Douglas, Kimberly; Wismer, Don
Computer-Based Indexing on a Small Scale: Bibliography.
Jan 80
ME01/PC01 Plus Postage.
*Automatic Indexing: Computers: Indexing:
*Microforms: *Permuted Indexes: Special Libraries:
*Thesauri

The 131 references on small scale computer-based indexing cited in this bibliography are subdivided as follows: general, general (computer), index structure, microforms, specific systems, KWIC KWAC KWOC, and thesauri. (RAA)
COMPUTER-BASED INDEXING ON A SMALL SCALE:

BIBLIOGRAPHY

by

Kimberly Douglas

and

Don Wismer

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Don Wismer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."
COMPUTER-BASED INDEXING ON A SMALL SCALE: BIBLIOGRAPHY

prepared by Kimberly Douglas
Librarian, Bigelow Laboratory for
Ocean Sciences/Maine State Department
of Marine Resources

and

Don Wismer
State Agencies Librarian, Maine State
Library

January, 1980

Prepared for the Special Library Group of the Maine
Library Association's workshop on small-scale
indexing by computer, held at the Maine State Library
in Augusta, Maine on January 14, 1980.

Contents: General, General (Computers), Index
Structure (Citation Indexing, Other Aspects),
Microforms, Specific Systems (PRECIS,
FAMULUS, Other Systems), KWIC KWAC KWOC,
Thesauri.

-------------------------------------------------

GENERAL

1. American National Standards Institute (formerly United States
of America Standards Institute).
   Basic criteria for indexes, ANSI Z39.4. NY, The Institute,
   1968.

   "Specifications for printed indexes". The indexer 9,3

3. American Society of Indexers, Committee on Ethics, Standards,
   and Specifications.

4. Atkinson, D.
   "Indexer-publisher relations: a two-way street".
   The indexer 8,3 (1973), pages 172-174.

5. Bakewell, K.G.B.
   Classification and indexing practice. London, C. Bingley;

7. Bernier, C.L.  
"Index specifications". The indexer 9,3 (1974), pages 9-12.

8. Bhattacharyya, K.  
"Some indexing/searching characteristics of special libraries in science and technology in the U.K.". ASLIB proceedings 26,3, pages 109-129.

This book evolved from the authors' exchange of ideas on the goals and content of library school abstracting courses in the U.S. Covers the following: characteristics and types of abstracts; instructions and standards; historical review of abstracting services; contents and format; editing, classification and indexing; publishing; automatic abstracting; abstracting services; career opportunities.

(MSL call number 029.5 B734i 1978.)

11. Bourne, C.P.  


14. Brown, A.G.  

15. Buchanan, Brian.  


17. Collison, Robert Lewis.  

18. Collison, Robert Lewis.  
19. Conrad, Carleton C.  
"Status of indexing and classification systems and potential future trends". J. Chem. Inf. Comp. Sci. 16,4 (Nov 1976), pages 197-201. The expansion of industry and government research programs since 1920, resulting in the production of a large body of research reports, has provided the impetus for the development of new methods of handling the increased volume of information. The following developments in indexing are reviewed: punched card indexes; uniterm indexing; specialized thesauri; vocabulary control; file manipulation via computer. Outlines future trends.

20. Dataflow Systems, Inc.  

21. Drage, J.F.  

22. Foskett, Anthony Charles.  


24. Gould, A.M.  


27. Kennedy, R.A.  

28. Knight, G. Norman, ed.  

30. Jahoda, Gerald.

31. Lancaster, F.W.

32. Lancaster, Frederick Wilfred.

33. Langridge, Derek Wilton.

34. Maizell, Robert Edward et al.

35. Maltby, Arthur.

36. McCollin, L.R.
   "The purpose of indexing". The indexer 1,2 (1958), pages 31-35.

37. National Technical Information Service (NTIS).

38. Oliver, L.H.

39. Salton, G.

40. Salton, Gerard.

41. Salton, Gerard.

42. Speight, F.Y.

44. United States of America Standards Institute.  
USA standard basic criteria for indexes. NY, USASI, 1969.  
12p. USAS 239.4-1968.


