Hans Peter Luhn and Herbert M. Ohlman: Their Roles in the Origins of Keyword-in-Context/Permutation Automatic Indexing

Robert V. Williams
School of Library and Information Science, University of South Carolina, Columbia, SC 29208. E-mail: bobwill@sc.edu

The invention of automatic indexing using a keyword-in-context approach has generally been attributed solely to Hans Peter Luhn of IBM. This article shows that credit for this invention belongs equally to Luhn and Herbert Ohlman of the System Development Corporation. It also traces the origins of title derivative automatic indexing, its development and implementation, and current status.

Introduction

In November 1958, at the International Conference on Scientific Information (ICSI) in Washington, DC, Hans Peter Luhn, of the International Business Machines (IBM) Company, and Herbert M. Ohlman, of the System Development Corporation (SDC), simultaneously and independently presented their new systems for automatic indexing. Luhn called his system "Keyword-in-Context (KWIC) indexing" and Ohlman called his "permutation indexing." Within a few years, despite the similarities and obvious lack of advantages of one system over the other, Luhn's KWIC became the de facto standard for title derivative automatic indexing, and Ohlman's permutation indexing system largely disappeared. A 1965 state-of-the art report on automatic indexing (Stevens, 1965), with an extensive discussion of KWIC and related concepts, did not discuss Ohlman's work, though it is cited in the bibliography. A 1966 retrospective review of the KWIC concept gave a slight mention of Ohlman's work on permuted indexes, but uses the phrase "… when Luhn invented the KWIC index…." (Fischer, 1966, p. 58). By 1995, Wellisch's (1996) handbook on indexing had no mention of Ohlman and stated "Luhn, who first published the successful application of his idea in 1958, called his method KWIC … and it became, indeed, the first and to this day the only fully automatic indexing method" (pp. 258–259). Lancaster's discussion of KWIC and related techniques in his 1998 and 2003 overviews of indexing theory and practice referred only to Luhn (Lancaster, 1998, pp. 48–50; Lancaster, 2003, pp. 54–56). Chu (2007), likely echoing Wellisch, also called Luhn's KWIC "… the only fully automatic indexing method." (p. 9). The most significant recent mention of Ohlman's work was Eugene Garfield's (1976) acknowledgment of the influence of his ideas as he and Irving Sher developed the *Permuterm Subject Index* used in the *Science Citation Index*. Today, very few people in information studies have any memory of the permutation indexing system and Herbert Ohlman's name is seldom recognized while KWIC is almost instantly recognized and Luhn's name is well-known.

This essay explores how and why this happened, the origins and development of automatic indexing, with particular attention to the history and development of the systems of Luhn and Ohlman, the closely associated systems of automatic abstracting and selective dissemination of information (SDI), and the early years of machine-based information retrieval. It also traces the careers of Luhn and Ohlman to better understand the reasons for the success of KWIC (and related systems) indexing and the disappearance of permutation indexing.

In his masterful 1963 text, *Methods of Information Handling*, Charles P. Bourne noted: "The first published reports to describe the implementation of this technique [KWIC and permutation] with data processing equipment came independently and simultaneously in 1958 from IBM and SDC …" (p. 17). Bourne (1963) discussed the two systems under the heading "Fundamental Methods of Indexing Subject Content" and the subheading "Words Chosen from Title or Text but Common Words Omitted" and noted that it "… is currently in use by many companies and publishers for indexes and title announcement bulletins" (p. 17). Bourne's 1963 description of the use of these systems is accurate for the time, but significantly understates the impact that these systems, particularly KWIC, would have on the field of information retrieval in the next few years.

Received August 13, 2009; revised October 5, 2009; accepted October 5, 2009

 • Published online 10 December 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21265

