

LITERATURE CITED

- (1) Davis, Charles, H., "An Approach to Automated Vocabulary Control in Indexes of Organic Compounds," *J. CHEM. DOC.* **7**, 131-34 (1967).
- (2) Simmons, Robert F., "Automated Language Processing," *Ann. Rev. Inform. Sci. Technol.* Vol. 1, p. 137-169, Wiley, 1966.
- (3) Bobrow, D. G., J. B. Fraser and M. R. Quillian, "Automated Language Processing," *Ann. Rev. Inform. Sci. Technol.* Vol. 2, p. 161-86, Wiley, 1967.
- (4) Salton, Gerard, "Automated Language Processing," *Ann. Rev. Inform. Sci. Technol.* Vol. 3, p. 169-99, Encyclopaedia Britannica, 1968.
- (5) Garfield, Eugene, "An Algorithm for Translating Chemical Names to Molecular Formulas," *J. CHEM. DOC.* **2**, 177-79 (1962).
- (6) Vander Stouw, G. G., I. Naznitsky, and J. E. Rush, "Procedures for Converting Systematic Names of Organic Compounds into Atom-Bond Connection Tables," *J. CHEM. DOC.* **7**, 165-69 (1967).
- (7) Chemical Abstracts Service, "The Naming and Indexing of Chemical Compounds," p. 27N, ACS, 1962.
- (8) COMIT II: "Computer Programming for Primarily Non-Numerical Tasks," p. 101, Winter 1966. (Unpublished, but available from Victor Yngve *et al.*, currently of the Linguistics Department and the Graduate Library School, University of Chicago.)
- (9) Davis, Charles H., "The Binary Search Algorithm," *Am. Doc.* **20**, 167 (1969).
- (10) Tate, Fred, A., "Handling Chemical Compounds," *Ann. Rev. Inform. Sci. Technol.* Vol. 2, p. 292, Wiley, 1967.

The Double-KWIC Coordinate Index

A New Approach for Preparation of High-Quality Printed Indexes by Automatic Indexing Techniques*

ANTHONY E. PETRARCA and W. MICHAEL LAY
Department of Computer and Information Science, The
Ohio State University, Columbus, Ohio 43210

Received June 26, 1969

An indexing scheme is described which lends itself to production of high-quality printed indexes by relatively simple automated techniques. The scheme, an extension of the key-word-in-context (KWIC) indexing concept, leads to the generation of a new type of index called the "Double-KWIC Coordinate Index." This new index has many of the qualities of an articulated subject index; however, it can be produced much more simply and less expensively by automated techniques, because it does not require the depth of syntactic analysis usually needed to produce articulated subject indexes. A prototype double-KWIC coordinate index has been prepared for Volume 7 of the JOURNAL OF CHEMICAL DOCUMENTATION. The computer program for producing this index was written in PL/1. Some of the advantages and disadvantages of the Double-KWIC coordinate index are discussed, together with some areas of immediate useful application.

The need for high-quality printed indexes to facilitate manual retrieval of information has not diminished, despite the strides that have been made in the development of automatic information retrieval systems. Nevertheless, attempts to produce high-quality indexes by automated techniques have only recently begun to merit serious attention. Perhaps the most significant breakthrough in this area occurred when Luhn and others^{1,2} successfully applied the key-word-in-context (KWIC) indexing concept as an automated indexing technique. The widespread use of KWIC indexes since that time and the variety of formats in which they have appeared have been reviewed by Fischer³ and others.⁴

The rapid rise in popularity of KWIC indexes apparently has been due to the high speed and low cost of producing them. However, as noted by Fischer,³ there has been some dissatisfaction with the quality of KWIC indexes. Most of the attempts to improve quality have dealt with variations in format to improve readability or with enrichment of titles to provide additional index

entries which otherwise would not have been derived from words in the titles.

The enrichment of titles improves the quality of KWIC indexes by increasing the breadth of indexing. An equally attractive possibility, which appears to have been little explored, involves extension of the KWIC indexing principle to provide for an increased depth of indexing. If a greater depth of indexing were possible, it would help to overcome one of the major drawbacks of KWIC indexing, viz., a large number of index entries under a given keyword.