46. Vickery, Brian Campbell.  

47. Vickery, Brian Campbell.  


GENERAL (COMPUTERS)

49. Borko, Harold, ed.  

50. Borko, Harold.  

51. Braun, Stephan and Camilla Schwind.  
The extraction of index-phrases from texts with the help of a single word concept dictionary and a thesaurus containing relations among concepts is described. Results obtained show that phrase extraction from texts by this semantic method is possible and offers many advantages over other purely semantic or statistic methods concerning preciseness and completeness of the meaning with relation to the text.

52. Field, B.J.  
Describes techniques for the automatic assignment of controlled language subject headings from free indexing, using both English language and French language free indexing and thesauri. These techniques involve the automatic truncation, manipulation and fragmentation of the free index terms for their comparison with the thesaurus. Also discusses a study of methods for the automatic assignment of classification codes from English language free indexing.
53. Field, B.J.


A number of techniques have been studied for the automatic assignment of controlled subject headings and classifications from free indexing. These techniques involve the automatic manipulation and truncation of the free-index phrases assigned to a document and the use of a manually-constructed thesaurus and automatically-generated dictionaries together with statistical ranking and weighting methods. These are based on the use of a statistically-generated 'adhesion coefficient' which reflects the degree of association between the free-indexing terms, the controlled subject headings, and the classifications. By the analysis of a large sample of manually-indexed documents the system generates dictionaries of free-language and controlled-language terms together with their associated classifications and adhesion coefficients. Having learned from the manually-indexed documents the system uses these dictionaries in the subsequent automatic classification procedure. The accuracy and cost-effectiveness of the automatically-assigned subject headings and classifications has been compared with that of the manual system. The results were encouraging and the costs comparable to those of a manual system.


Several factors relevant to the evaluation and selection of cost-effective computer software are discussed. Topics considered include: usage rights, disclosure privileges, delivery and warranty terms, maintenance agreements, program releases and modifications, installation, and remote versus on-site usage.

55. Hines, T.C. and J.L. Harris.

56. Luhn, H.P.

57. Lynch, M.F.
"Computer-organized display of subject information". The indexer 7,3 (1971), pages 94-100.

"A program suite for the production of articulated subject indexes". Computer journal 16,1 (1972), pages 47-51.

60. Pao, Miranda Lee.


Goffman suggested an automatic indexing method based on the frequency of word occurrences in any given text. The most content-bearing words of the text would occur at the transition region at which Zipf's First Law of words of high frequency of occurrences begins to take on properties of words of low frequency of occurrences. An experiment was conducted to test this hypothesis. Word frequencies of 2 articles were analysed; results bear good semantic similarity to subjective human intellectual indexing of the same articles. This is a simple method which may be automated easily.

61. Ryan, V.J. and V.A. Dearing.

"Computerized text editing and processing with built-in indexing". Information storage and retrieval 10,5/6 (1974), pages 211-228.

62. Salton, G. et al.


A variety of abstract automatic indexing models have been developed in recent times in an effort to produce indexing methods that are both effective and usable in practice. Among these are the term discrimination model and the term precision system. These indexing systems are briefly described and experimental evidence is cited showing that a combination of both theories produced better retrieval performance than either one alone. Appropriate conclusions are reached concerning viable automatic indexing procedures usable in practice.

63. Salton, G.


64. Sharp, J.R.


Indexing for retrieval includes conventional cataloging, author entry, title entry, series entry, conventional subject indexing using classified and alphabetical catalogs, cataloging and indexing of non-bibliographic material, physical production of the catalog, Boolean retrieval, indexing vocabularies, indexing languages ancillary devices (hierarchical arrangements), thesaurus updating by computer, and automatic indexing.
65. Somerfield, G.A.
Discussess the production of abstracting and indexing (Ai) services, with special reference to the experience of the Commonwealth Agricultural Bureaux (CAB). The structure of a machine-readable bibliographic record is described. Reviews input techniques, including the use of visual display units, and the advantages of optical character recognition. Proof-reading, error correction and validation procedures are described, together with methods for selecting the records to be included in a particular Ai serial. For economy and accuracy subject indexing can be based on a computer-held authority file of authorised index terms and phrases.