Background: Indexing, Automatic Indexing, and Information Retrieval as of 1958

Bourne (1963) began his text with a review of the "information problem" that existed about 1955 to 1962: 30,000 to 50,000 scientific and technical journals with 900,000 to 2.1 million articles published each year; 3,500 abstracting and indexing services in the world (with 10,000 items added per week); 3,000 specialized information centers in the United States trying to cope with these problems; and the increasing growth of academic libraries serving the sciences. The purpose of his book was to review a large variety of approaches, systems, and equipment to deal with this information retrieval problem, which he confined to document retrieval. Bourne (1963) first reviewed classification and indexing systems, such as subject headings, descriptors, Uniterms, and others using an organizational approach by "... degree of control exercised on the growth and use of the indexing language" (p. 13). Thus, he came very quickly to a discussion of KWIC/permutation indexing, noting that "Its distinguishing feature is its form of display. Because the index term is shown in its context, its indexing power is slightly greater (but only slightly) than the simplest of indexing schemes that lift words from the text" (Bourne, 1963, p. 16). Despite this less than enthusiastic recommendation of KWIC/permutation, Bourne nevertheless spent five pages discussing the various approaches in current use, their advantages and disadvantages, and variations employed such as enriching title words with keywords from the document. He also discussed KWIC/permutation as complementary to the early "automatic extraction" systems just being developed.

By 1965, Stevens had devoted 50 pages to KWIC and related techniques in her state-of-the-art review of automatic indexing techniques and listed more than 30 separate systems currently in use or near deployment (pp. 51–91). To say that KWIC had caught on is, as noted earlier, an understatement. When *Chemical Titles*, the first computer-produced periodical, and the first large-scale use of KWIC, was published in 1961, Dale Baker, Director of Chemical Abstracts Service, called it "the miracle of the decade" and "the greatest thing to happen in chemistry since the invention of the test tube" (as cited in Stevens, 1965, p. 40). Stevens went on to note that this optimism about KWIC was caused by the speedy production of an entire index to a collection of documents so that dissemination can be rapidly accomplished. *Chemical Titles* was a great success, as was *Biological Abstracts-Subjects-in-Context* (*BASIC*) published by Biological Abstracts, Inc. Other organizations, both in the United States and around the world, soon began their own KWIC publications.

Variations on KWIC were soon to follow, such as keyword out of context (KWOC), keyword alongside context (KWAC), key phrase in context (KPIC), and many others. The major variations, however, were oriented toward enrichment of the (usually) title words in the index: authors names, keywords from the abstract, and word frequency count. Lesser used variations included adding the most commonly mentioned word to the index, relative word frequency (the less mentioned, the more important), and special clues (e.g., capitalization, end or beginning of a sentence, word co-occurrence, semantic and syntactical relationships). Stevens (1965) reviewed more than 50 experimental and quasi-experimental studies of automatic derivative indexing and automatic classification and categorization techniques used from 1958 to 1965 and concluded that "In general, the results of automatic index-term assignment procedures appear to run in the area of 45–75 percent agreement with prior human indexing" (p. 178). (Weinberg, 1981, however, reported a much lower rate of agreement.) Stevens also noted the lack of a solid body of research on indexing effectiveness, whether automatic or assigned, to answer the questions related to how "good retrieval" can be evaluated (p. 151).

A major issue continually addressed by Stevens' (1965) review is whether KWIC-type indexes can actually be called indexes and whether the process of creating one can be called indexing. Stevens hedged only a little in her summary appraisal of automatic indexing:

> "There can be little or no doubt that the results of automatic assignment indexing experiments to date, (if extrapolation from the small and often highly specialized samples so far used in actual tests is in fact warranted) do suggest that an indexing quality generally comparable to that achievable by run-of-the-mill manual operations, at comparable costs and with increased timeliness, can be achieved by machine." (p. 178)

Quoting M.E. Maron, one of the foremost researchers on automatic indexing of this period, Stevens noted that "automatic indexing represents the opening wedge in a general attack at not only the problems of identification search and retrieval, but also the problem of automatically transforming information on the basis of its content." (p. 182)

Maron's statement (as cited in Stevens, 1965) is prescient because it superbly described not only current research but predicted the coming focus of information retrieval work for the next 20 years. Automatic indexing continued to be of great interest to researchers and funders. Stevens' 1965 report cited 662 references to automatic indexing studies, but they represented only the early work on the problem. Following partially in the footsteps of Luhn (1958a) and Phyllis B. Baxendale (1958), statistical analysis of terms, phrases, keywords, and so on would predominate in the years immediately after 1965. Studies using automatic assignment indexing techniques, including using humans to analyze statistical data on word occurrence, would become common, particularly in efforts to automatically compile thesauri. When used in concert with efforts at automatic extraction (a descendant of automatic abstracting), or text summarization, these approaches eventually led to efforts at genuine information retrieval systems, most notably Gerard Salton and McGill's (1983) SMART system. Anderson (1994), discussing issues relating to revising the National Information Standards Organization ANSI Z39.4 standard on criteria for indexers, noted that "... a simple KWIC, with vocabulary management, can be a very effective index..." (p. 633), but added the proviso

DOI: 10.1002/asi

that this is true only if the information on which it is based (usually titles) are good representations of the text indexed. Lancaster (1998) wrote a concise summary of these trends and an overview of work current at the time. His summary showed that even though KWIC-type approaches, as a single technique, had mostly disappeared by the 1990s, both research and practice still utilized some of the basic ideas in them.