One of the difficulties encountered in such a situation is illustrated by the set of KWIC index entries shown in Figure 1, which are taken from a KWIC index derived from titles appearing in Volume 7 of the JOURNAL OF CHEMICAL DOCUMENTATION. Because these index entries are subordered on the basis of words immediately following the word in the index column, the resulting order differs markedly from the usual order one would find in a back-of-the-book index or an articulated subject index. For example, several of the entries indexed under "INFORMATION" indicate that the titles deal with

* Presented in part before the Division of Chemical Literature, 157th Meeting, ACS, Minneapolis, Minn., April 1969.

ERIFICATION OF STRUCTURAL INFORMATION	= +SYSTEM, I. STORAGE AND V	437
G TO THE COMMUNICATION OF INFORMATION	= -AFHY OF RESEARCH RELATIN	823
L KEYBOARDING CHEMICAL INFORMATION		2
LARGE COMMUNITY SELECTED INFORMATION		242
3 BOOK REVIEW: TECHNICAL INFORMATION	= ANNOUNCEMENT SYSTEMS FOR A	126
+VAL SYSTEM AND AUTOMATIC INFORMATION	CENTER ADMINISTRATION, VOL	1257
+INSTRATION OF TECHNICAL INFORMATION	DISTRIBUTION USING COMPUTE+	824
UCATION = BOOK REVIEW: INFORMATION	GROUPS, INTRODUCTORY REMAR+	822
N BUILDING AN OPERATIONAL INFORMATION	MANAGEMENT IN ENGINEERING E	107
ITIC IN+ THE B.F. GOODRICH INFORMATION	PROGRAM = FACTORS I	126
SYSTEM FOR I+ BIOMEDICAL INFORMATION	RETRIEVAL SYSTEM AND AUTOMA	107
REVIEW: ANNUAL REVIEW OF INFORMATION	RETRIEVAL: A COMPUTER-BASED	83-2
MC TRAINING PROGRAMS FOR INFORMATION	SCIENCE AND TECHNOLOGY = +K	83-2
NG EDUCATION IN TECHNICAL INFORMATION	SCIENTISTS = +IES AND ACADE	118
INTEGRATION OF TECHNICAL INFORMATION	SERVICES = CONTINUUM	115
M.+ A CHEMICALLY ORIENTED INFORMATION	SERVICES = COORDINATION AND	111
EDITORIAL: A NATIONAL INFORMATION	STORAGE AND RETRIEVAL SYSTE	433
IN A LARGE-SCALE CHEMICAL INFORMATION	SYSTEM =	5
DETERMINING COSTS OF INFORMATION	SYSTEM = +UNIQUE NOTATION	192
SCIENTIFIC AND TECHNICAL INFORMATION	SYSTEMS =	101
	SYSTEMS IN CURRENT USE = +L	83-2

Figure 1. Portion of a conventional KWIC index, illustrating entries for a high-density keyword