66. Sparck-Jones, K.

67. Sparck-Jones, K.

68. Spranger, U.

69. Stevens, M.E.

70. Stibic, V.
The Information Systems and Automation department of Phillips, Eindhoven, is experimenting with automatic indexing using DIRECT, an in-house on-line retrieval system. The study involves two fields: patents on color television and information systems and automation. Some conclusions are: automatic indexing is much cheaper than manual indexing but the higher noise increases retrieval costs; the environment (document collection and enquiries) influences the feasibility of automatic indexing; noise is mainly due to lack of syntax; sophisticated syntactic devices are inappropriate but a simple syntax may be required to reduce noise; the ranking of output improves performance and has economic advantages; interactive methods also improve performance, but little is known about their costs; a combination of automatic operations with some human decisions seems the optimum; automated methods of analyzing retrieval results can improve the efficiency of information retrieval research.
INDEX STRUCTURE: CITATION INDEXING

71. Garfield, Eugene.
    "Citation indexing: its theory and application in science, technology, and humanities.

72. Garfield, Eugene.

73. Margolis, J.
    "Citation indexing and evaluation of scientific papers".

74. Weinstock, M.
    "Citation indexes". Encyclopedia of library and information science 5 (1971), pages 16-40. Published by Marcel Dekker, New York.

INDEX STRUCTURE: OTHER ASPECTS

75. Armitage, J.E. and M.F. Lynch.
    "Some structural characteristics of articulated subject indexes". Information storage and retrieval 4,2 (1968), pages 101-111.

76. Bernier, C.L.

77. Bernier, C.L. and E.J. Crane.
    "Correlative indexes VIII: subject-indexing vs. word-indexing".

78. Bernier, C.L.

79. Campbell, D.J.
    "Making your own indexing system in science and technology: classification and keyword system". Aslib proceedings 15,10 (1963), pages 282-303.

80. Farradane, Jason.

81. Farradane, Jason and Peter Gulutzan.
    Relational indexing has been tested for any distortion of meaning in the course of the indexing by constructing a computer program for conversion of the diagrams to a permuted alphabetical index. The
INDEX STRUCTURE: OTHER ASPECTS

methods involve 'translation' of the diagrams into English sentences which include all words and interrelations in each entry, without any alterations of any kind to the words or their interrelations. The results have fully confirmed the expectation that relational indexing does not distort the meaning of the subject matter in any essential way.

82. Veilleux, M.P.

MICROFORMS

83. Bolnick, Franklin I.

SPECIFIC SYSTEMS: PRECIS

84. Austin, D.

85. Austin, D.

86. Austin, D. and J.A. Digger.
"PRECIS: the preserved context index system". Library resources and technical services 21,1 (1977), pages 13-30.

87. Bakewell, K.G.B.
"The PRECIS indexing system". The indexer 9,4 (1975), pages 160-166.

88. College Bibliocentre, Toronto.

89. International PRECIS Workshop, University of Maryland, 1976.

90. Preschel, Barbara M.
Briefly describes PRECIS's aims and its two aspects, the syntax and the semantics. Those people attending the Maryland workshop (see 89) who are experienced indexers appeared to have much less trouble doing the exercises than those who were primarily librarians. The interesting point arising from the contributed papers was the number of times PRECIS was compared to the Library of Congress Subject Headings, the latter almost always coming off second best. The management of PRECIS indexing is also
mentioned. The individual aspects of PRECIS are not things that have never been done before; what makes the system innovative and powerful is that they have been fused to produce a flexible and integrated system. The mere fact that PRECIS exists makes it a candidate for a future universal computerized system of bibliographic control.


SPECIFIC SYSTEMS: FAMULUS


The document describes how the FAMULUS system (obtained free from the Pacific Southwest Forest and Range Experiment Station) was used to build a 4000+ reference file, which was published as "A Bibliography of Quantitative Ecology" (Schultz, 1976). Examples from the main programs show the system of checks, cross-checks, information retrieval, and printouts used for editing the bibliography prior to publication. Further it is hoped that the step by step presentation of the various main programs of FAMULUS, using actual examples, will both stimulate others to attempt bibliographies of their own specialized fields and show that FAMULUS is a worthwhile system to use in such an endeavor. The examples shown as well as the System User's Manual (see no.93) should make the FAMULUS system readily understandable to interested indexers.