Somewhat lost in the excitement over automatic indexing and abstracting was information retrieval work that had begun by the early 1950s on machine-based retrieval using edge-notched cards, punched cards and sorters, microfilm-based optical coincidence systems, and others such as Mooers' Zatocoding system (Mooers, 1951). Most of these systems used assigned indexing terms (descriptors, Uniterms, etc.) and employed Boolean logic for the actual searching and retrieval.

KWIC/Permutation Before KWIC and Permutation

Three years after Luhn and Ohlman publicized their new systems, Mary P. Veilleux, of the U.S. Central Intelligence Agency (CIA), presented a paper at the 1961 Third Institute on Information Storage and Retrieval (Veilleux, 1963). She disclosed that a permuted title word indexing system had been in use for internal reports at the CIA since about 1953. Her presentation detailed a system using IBM punched cards, sorting equipment, authority file, thesaurus, indexer enhancement, and optical character recognition equipment to produce an internal reference aid for the intelligence community (Veilleux, 1963). However, because it was an internal CIA operation, and this was the first public presentation about it, the system was unknown outside the agency. While Veilleux did not discuss the systems at the CIA prior to about 1953, Burke (1994), in his *Information and Secrecy* book, briefly discussed a similar system in use by the Office of Naval Research (ONR), OP-20-G, where cryptanalysis was being done during World War II (p. 64). Using tabulators and punched cards, the analysts were sorting and correlating code words in various reports. Burke (personal communication, May 10, 2005) later noted that ONR was using these techniques to search coded messages for critical identifying information words, names, and events.

Shortly after the introduction of Ohlman's (1958) permutation index at the ICSI, he began to delve into the history of related systems and presented a paper at the 1960 meeting of the American Documentation Institute (ADI), titled "Mechanical Indexing: Historical Development, Techniques, and Critique." In this paper, he traced the origins back to the first card catalog, which he dated as 1775, at the Paris Academy of Sciences, where it was used to intercollate new materials. He traced mechanical indexing and the use of keywords to Andrea Crestadoro's 1856 book, *The Art of Making Catalogues of Libraries* (as cited in Ohlman, 1960), and to even earlier concordances, where the full use of keyword and context were fully displayed. Wellisch (1996, p. 183) traced the first literary concordance, to the Bible, to 1230 and the first machine-based concordance (using IBM punched cards) to the early 1950s. Weinberg (2001) showed that features such as permutation indexing date even earlier, around the 9th or 10th century.

Ohlman followed up his 1960 presentation with a draft chronology, then available from the National Bureau of Standards, of the concept of permutation indexing (Ohlman, 1961?). Reading the 1960 ADI paper in 2009, one has the feeling that Ohlman was trying to assert that Luhn and IBM did not own the idea of KWIC indexing; that it was much older and was a commonly used technique in information retrieval and literary analysis.

Herbert M. Ohlman and his Permutation Index: Origins, Development, and Disappearance

Herbert Marvin Ohlman (Figure 1) was born in New York City in 1927. He earned a B.S. degree in Physics from Syracuse University and an M.S. in Applied Mathematics and Computer Science from Washington University in St. Louis. He took additional graduate work in Physics, Mathematics, Operations Research, and Library Science at several different schools. From 1950 to 1980, he was a staff member at the National Bureau of Standards, Carrier Corporation, Battelle Memorial Institute, System Development Corporation, Lockheed Electronics, IBM Advanced Systems Division, Itek, Xerox, Washington University, and the World Health Organization. After his retirement, he did consulting work in telecommunications, computer-communications systems, educational technology, and information technology. He was an active member of the American Society for Information Science (ASIS) for 30 years and authored more than 40 reports and articles during his career, most on information technology. He died on May 27, 2002, following complications from heart surgery.