INFORMATION	
RETRIEVAL SYSTEM. I. STORAGE AND VERIFICATION OF STRUCTURAL *	43
A CHEMICALLY ORIENTED * STORAGE AND RETRIEVAL SYSTEM. I. STORAG	43
BIOMEDICAL * RETRIEVAL: A COMPUTER-BASED SYSTEM FOR INDIVIDUAL	98
DETERMINING COSTS OF * SYSTEMS =	101
FACTORS IN BUILDING AN OPERATIONAL * PROGRAM =	107
SYMPOSIUM ON ADMINISTRATION OF TECHNICAL * GROUPS. INTRODUCTORY	110
COORDINATION AND INTEGRATION OF TECHNICAL * SERVICES =	111
CONTINUING EDUCATION IN TECHNICAL * SERVICES =	115
SALARIES AND ACADEMIC TRAINING PROGRAMS FOR * SCIENTISTS =	118
THE B.F. GOODRICH * RETRIEVAL SYSTEM AND AUTOMATIC INFORMATION	124
THE B.F. GOODRICH INFORMATION RETRIEVAL SYSTEM AND AUTOMATIC *	124
SELFTIVE * ARRANGEMENT SYSTEMS FOR A LARGE COMMUNITY OF USERS	142
USE OF NONUNIQUE NOTATION IN A LARGE-SCALE CHEMICAL * SYSTEM	192
KEYBOARDING CHEMICAL *	232
BOOK REVIEW: * MANAGEMENT IN ENGINEERING EDUCATION =	B2-2
BOOK REVIEW: ANNUAL REVIEW OF * SCIENCE AND TECHNOLOGY =	B3-2
CONVENTIONAL SCIENTIFIC AND TECHNICAL * SYSTEMS IN CURRENT USE =	B3-2
W BIBLIOGRAPHY OF RESEARCH RELATING TO THE COMMUNICATION OF * =	B2-5
BOOK REVIEW: TECHNICAL * CENTER ADMINISTRATION, VOL 3 =	B2-5
EDITORIAL: A NATIONAL * SYSTEM =	E 6

Figure 2. Variant form of a KWIC index for the same entries illustrated in Figure 1

"TECHNICAL INFORMATION," but the entries are scattered because of the ordering principle just described.

In another format for the KWIC index (Figure 2), the situation is even worse. In this format, the index word is extracted from the title and replaced by an asterisk to indicate its location in the title. All of the titles, or portions thereof, from which a given index term is extracted are then grouped together under that index term and are subordered on the basis of the accession numbers for the titles from which they are derived. This method of ordering is worse than the first, because of complete randomization of the words to the right as well as to the left of the index words. Also, this second format makes it more difficult to determine the immediate context about the keyword when scanning the individual entries, since the keyword—in this case, its identifying asterisk—no longer appears in a fixed position.

To overcome some of the difficulties of the KWIC indexing approach, studies have been initiated by Lynch,⁵ Dolby,⁶ and others to analyze the characteristics of traditional subject indexes. These approaches tend to require linguistic analysis of titles and title-like phrases to effect the transformations required to produce such higher-quality indexes by automated techniques. This paper reports on a more simplified approach to automatic preparation of higher-quality indexes, based on an extension of the KWIC indexing technique concept. For reasons which will soon become apparent, we have chosen the name "Double-KWIC Coordinate Index" for the printed index produced by this new approach.

CONSTRUCTION OF THE DOUBLE-KWIC COORDINATE INDEX

As illustrated in Figure 3, the double-KWIC coordinate index is constructed as follows:

1. The first significant word in a title is extracted as a main index term and replaced by an asterisk (*) to indicate its position in the title.
2. The remaining words in the title are then rotated, so as to permit each significant word to appear as the first word of a wrap-around subordinate entry under the main index term.

Steps 1 and 2 are repeated until all of the titles of a given bibliographic listing are processed. The index entries so created are then sorted alphabetically, both with regard to main terms (primary sort) and subordinate terms (secondary sort). Word significance for selection of main index terms and subordinate index terms is established on the basis of stop lists, discussed later. Also, main index terms are not restricted to single words, but may consist of multi-word terms derived from contiguous sets of words in the titles.

The computer programs for carrying out all of these operations have been written in PL/1. They have been

