal efficacy or likelihood of satisfying queries.


How the physical medium affects the way indexes and abstracts are being designed and used.


Argues that Salton & Zhang's negative retrieval results using indexing by title and abstract only vs. indexing by terms from bibliographically related titles (see 238) are overly pessimistic. Discusses some strategies for term selection and gives reasons why cited titles are more useful than other types of related titles.

COMPUTER-ASSISTED INDEXING


The Central Abstracting & Indexing Service of the American Petroleum Institute (API-CAIS) has successfully applied expert system techniques to the job of selecting index terms from abstracts. Using the API Thesaurus as a base, a rule-based system has been created that has been in productive use since February 1985. The index terms selected by computer are reviewed by a human index editor, as are the terms selected by CAIS's human indexers.


A general discussion of computer-aided indexing, with examples drawn from the legal literature.

AUTOMATIC INDEXING

(see also 513)


Reports work on automatic text processing by a group of linguists at the Charles University of Prague, including the MOZAIC system of automatic indexing for Slavic and other highly inflected languages, and the ASIMUT system, which is claimed to be able to extract factual information from full text and to convert it automatically to a database.


A review of the development of the attempts to use machines for indexing. Stipulates nine 'postulates of impotence' which show why machines cannot formulate search requests or recognize meaning, and that fully automatic indexing and retrieval is not possible in principle. Also stated are three 'postulates of fertility', suggesting ways in which online systems could be used in an innovative way to retrieve information from literatures that are ostensibly unconnected yet may yield hitherto unnoticed relationships.

SUBJECT INDEXING

(see also 538, 561, 574)


Library of Congress subject headings reflect American society's current concerns and crazes. Dozens of new headings from the 1980s are cited, from Baby boom generation through Junk bonds and Pita bread to Taxpayer compliance and beyond.


LC's decision to replace see and see also references and their reciprocal x and xx codes with conventional thesaurus codes (USE, UF, BT, NT, etc.) does not change the chaotic nature of LCSH, and is even irresponsible and misleading. Future text analysis systems will be expert systems and as such must be rule-based. Solutions to the problems of LCSH will be discussed in a follow-up article.


A critique of the shortcomings of LCSH and suggestions for the 'killing of errors' (to which the 'euthanasia' of the title refers). Historical and geographical subjects are the focus of the critique, demonstrated by headings for Indian history. The Library of Congress replies (rather unconvincingly) on p. 21–23.


Subjects that are materials may be related to other such subjects in a generic or partitive type of relation, and also as collections,
all of which are static relations. Dynamic relations involve a time factor (processes and operations). A method of finding the subjects related to a given material subject is proposed.


Two subject heading lists for Catholic topics, published in 1937 and 1942 respectively, are now obsolete but are still used by some libraries. In light of bibliographic control by OCLC and other networks the validity of specialized subject heading lists is questioned.


THESOURI
(see also 542)


The co-occurrence frequency of natural language words and controlled keywords formed the basis for the latest revision of the JICST Thesaurus of Science and Technology, and was also useful as a tool for searching by natural language.


A study of the methods used to update a thesaurus on hand-capped persons by means of comparative evaluations of other thesauri and various sources of new terms and relationships.


The Chinese Thesaurus of synonyms classifies about 70,000 terms into three grades: the first consists of 12, the second of 94, and the third of 1,428 classes.


The Japan Information Center for Science and Technology (JICST) thesaurus has been updated and restructured by adding new terms, English translations of Japanese keywords, some Chinese terms, and indication of preferred synonyms.


Report on a conference on problems in thesaurus construction, maintenance, and their use in automatic indexing, retrieval, and artificial intelligence.


Maintenance and updating of a thesaurus entails comparisons and evaluations of other thesauri in the same or related field. ‘Connecting’ thesauri includes not only determining which terms in two thesauri are the same but also what kinds of relationships can be transferred from one thesaurus to another. Connecting is exemplified by the National Library of Medicine’s MeSH thesaurus which is now experimentally used to index medical titles, research and clinical titles, representation of allied health fields, and level of faculty involvement.