In a paper presented at the 1998 Conference on the History and Heritage of Science Information Systems, Ohlman described in some detail how he became interested in what he came to call "permutation" indexing (Ohlman, 1999).

c. 1965

FIG. 1. Herbert Ohlman, about 1965. Used with permission of his family.

DOI: 10.1002/asi

He had joined the System Development Division of the Rand Corporation [which later became separate as System Development Corporation (SDC)] sometime in early 1957 to work on efforts to control the documentation of the very large defense project, Semi-Automated Ground Environment (SAGE). The librarian, Malcolm Ferguson, had developed a Jonkers Termatrex "peek-a-boo" optical coincidence system for indexing the SAGE documents. Noticing that much of the space on these "peek-a-boo" cards was wasted and that the system also isolated terms from their natural language context, Ohlman (1999) thought this would reduce retrieval effectiveness:

> To provide contextual retrieval, I devised a new method based on IBM punched cards and tabulating machines. Significant title words were keypunched, one card to a document title. Tabulating machines were "programmed" by Lewis Hart of SDC . . . to punch duplicate cards. In each duplicate card, title words were cycled to the left; the expanded deck was then sorted, and this final deck run through a printer to produce the index. (p. 188)

Lauren Doyle's 1966 lengthy internal SDC memo provided additional details on how the "peek-a-boo" system came to SDC, noting that Dave Sternlight, the director of the recently formed "documentation section," had attended meetings in Cleveland and Washington that focused on documentation where he learned of this new system. Ohlman then inherited the documentation problem when Sternlight decided he did not want to move to California and left for a job at IBM. Ohlman began his work on the permutation system in Lexington, MA and continued when SDC moved to Santa Monica, CA. Doyle (1966) described Ohlman as ". . . undoubtedly the key figure during 1958 in the unfolding of the language processing and retrieval story [at SDC]" (p. 5). Doyle described Ohlman as an enthusiastic leader in the work developing the permutation system and credits him with the beginning of SDC's later work in information retrieval (p. 7).

The first edition of the index to the SAGE documents, titled "Permutation Index No.1," indexed 1,800 reports and was produced sometime shortly after June 1957; the second edition indexed 4,000 SDC and Lincoln Lab reports and was produced sometime shortly after June 1958. The first published announcement of the permutation indexing system came in October 1958 in the third issue of the National Science Foundation's (NSF) *Current Research and Development in Scientific Documentation* (*CRDSD*). This relatively new publication was created by NSF to keep researchers and practitioners aware of new developments in documentation. The note, by Ohlman, Lewis Hart, and Lauren Doyle (1958), all of SDC, stated:

> This group is developing a system of "permutation indexing," which consists of an index to document titles, organized in such a way that all titles containing a particular word are automatically grouped together. This operation is actually a cyclic permutation of words. (p. 34)

The note went on to detail exactly how the cards are keypunched, sorted, and words kept in context in the printed index. At the same time, during the summer of 1958, Ohlman was advising staff at the Rocketdyne Company of North American Aviation, Inc., Canoga Park, CA, in their work to prepare an "electronic permutation index" to material handling machines and equipment. Rocketdyne published this in August 1958 and acknowledged Ohlman's help (Rocketdyne Corp., 1958[1]).

As this work was going on at SDC, Ohlman also submitted a paper, later accepted, on superimposed coding for the upcoming ICSI conference [Interestingly, Luhn was the organizing chair of the Area 5 section of ICSI, where Ohlman presented his paper, and Luhn and Ohlman exchanged several letters regarding the paper and the refereeing process (Ohlman files, 1957–1958)]. Acceptance of his paper entitled Ohlman to receive the ICSI papers preprints. Olhman (1999) later recalled, "Here was a perfect way to demonstrate the speed and automation features of permutation indexing to information science and technology colleagues" (p. 188). Ohlman and his colleagues, Hart, of the SAGE Computing Project, and Joan Citron, then prepared this demonstration index, titled "A Permutation Index to the Preprints of the International Conference on Scientific Information" (Citron, Hart, & Ohlman, 1958) that Doyle (1966, p. 5) would later describe this as ". . . an all-out effort to 'scoop IBM' . . .," implying that Ohlman's team knew in advance about IBM's work on KWIC. Exactly what they knew about IBM and Luhn's work on KWIC, however, is not clear. The index was distributed at the conference to all participants. The index contained more than title words: "Entries were selected not just from titles; they included author names and affiliations, headings, captions, sentences, and even phrases selected for their significance as thought units" (Ohlman, 1999, p. 188). According to Ohlman (1999), an average of five entries for each of 1,400 pages of preprints was provided by the index, additional contextual information was added, and the keywords were printed in the middle of the index page instead of the left-hand side, as it had with the SDC documents indexes. The bibliography of the papers included what served as a summary of each article's contents, including any tables, author names, and affiliations. All work was done with tabulating equipment and punched cards (see Figures 2 and 3).