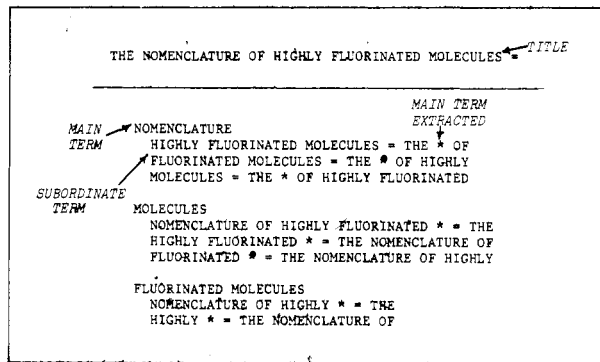


Figure 3. Construction of double-KWIC coordinate index entries

1 BOOK REVIEW

2

3 ADMINISTRATION, VOL 3 = *: TECHNICAL INFORMATION CENTER D257

4 ANALYSIS = *: CONTINUING NUMERICAL DATA PROJECTS A SURVEY AND M2-2

5 ANALYSIS, VOL 4 = *: ENCYCLOPEDIA OF INDUSTRIAL CHEMICAL D258

ANALYSIS, VOLS 2 AND 3 = +ENCYCLOPEDIA OF INDUSTRIAL CHEMICAL B 59

ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY = *: B3-2

APPLICATIONS = *: COMPUTER D258

BASIC PRINCIPLES OF CHEMISTRY = *: B3-2

BIBLIOGRAPHY REVIEW = *: SALICYLATES, A CRITICAL B 59

BIBLIOGRAPHY OF RESEARCH RELATING TO THE COMMUNICATION OF INFO B337

BIOCHEMICAL PREPARATIONS = *: B3-2

BOOK * *: CHEMICAL DATA B3-2

BOOK OF CHEMISTRY = *: REFERENCE B2-2

CAS TODAY = *: B182

CATALOG ENTRIES = +COMPUTER FILING OF INDEX, BIBLIOGRAPHIC, AND D2-2

CENTER ADMINISTRATION, VOL 3 = *: TECHNICAL INFORMATION D257

CHEMICAL ANALYSIS, VOL 4 = *: ENCYCLOPEDIA OF INDUSTRIAL B258

CHEMICAL ANALYSIS, VOLS 2 AND 3 = +ENCYCLOPEDIA OF INDUSTRIAL B 59

6

- 1 - Main index term.
- 2 - Location of main index term in title being permuted (rotated) for creation of subordinate entries.
- 3 - Subordinate index term.
- 4 - Word in wrap-around title which immediately precedes subordinate index term.
- 5 - Truncation symbol used when words in wrap-around title do not fit in allotted field.
- 6 - Page number. Alphabetic characters preceding the page number represent the following: B - book review; E - editorial. Also, the two page-numbering systems used by the Journal are represented by the following formats: (1) Unhyphenated - arabic numbered pages used for sequential numbering of the pages for Volume 7; (2) Hyphenated - Roman numeral pages for the individual issues of Volume 7. The number preceding the hyphen is the issue number.

Figure 4. Annotated description of a prototype double-KWIC coordinate index derived from titles in *J. Chem. Doc.*, Volume 7

JOURNAL OF CHEMICAL DOCUMENTATION, VOL. 9, NO. 4, NOVEMBER 1969

DOUBLE-KWIC COORDINATE INDEX

compiled and executed on an IBM 360/75 computer at the Ohio State University Computer Center. An annotated description of a prototype index prepared for Volume 7 of the JOURNAL OF CHEMICAL DOCUMENTATION is illustrated in Figure 4.

NOTE ADDED IN PROOF: At the suggestion of the editor of the JOURNAL OF CHEMICAL DOCUMENTATION, only a portion of the complete prototype index is included with this article (Figure 5). Instead, the authors have agreed to prepare the subject index to this volume of the journal, using a modified form of the double-KWIC coordinate index which has been described in part in *ASIS Proceedings* 6, 277 (1969) and in the foreword to "A Modified Double-KWIC Coordinate Index for *ASIS Proceedings*, Volume 6 (1969)" which was distributed at the 32nd Annual Meeting of the American Society for Information Science (ASIS), San Francisco, October 1-4, 1969. Copies of the complete prototype index and the last named document have been deposited in the National Auxiliary Publication Service (NAPS) of the ASIS. Accession numbers and procurement information are listed at the end of this article.