INDEXING AND ABSTRACTING BIBLIOGRAPHY PART 6

INDEXING SPECIFIC FORMATS
(see also 544)


On subject headings assigned to books of scientific illustration. Questions whether scientific illustrations are recognized as art by libraries and whether the subject headings are those that an art historian might expect.

SERIALS & NEWSPAPER INDEXING
(see also 547, 560, 569)


A report on a one-day seminar sponsored by the British Library, held in York in April 1987.


An expanded version of 257.

INDEX PRODUCTION


The State University of New York at Albany University Libraries Computer Search Service developed an index to terms used in six different databases to which the service has access. The index was compiled by using the Personal Pearl database management system.


The use of the computer-aided Lexical Dictionary developed at NASA for the compilation, editing and maintenance of the NASA Thesaurus, and in the indexing and abstracting of articles and reports listed in NASA STAR.


Discussing the indexing of decisions by tribunals and boards of the Ontario Ministry of Labour on cards and in printed indexes, and their conversion to an online system using BASIS software.


Census data recorded in the printed 1980 census reports, on magnetic tape and on microfiche were combined in an index produced with dBASE II software.


Describes the process of building a medical library’s own in-house database derived from MEDLINE tapes, decisions on inclusion of journals and years covered, balance between research and clinical titles, representation of allied health fields, and level of faculty involvement.

Medical Subject Headings (MeSH), Standardized Nomenclature of Medicine (SNOMED), and Current Medical Information and Technology (CMIT) are mapped onto a relational data model and stored in an Intelligent Database machine. Concepts common to the three knowledge bases are semi-automatically merged.


**FILING**


Discusses filing problems, sorting routines of computer software, and the shortcomings of the *Chicago Manual of Style* filing rules.

**USE OF INDEXES & ABSTRACTS**

(see also 508, 516, 517)


Six user studies conducted at Louisiana State University Libraries showed that most users agreed that InfoTrac could be used without training or instruction; 73% of users received successful answers; 88% considered InfoTrac preferable to other reference tools. Criticisms of InfoTrac in the literature are surveyed and analyzed.


Hardware (terminals and printers) outside of Japan are incapable of displaying Kanji and Kana characters, and sometimes even numerical information cannot be displayed. However, standard PC equipment can be programmed to display and print Kanji online data, and standardized text can be displayed in Roman letters.


In an experiment to determine the relative effectiveness of InfoTrac II on CD-ROM vs. Readers' Guide, students at a community college searching for answers to the same questions had 83% success with Readers' Guide, vs. 63% with InfoTrac.


The use of VU/TEXT in a public library to answer a tricky reference question within 14 minutes.


Concludes that national A&I services have only limited usefulness in a research library.


An expanded and revised version of 464.


The wide range of subject access methods available to Prestel users, options and constraints faced by information providers and by Prestel's information management team, and restrictions of the system itself.

**ECONOMICS OF A&I**


Examines the principles of controlled indexing in A&I services from the publisher's point of view and the need to minimize costs.

**INDEXERS**


An account of the publisher of reference books and indexes, its production methods and staffing, and a portrait of its founder.


Document analysts at Chemical Abstracts Service (CAS) must have a degree in chemistry, be able to discern new chemistry, be proficient in at least one other language, and have idiomatic English writing skills. The recruiting practices of CAS are discussed.


**HISTORY OF A&I**

(see also 524)


561 WYKOFF, L. Subject headings: a brief online history. *Medical Reference Services Quarterly* 6 (3) (Fall 1987): 69–73.

Discusses Boolean searches, text word searching, free-text searching, word adjacent searching, and proximity searching for headings and words.

**ABSTRACTS & ABSTRACTING**


Methods of eliminating syntactic and lexical redundancy in abstracts are discussed. Establishes rules of algorithms for the automatic editing of abstracts and elimination of redundancies.


Condemns the type of abstract that is really only a kind of table of contents and does not give specific information on results. Journal editors should reject such worthless abstracts and insist that authors of papers submit abstracts giving results.

Published as a supplement to *The Indexer*