The edition of the ICSI preprints index given out at the conference was a spiral-bound, probably offset-produced, typescript of about 35 pages. It contained the following statement about the origins of the index:

> This tool was also developed independently by the Rocketdyne Division of North American Aviation as "rotational indexing," and by IBM as "keyword in context." It is also related to the "rotated file" of the Chemical Biological Coordination Center, and the "correlative indexing" of Chemical Abstracts. It employs the principles of "extraction" developed by Battelle Memorial Institute. The earliest reference to the

[1] In a note written in Ohlman's hand on the letter of thanks from Rocketdyne, he says that he thinks the Rocketdyne staff was not aware of Luhn's work on KWIC.

DOI: 10.1002/asi

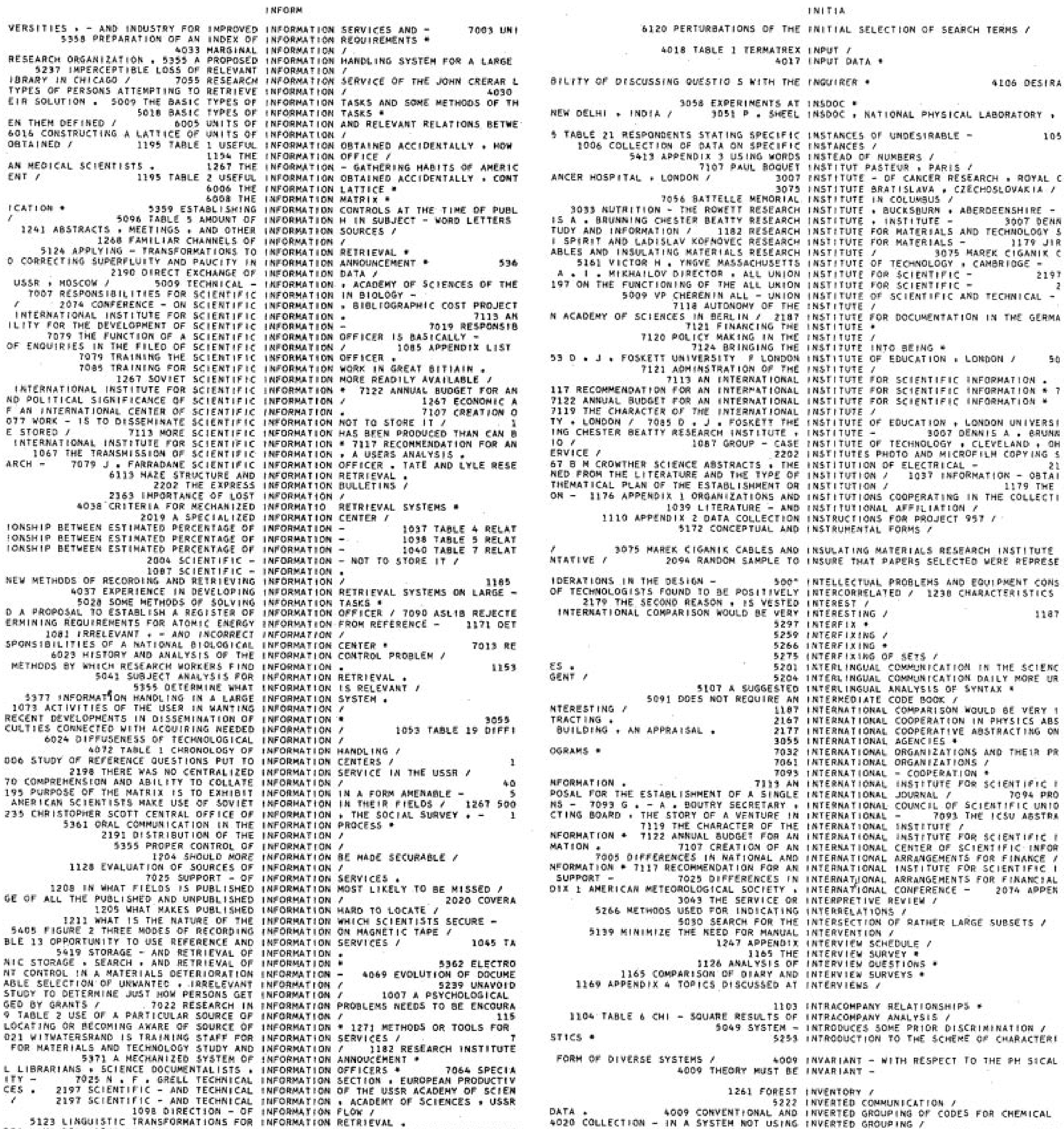