UTILITY OF THE DOUBLE-KWIC COORDINATE INDEX

To illustrate some of the advantages of the double-KWIC coordinate index, let us compare some segments of the conventional KWIC index and a double-KWIC coordinate index prepared from the titles contained in Volume 7 of the JOURNAL OF CHEMICAL DOCUMENTATION. Figure 6 illustrates that portion of each index for the entries under "NOMENCLATURE." In the KWIC index, random ordering of subordinate concepts would require examination of the entire context for each entry if one were interested in a specific type of nomenclature. The double-KWIC coordinate index eliminates this problem, since alphabetic ordering of subordinate concepts enables one quickly to scan the subordinate index terms to find the particular subordinate concept—i.e., type of nomenclature—in question.

Both indexes would permit one to locate equally as well those precoordinated index terms under the heading for the modifier immediately preceding the word nomenclature—e.g., "BIOCHEMICAL NOMENCLATURE." However, as illustrated in Figure 7, the double-KWIC coordinate index would enable one quickly to coordinate such precoordinated terms with other subordinate concepts, as well.

There is a practical upper limit of three or four to the length of multi-word terms that may be extracted as main terms for the double-KWIC coordinate index. Figure 8 illustrates how some useful three-word main terms describing concepts which were scattered under the index term "INFORMATION" in Figures 1 and 2 can be coordinated quickly with other concepts, as well.

The use of enrichment terms to enhance the quality of KWIC indexes applies even more so to double-KWIC coordinate indexes. Two enrichment terms were used for the prototype index prepared for Volume 7 of the JOURNAL OF CHEMICAL DOCUMENTATION—one for book reviews and one for editorials, primarily to differentiate these from the articles published in the journal. Figure 5 illustrates the subordinate entries under the main term "BOOK REVIEW." Note how the subordinate entries enable one immediately to locate entries for those books whose titles contain keywords of particular interest. Furthermore, as

KWIC INDEX	
CARBOHYDRATE NOMENCLATURE =	78
SOME PROBLEMS IN POLYMER NOMENCLATURE =	74
FUNCTIONING OF BIOCHEMICAL NOMENCLATURE =	72
INORGANIC NOMENCLATURE IN 1966: PROGRESS AND PROBLEMS =	67
THE NOMENCLATURE OF HIGHLY FLUORINATED MOLECULES =	82
THE NOMENCLATURE OF ORGANIC CHEMISTRY =	64

DOUBLE KWIC COORDINATE INDEX	
NOMENCLATURE	
BIOCHEMICAL * =	THE ORGANIZATION AND FUNCTIONING OF 72
CARBOHYDRATE * =	78
CHEMISTRY * =	THE * OF ORGANIC 64
FLUORINATED MOLECULES * =	THE * OF HIGHLY 82
FUNCTIONING OF BIOCHEMICAL * =	THE ORGANIZATION AND 72
HIGHLY FLUORINATED MOLECULES * =	THE * OF 82
INORGANIC * IN 1966: PROGRESS AND PROBLEMS =	67
MOLECULES * =	THE * OF HIGHLY FLUORINATED 82
ORGANIC CHEMISTRY * =	THE * OF 64
ORGANIZATION AND FUNCTIONING OF BIOCHEMICAL * =	THE 72
POLYMER * =	SOME PROBLEMS IN 74
PROBLEMS =	INORGANIC * IN 1966: PROGRESS AND 67
PROBLEMS IN POLYMER * =	SOME 74
PROGRESS AND PROBLEMS =	INORGANIC * IN 1966: 67

Figure 6. Comparison of conventional KWIC entries and double-KWIC coordinate index entries for a given index term derived from the same set of titles

BIOCHEMICAL NOMENCLATURE	
FUNCTIONING OF * =	THE ORGANIZATION AND 72
ORGANIZATION AND FUNCTIONING OF * =	THE 72

INORGANIC NOMENCLATURE	
PROBLEMS =	* IN 1966: PROGRESS AND 67
PROGRESS AND PROBLEMS =	* IN 1966: 67

POLYMER NOMENCLATURE	
PROBLEMS IN * =	SOME 74

Figure 7. Illustration of increased depth of indexing provided by the extraction of two-word terms