FAMULUS is a computer-based system designed to support the documentation activities of the individual scientist with minimum interference in his information-organizing habits and instincts. It operates economically on the IBM 360/40+, the CDC 6400 and 6600, the Univac 1108 and Xerox's Sigma 7 computers. The user is offered editing, sorting, indexing, vocabulary-building, searching and file revision features in a package which leaves him or her free to structure input according to the local idiosyncratic needs. The Manual contains examples of all control cards and a description of the search strategy. Operation of the system sub-programs is described. The eight main features: 'Edit', writes punched card input onto tape, allows the user to make corrections, additions, and deletions; 'Sort', rearranges file order by changing the order of fields within records so the file can be realphabetized; 'Merge', provides updating facilities and permits enlargement of the files through merging two individual files into one master file; 'Galley', prints the file in any of several formats; 'Vocab', prints in alphabetic order the words in any given field of a file, making lists of index terms, keywords in title, etc.; 'Index', lists keywords and tells what records they come from, thus indexing the file; 'Search', scans stipulated field(s) in the records of a file, matching them against a user-prepared search question; 'Ossify', transfers tapes to punched cards.

ACCESS is a program for the catalog and access of information. The program is primarily designed for and intended to handle a personal library, although larger applications are possible. ACCESS produces a listing of all entries by locator code, so that one knows where to find the entry in his or her library (i.e., call or accession number), a listing of entry titles by user-specified category codes (enabling a hierarchical approach to subject classification), and a keyword-in-context (KWIC) listing with each keyword specified by the user. ACCESS is presently programmed in FORTRAN and operates on any IBM System 360 under OS (it uses the IBM SORT MERGE package). It is anticipated a machine language version, soon to be implemented, will greatly decrease the running time of the program.


The cataloging system of the document center of the Directorate of Aerospace Studies has been computerized on the CDC 6600 computers of the Air Force Weapons Laboratory. The system is unusual in that catalog listings of the holdings may be arbitrary in informational content, i.e. with regard to what types of information appear and their order of appearance. Although its initial application was to library cataloging it may be used to catalog any collection of objects which can be described by at most 20 categories of information. This document describes ADM, the FORTRAN program which originally creates and then maintains the database from which catalog listings are made. This is a users manual and a programmers manual. ADM is described in sufficient detail to permit its adoption to other high level computer languages and/or other large, scientifically oriented computers. The second volume describes LIBLIST, the FORTRAN program which produces printed catalogs from databases maintained by its companion program ADM.


The application description is directed to those desiring to acquaint themselves with the characteristics of GIPSY. It provides guidelines and aids for the preparation and use of the system, and covers such areas as 1) information collection and creation; 2) information retrieval; 3) maintenance facilities and 4) utility programs. The appendices include sample input documents, form descriptions, keypunched input, error records, printed documents and questions.
The first part of this article reviews attempts to use UDC as a computerized indexing language. The second part describes INDECLASS, a method of using UDC numbers in conjunction with the terms and roles of a coordinate indexing system.

NEPHIS is a computer-assisted indexing system developed by the author at the Western Ontario University School of Library and Information Science. It uses a program which transforms a file of input strings into a file of output strings. The latter is sorted and printed out to form a printed index (a sample page is shown). The system requires minimal typing-in. Experiments which explore the possible teaching of NEPHIS are described. The results showed a need for the analysis of mistakes and new instructions on on-line editing of input strings.


**********************

A library specialist and a computer specialist were asked to collaborate to solve the problem of a growing and diversified in-house, technical library in a diversified agribusiness organization. A computer program was written to: 1) generate alphabetical catalogs of collections; 2) index documents by index terms; 3) index documents by uniterms. Books, periodicals, and documents can be rapidly indexed, updated, added, or deleted; and printouts of the latest holdings can be obtained instantaneously.

102. Dykstra, Mary.
Access to film information: an indexing and retrieval system for the National Film Board of Canada. Halifax, NS, Dalhousie University, School of Library Service, 1977. 73 leaves. Occasional paper 15 x0318-7408.
103. Gardiner, George L.  
Computer assisted indexing in the Central State University library. Champaign, IL, University of Illinois Graduate School of Library Science, 1975. Occasional papers no.120 x0073-5310. 24p.


105. Mischo, William H.  
"An augmented keyword index to abstracting and indexing services". Journal of academic librarianship 3,3 (July 1977), pages 141-145.  
It has become increasingly difficult for the librarian/information specialist, and impossible for the researcher, to keep abreast of available abstracting and indexing tools with resulting retrieval problems as cross-disciplinary studies increase. Discusses the Iowa State University Library Reference Department's computer-organized augmented keyword index to its A I holdings. The guide is comprised of title words and added descriptors, and directs users to designated tools in specific fields, as well as other complementary indexes of which they may be unaware. The listing is arranged in two sections: an alphabetical title listing and a subject arrangement. Details are given of coverage, indexing method and cost.