FIG. 2. Sample permutation index page from J. Citron, L. Hart, & H. Ohlman, *A Permutation Index to the Preprints of the International Conference on Scientific Information.* Santa Monica, CA: System Development Corp., December 15, 1959 (SP-44), p. 18. Numbers to the left of the index entry and contextual information are area and page numbers in the ICSI preprints volume. Reproduced from the original publication.

DOI: 10.1002/asi

1003 LITERATURE AND REFERENCE NEEDS OF SCIENTISTS .
1003 KNOWLEDGE NOW AVAILABLE AND METHODS OF ASCERTAINING REQUIREMENTS *
1006 CASE STUDIES OF ACTUAL RESEARCH PROJECTS /
1006 COLLECTION OF DATA ON SPECIFIC INSTANCES /
1006 STUDY OF REFERENCE QUESTIONS PUT TO INFORMATION CENTERS /
1006 STUDIES OF THE RESULTS OBTAINED AND CHANGES IN THE WORK HABITS /
1007 A PSYCHOLOGICAL STUDY TO DETERMINE JUST HOW PERSONS GET INFORMATION /
1007 TIME WORTHWHILE TO SPEND IN A LITERATURE SEARCH /
1007 ASSIGNMENT OF THE SAME RESEARCH PROBLEM TO TWO DIFFERENT GROUPS /