TECHNICAL INFORMATION CENTER	
ADMINISTRATION, VOL 3 =	BOOK REVIEW: * B257
BOOK REVIEW: * ADMINISTRATION, VOL 3 =	B257

TECHNICAL INFORMATION GROUPS	
ADMINISTRATION OF * INTRODUCTORY REMARKS =	SYMPOSIUM ON 110
INTRODUCTORY REMARKS =	SYMPOSIUM ON ADMINISTRATION OF * 110
REMARKS =	SYMPOSIUM ON ADMINISTRATION OF * INTRODUCTORY 110
SYMPOSIUM ON ADMINISTRATION OF * INTRODUCTORY REMARKS =	110

TECHNICAL INFORMATION SERVICES	
CONTINUING EDUCATION IN * =	115
COORDINATION AND INTEGRATION OF * =	111
EDUCATION IN * =	CONTINUING 115
INTEGRATION OF * =	COORDINATION AND 111

Figure 8. Some three-word main terms of the prototype double-KWIC coordinate index

BOOK REVIEW	
ADMINISTRATION, VOL 3 =	*: TECHNICAL INFORMATION CENTER B257
ANALYSIS =	*: CONTINUING NUMERICAL DATA PROJECTS A SURVEY AND B2-2
ANALYSIS, VOL 4 =	*: ENCYCLOPEDIA OF INDUSTRIAL CHEMICAL B258
ANALYSIS, VOLS 2 AND 3 =	+ENCYCLOPEDIA OF INDUSTRIAL CHEMICAL B 59
ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY =	*: B3-2
APPLICATIONS =	*: COMPUTER B258
BASIC PRINCIPLES OF CHEMISTRY =	*: B3-2
BIBLIOGRAPHIC REVIEW =	*: SALICYLATES. A CRITICAL B 59
BIBLIOGRAPHY OF RESEARCH RELATING TO THE COMMUNICATION OF INFORMATION =	B257
BIOCHEMICAL PREPARATIONS =	*: B3-2
BOOK =	*: CHEMICAL DATA B3-2
BOOK OF CHEMISTRY =	*: REFERENCE B2-2
CAS TODAY =	*: B182
CATALOG ENTRIES =	+COMPUTER FILING OF INDEX, BIBLIOGRAPHIC, AND B2-2
CENTER ADMINISTRATION, VOL 3 =	*: TECHNICAL INFORMATION B257
CHEMICAL ANALYSIS, VOL 4 =	*: ENCYCLOPEDIA OF INDUSTRIAL B258
CHEMICAL ANALYSIS, VOLS 2 AND 3 =	+ENCYCLOPEDIA OF INDUSTRIAL B259

Figure 9. Illustration of how the word "REVIEW" maintains its separate identity as an index term when it appears independently of "BOOK REVIEW"

illustrated in Figures 8 and 9, access can be gained through the keywords of the book titles themselves—e.g., "TECHNICAL INFORMATION CENTER" in Figure 8.

A special technique involves the use of an underscore in the enrichment term "BOOK__REVIEW," which allows it to be treated as a single word by the algorithm for constructing the index. This technique was employed so that the words "BOOK" and "REVIEW" would maintain separate identities, to permit them to be used as meaningful index terms (Figure 9) when they appeared independently of the enrichment term.

STOP LISTS FOR THE DOUBLE-KWIC COORDINATE INDEX

Three stop lists were used to preclude the appearance of nonsignificant main terms and subordinate terms in the prototype double-KWIC coordinate index.

Stop list A consists of low index-value words which should never appear as the first word of a main index term, but which might appear in other positions of a main term. Included on this list are such words as *activities*, *announcement*, *applications*, *approach*, *assisted*, etc.; all prepositions, articles, and conjunctions; and all character strings less than three.

Stop list B consists of words which should never appear as subordinate index terms or as the final word of a multiple-word main index term. Included on this list are all prepositions, articles, and conjunctions; all character strings less than three; and a few words of extraordinarily low index value, such as *some*, *such*, etc.