107. Perica, Esther.  

KWIC, KWAC, KWOC  
See also 94.

108. Fischer, M.  

109. Lewis, R.F.  
"KWIC...is it quick?" Medical Library Association: Bulletin 52,1 (1964), pages 142-147.

110. Luhn, H.P.  
111. Youden, W.W.

112. Lay, Michael W. and Anthony E. Petrarca.

A new approach to creating the double KWIC coordinate index consists of extracting potential main terms directly from the titles (or title-like phrases) instead of from a KWIC index of the titles, and sorting and temporarily retaining the potential index entries in a KWOC type format until other conditions are examined. After all of the titles have been processed and the actual main terms have been selected, if the number of titles containing a particular main term exceeds an arbitrarily assigned threshold value, conventional double KWIC (permuted) subordinate entries are created. If the threshold value is not exceeded, KWOC-type (non-permuted) subordinate entries are created. Other refinements which improve the overall quality of the index have also been introduced. In addition, the new approach facilitates gathering of statistics which will help to reduce the manual and intellectual effort previously required for selection of main terms. The system design for creating the modified double KWIC coordinate index is discussed together with some operating statistics for producing such an index. The use of statistical relationships to create cross references between certain types of terms is also explored.

113. Conaway, Charles W.
The KWAC index generation program was implemented at the State University of New York at Buffalo. It consists of only 252 statements for the COBOL compiler edition V310222 on the CDC6400 computer under the KRONOS operating system in the batch mode. The KWAC program is essentially a KWIC index generator designed for a special purpose for use in a particular course, though it has sufficient flexibility to be used in other similar contexts. The program takes free form natural language input, and generates an index in alphabetical order of each significant word appearing in the input alongside which appears the bibliographical description of the document in which the word was located. Output options permit free form title page, introduction, and epilog. Page headers and footers are permitted, and index page numbers are assigned and printed automatically. Determination of significant words is made in two ways: 1) a word is determined to be significant by default, if it does not appear in a user-input stop list and provided it begins with an alphabetic character and it is more than one character long; and 2) an otherwise significant word may be eliminated from the indexing by a simple procedure at the time of input.
114. Lane, B.B.
   "Keywords in-and out-of-context". Special libraries 55,1 (1964), pages 45-46.

THESAURI

115. Aa, H.J. van der.


   Guidelines for thesaurus structure, construction, and use.
   NY, The Institute, 1974.

118. Angell, R.S.

   "Correlative indexes III: semantic relations among semantemes -- the technical thesaurus". American documentation 8,3 (1957), pages 211-220.

120. Bernier, C.L.

121. Blagden, J.F.
   "Thesaurus compilation methods: a literature review".
   Aslib proceedings 20,8 (1968), pages 345-359.

   Use of descriptors and thesauri. 12 Aug 1971. 60 pages.
   Available from ERIC, ED-059 727.

An attempt is made to provide a brief outline of the present situation with regard to the language used for computer retrieval. This language is increasingly based on descriptors and thesauri, which contain the keywords for manual, mechanical or electronic data retrieval. After a review of the general problem, the document deals with techniques of preparing standardized lists of descriptors, considering them separately from thesauri which represent a higher and more complex level of elaboration which is not always required. Consideration is then given to the classification and construction of descriptors and thesauri, and it is also recommended that this vital aspect of data retrieval should be standardized. Lastly, an outline of some systems of descriptors and thesauri used in the international field is presented, ending with an explanation of the project which, as a subsystem, is being implemented by CLADES.
123. Greene, Nicholas M.
   *Key words and cross references for indexing and standardization of nomenclature in anaesthesiology.* Philadelphia, PA, Lippincott, 1976. 40p.

124. Heald, J.H.

125. Kim, C.

126. Lancaster, Frederick Wilfred.

127. Marshall, Joan K.

128. Soergel, D.

129. Soergel, D.

130. Spang-Hanssen, Henning.
   *Roles and links compared with grammatical relations in natural languages.* Lyngby, Dansk Teknisk Litteraturselskab, 1976. 33 leaves.