1009 STUDY ON THE USE OF SCIENTIFIC LITERATURE AND REFERENCE SERVICES -
1009 SERVICES - BY SCANDINAVIAN SCIENTISTS AND ENGINEERS ENGAGED IN RESEARCH -
1009 RESEARCH - AND DEVELOPMENT .
1009 EARLIER STUDIES IN THE FIELD *
1009 ELIN TORNUDD SECRETARY OF THE SCANDINAVIAN COUNCIL FOR APPLIED RESEARCH /
1011 BIBLIOGRAPHY OF THE USE OF SCIENTIFIC LITERATURE AND REFERENCE SERVICES -
1011 SERVICES - AS REVEALED BY STUDIES DIRECTED TO THE SCIENTIST /
1022 BIBLIOGRAPHY OF EARLIER STUDIES ON THE USE OF SCIENTIFIC LITERATURE -
1022 LITERATURE - AND REFERENCE SERVICES ARRANGED BY SUBJECT FIELD /
1032 THE SCANDINAVIAN STUDY *
1034 TABLE 1 DISTRIBUTION OF RESPONDENTS BY FIELD OF RESEARCH /
1035 TABLE 2 DISTRIBUTION OF RESPONDENTS BY ACADEMIC DEGREE /
1036 TABLE 3 ESTIMATED ABILITY IN KEEPING UP WITH THE NEW DEVELOPMENTS /
1037 TABLE 4 RELATIONSHIP BETWEEN ESTIMATED PERCENTAGE OF INFORMATION -
1037 INFORMATION - OBTAINED FROM THE LITERATURE AND THE TYPE OF INSTITUTION /
1038 TABLE 5 RELATIONSHIP BETWEEN ESTIMATED PERCENTAGE OF INFORMATION -
1038 INFORMATION - OBTAINED FROM THE LITERATURE AND THE FIELD OF RESEARCH /
1039 TABLE 6 RELATIONSHIP BETWEEN ESTIMATED TIME SPENT ON LITERATURE -
1039 LITERATURE - AND INSTITUTIONAL AFFILIATION /
1040 TABLE 7 RELATIONSHIP BETWEEN ESTIMATED PERCENTAGE OF INFORMATION -
1040 INFORMATION - OBTAINED FROM THE LITERATURE AND THE ESTIMATED TIME /
1040 TABLE 8 NUMBER OF JOURNALS PERSONALLY RECEIVED /
1041 TABLE 9 NUMBER OF JOURNALS REGULARLY READ /
1042 TABLE 10 RELATIVE IMPORTANCE OF DIFFERENT PUBLICATIONS /
1044 TABLE 11 NUMBER OF FOREIGN LANGUAGES READ /
1045 TABLE 12 KEEPING PERSONAL INDEXES /
1045 TABLE 13 OPPORTUNITY TO USE REFERENCE AND INFORMATION SERVICES /
1046 TABLE 14 USE OF DIFFERENT LIBRARY SERVICES /
1049 TABLE 15 UNAVAILABLE SERVICES FOR WHICH THE NEED IS GREATEST /
1050 TABLE 16 NUMBER OF PAPERS PUBLISHED DURING THE PAST 3 YEARS /
1051 TABLE 17 NATURE OF PUBLISHING MEDIA USED FOR OWN PAPERS /
1052 TABLE 18 DIFFICULTIES WITH OWN PUBLICATIONS /
1053 TABLE 19 DIFFICULTIES CONNECTED WITH ACQUIRING NEEDED INFORMATION /
1054 TABLE 20 SPECIFIC CASES OF UNDESIRABLE DUPLICATION OF RESEARCH /
1055 TABLE 21 RESPONDENTS STATING SPECIFIC INSTANCES OF UNDESIRABLE -
1055 UNDESIRBALE - DUPLICATION /
1056 TABLE 22 NUMBER OF RESPONDENTS PRESENTING CONCRETE SUGGESTIONS -
1059 TABLE 23 RESPONDENTS HAVING FELT A NEED FOR IMPROVING -
1059 IMPROVING - THEIR SKILL IN THE USE OF LITERATURE AFTER GRADUATION /
1063 APPENDIX 1 ENGLISH TRANSLATION OF THE CIRCULAR LETTER MAILED /
1064 APPENDIX 2 ENGLISH TRANSLATION OF THE QUESTIONNAIRE FOR RESEARCH WORKERS /

1067 THE TRANSMISSION OF SCIENTIFIC INFORMATION , A USERS ANALYSIS .
1067 J . D . BERNAL PROFESSOR OF PHYSICS , BIRBECK COLLEGE , -
1067 COLLEGE , - UNIVERSITY OF LONDON /
1070 THE AMOUNT TO BE READ INCREASES EXPONENTIALLY /
1072 ANALYSIS BASED ON THE NATURE OF THE MATERIAL USED /
1073 ACTIVITIES OF THE USER IN WANTING INFORMATION /
1077 THE PRIMARY REASON FOR PUBLISHING ORIGINAL SCIENTIFIC WORK -
1077 WORK - IS TO DISSEMINATE SCIENTIFIC INFORMATION NOT TO STORE IT /
1078 THOSE WHO NEED TO FOLLOW SCIENCE HAVE LEAST OPPORTUNITY /
1079 PRESERVING ACCESS TO ORIGINAL PAPERS /
1081 SEMI - AUTOMATIC FILTER AGAINST REPETITIVE , IRRELEVANT , -
1081 IRRELEVANT , - AND INCORRECT INFORMATION /
1085 3 REFERENCES /
1085 APPENDIX LIST OF ENQUIRIES IN THE FILED OF SCIENTIFIC INFORMATION /