Stop lists A and B are invoked by the algorithms which generated the main term and subordinate term entries. Consequently, these stop lists actually prevented generation of index entries containing the stop-list words in the positions indicated above.

Stop list C, on the other hand, was invoked just prior to the output formatting stage. Its function is to eliminate redundancy caused by generation of single-word and multi-word main terms which started with a common word. For example, the main terms AMERICAN and AMERICAN CHEMICAL were eliminated in favor of the more specific term AMERICAN CHEMICAL SOCIETY, since there was complete overlap in the titles from which they were derived. In other instances, the less specific term may have been retained if there was incomplete overlap.

ADVANTAGES AND DISADVANTAGES OF THE DOUBLE-KWIC COORDINATE INDEX

Some of the advantages of the double-KWIC coordinate index as compared to the conventional KWIC index have already been cited. Briefly, they may be summarized as follows: (1) The double-KWIC coordinate index provides a greater depth of indexing. (2) Coordinate searches can be performed more easily on the double-KWIC coordinate index, both because of its format and because of the alphabetic ordering of the subordinate terms under the main index terms. False coordinations are unlikely, because contextual relationships between the main terms and the subordinate terms are preserved in each index entry. (3) Class relationships can be expressed by the

use of enrichment terms. When these enrichment terms appear as main headings, the members of the class are differentiated on the basis of the subordinate index terms. Specific members of a class can also be accessed through main headings describing the specific members of the class. (4) The format of the double-KWIC coordinate index is more readable, because it closely resembles the format of a conventional subject index.

The major disadvantages of the double-KWIC coordinate index over the conventional KWIC index are the increased size and the higher costs of production. But cost-return benefits could well justify the use of double-KWIC coordinate indexes in place of conventional KWIC indexes in many instances. The prototype double-KWIC coordinate index contained 1630 entries derived from 91 titles, whereas the corresponding KWIC index contained only 388 entries—a ratio of approximately 4 to 1. Computing time for the prototype double-KWIC coordinate index was about twice that for the conventional KWIC index. With proper layout and 50% photoreduction (Figure 5), the prototype index occupies seven and one-half standard-size pages.

The real value of the double-KWIC coordinate index can be appreciated when it is compared with an articulated subject index automatically generated from a given set of titles or title-like phrases.⁷ The double-KWIC coordinate index approaches the quality of such an index, but more likely at a considerably reduced cost.

AREAS OF IMMEDIATE APPLICATION

Among the immediately useful areas of application, the double-KWIC coordinate indexing technique could be used for the preparation of volume indexes to journals, as typified by the prototype index (Figure 5) and the subject index to this volume of the journal (see NOTE ADDED IN PROOF). The availability of input from *Chemical Titles* for most chemical journals makes this idea unusually attractive. The double-KWIC coordinate index could also be used in place of conventional KWIC indexes—e.g., *Chemical Titles*, *Computing Reviews*, etc.—at least on a trial basis to assess the cost-benefit-performance relationships. The approximately four-fold increase in the number of index entries created for the double-KWIC coordinate index would only cause slightly more than a two-fold increase in the over-all size of any conventional KWIC index, since approximately half of the index contains the bibliography and an author index, neither of which would be altered in space requirements.

FUTURE CONSIDERATIONS

Among the ideas to be explored in future research are:

1. Statistical selection of multi-word main index terms.
2. Possible reduction in the over-all size of the index by automatic creation of "see" and "see also" references.
3. Use of automatic text-editing systems to eliminate or reduce the need for the various stop lists.
4. Elimination of scattering caused by main terms consisting of variant word forms—e.g., singular and plural forms of a main term.
5. Generation of double-KWIC indexes from abstracts which are available in computer-readable format.

CONCLUSIONS

The double-KWIC coordinate indexing technique appears to be a significant contribution to the field of automatic indexing. It is easy to envisage many useful applications of this indexing technique, in addition to the ones mentioned above. We hope that its usefulness will be further enhanced by improvements resulting from future research activities which we plan to pursue in this area.

ACKNOWLEDGMENT

Partial support of this work by a grant (G.N.-534) from the National Science Foundation (OSIS) is gratefully acknowledged. The authors also wish to thank S. M. Cohen, Chemical Abstracts Service, for valuable discussions, and the Office of Education for a fellowship to W.M.L.

LITERATURE CITED

- (1) Luhn, H. P., "Keyword-in-Context Index for Technical Literature (KWIC Index)," RC-127, IBM Corp., Yorktown Heights, N. Y., 1959; also, *Am. Doc.* 9, 288 (1960).
- (2) Citron, J., L. Hart, and H. Ohlman, "A Permutation Index to the Preprints of the International Conference on Scientific Information," SP44, System Development Corp., Santa Monica, Calif., 1959.
- (3) Fischer, M., "The KWIC Index Concept: A Retrospective View," *Am. Doc.* 17, 57 (1966).
- (4) Sharp, J. R., "Content Analysis, Specification, and Control," *Ann. Rev. Inform. Sci. Technol.* 2, 87 (1968).
- (5) Armitage, J. E., and M. F. Lynch, "Articulation in the Generation of Subject Indexes by Computer," *J. CHEM. DOC.* 7, 170 (1967).
- (6) Dolby, J. L., "The Structure of Indexing: The Distribution of Structure-Word-Free Back-of-the-Book Entries," *ASIS Proc.* 5, 65 (1968).
- (7) Armitage, J. E., and M. F. Lynch, "Evaluation of a Computer Technique for Generation of Subject Indexes," *Abstracts of Papers*, No. F-19, 155th National Meeting, ACS, San Francisco, Calif., April 1968.

NEWS AND NOTES

ACCESS

Compiled by CAS in cooperation with 397 libraries in 28 nations, *ACCESS* provides complete identification of source publications customarily cited in the scientific literature by abbreviated title and indicates which of the participating libraries maintain files of the cited publications. The 1969 edition covers some 21,000 periodicals, monographs, and conference proceedings volumes, including essentially all publications abstracted by *Chemical Abstracts* since 1907, those abstracted by *Chemisches Zentralblatt* since 1830, and those cited by *Beilstein's Handbuch der organischen Chemie* prior to 1907.

For each of these publications, *ACCESS* lists the full title and its standard abbreviation, the ASTM CODEN for the title, the language of publication, publication history, price, publisher, and the volumes of the publication available in each of the participating libraries. The 1969 edition also includes a directory of the participating libraries, indicating the lending or photocopying services each provides, a listing of the libraries' holdings of patent documents, and a directory of publishers and sales agents.

Information in the 1969 edition will be kept up to date through quarterly supplements, which will be offered on an annual subscription basis beginning in December 1969. Data from both the first edition and the supplements will also be available in computer-readable form.

ACCESS supersedes and replaces the *Chemical Abstracts Lists of Periodicals*, published by CAS since 1908.

The 1500-page 1969 edition of *ACCESS* is priced at \$100. An annual subscription to the quarterly supplements is \$75.

Conference on Image Storage and Transmission for Libraries

A Conference on problems and prospects for image storage and transmission systems for library applications will be held at the National Bureau of Standards, Gaithersburg, Md., on December 1-3, 1969. Cosponsors of the Conference will be the Federal Library Committee's Task Force on Automation, the Lister Hill National Center for Biomedical Communication, the Panel on Information Sciences Technology of COSATI, and the National Bureau of Standards.

The program will examine the development and use of graphic and textual image transmission systems, as well as the state of the advancing technology in equipment and techniques for such systems. Invited speakers will present underlying principles of the subject along with their appraisal of the more important aspects of applications. The technical sessions will include such topics as:

Survey of Transmission Systems
Operational Practice and Constraints
Microimage Storage and Transmission
Advanced Techniques in Image Transmission
Legal and Economic Problems

A copy of the Proceedings, in microfiche form, will be mailed to each registrant following the Conference.

For further information, contact Madeline M. Henderson, Center for Computer Sciences and Technology, National Bureau of Standards, Room B226-Instr., Washington, D. C. 20234.