1087 AN OPERATIONS RESEARCH STUDY OF THE DISSEMINATION OF SCIENTIFIC -
1087 SCIENTIFIC - INFORMATION .
1087 GENERAL DESIGN OF THE RESEARCH *
1087 MICHAEL H . HALBERT AND RUSSELL L . ACKOFF OPERATIONS RESEARCH GROUP -
1087 GROUP - CASE INSTITUTE OF TECHNOLOGY , CLEVELAND , OHIO /
1089 THE TIME AVAILABLE FOR SCIENTIFIC RESEARCH /
1090 HOW SCIENTISTS SPEND THEIR TIME *
1094 FIGURE 1 CHEMISTS TIME ALLOCATION SUMMARY /
1095 FIGURE 2 CHEMISTS TIME ALLOCATION . DETAIL /
1096 TABLE 1 PERCENT TIME ALLOCATION /
1096 TABLE 2 PERCENT SCIENTIFIC COMMUNICATION TIME ALLOCATION /
1096 TABLE 3 SENDING - RECEIVING VS . WRITTEN - ORAL , PERCENT /
1097 FIGURE 3 CHEMISTS COMMUNICATION TIME , ALLOCATION TO CHANNEL /
1098 FIGURE 4 CHEMISTS ALLOCATION OF COMMUNICATION TIME BY DIRECTION -
1098 DIRECTION - OF INFORMATION FLOW /
1099 FIGURE 5 CHEMISTS ALLOCATION OF COMMUNICATION TIME BY PERSON -
1099 PERSON - COMMUNICATED WITH /
1100 FIGURE 6 RELATIONSHIP BETWEEN EQUIPMENT USE AND COMMUNICATION /
1100 FIGURE 7 UNEXPLAINED EQUIPMENT USE /
1101 FIGURE 8 UNEXPLAINED SCIENTIFIC COMMUNICATION /
1102 TABLE 4 CHI - SQUARE TESTS ON DEVIATIONS FROM EXPECTED VALUE -
1102 VALUE - OF DEPENDENT ACTIVITY /
1103 TABLE 5 GROUP SIZE AS IT AFFECTS TIME ALLOCATION /
1103 INTRACOMPANY RELATIONSHIPS *
1104 TABLE 6 CHI - SQUARE RESULTS OF INTRACOMPANY ANALYSIS /
1110 APPENDIX 2 DATA COLLECTION INSTRUCTIONS FOR PROJECT 957 /
1117 APPENDIX 3 /
1118 APPENDIX 4 THE USE OF THE MULTINOMIAL ASSUMPTION /
1119 APPENDIX 5 SCIENTIFIC COMMUNICATION STUDY - QUESTIONNAIRE /

1121 INFORMATION AND LITERATURE USE IN A RESEARCH AND DEVELOPMENT -
1121 DEVELOPMENT - ORGANIZATION .
1121 ABSTRACT .
1121 I . H . HOGG AND J . ROLAND SMITH INDUSTRIAL GROUP , UNITED KINGDOM -
1121 KINGDOM - ATOMIC ENERGY AUTHORITY , RISLEY , WARRINGTON , LANCASHIRE ,
1121 LANCASHIRE , - ENGLAND
1125 MAIN SURVEY , COMPOSITION OF THE SAMPLE *
1126 ANALYSIS OF INTERVIEW QUESTIONS *
1127 ADEQUACY OF TIME FOR READING AT WORK /
1128 EVALUATION OF SOURCES OF INFORMATION /
1129 LITERATURE SEARCHING HABITS /
1130 PERSONAL INDEXES /
1130 CRITICISMS OF LIBRARY LISTS /
1131 FIGURE 1 LITERATURE SEARCHING HABITS /
1133 CRITICISMS OF THE LIBRARY SERVICE /
1133 PRIVATE PROVISION OF LITERATURE /
1134 VALUE OF PERIODICALS ACCORDING TO AGE /
1136 DOMESTIC AND FOREIGN PERIODICALS , PROPORTIONATE USE /
1136 VALUE OF PERIODICALS ACCORDING TO LANGUAGE /
1137 ANALYSIS OF READING DIARIES *
1138 FIGURE 3 VALUE OF PERIODICALS ACCORDING TO LANGUAGE /
1140 AMOUNT OF LITERATURE READ , AND WHERE READ /
1140 READING OF ABSTRACT JOURNALS /
1141 FIGURE 4 READING DIARIES - WHERE THE LITERATURE WAS OBTAINED /
1143 READING DIARIES - HOW THE LITERATURE REFERENCE WAS FOUND /
1143 USE MADE OF THE LITERATURE /
1145 4 REFERENCES /
1145 APPENDIX 1 SCORING SCALE FOR QUESTIONNAIRE /

1153 METHODS BY WHICH RESEARCH WORKERS FIND INFORMATION .
1153 ABSTRACT .
1153 R . M . FISHENDEN ATOMIC ENERGY RESEARCH ESTABLISHMENT , HARWELL , -
1153 HARWELL , - ENGLAND /
1154 SUMMARY OF SERVICES GIVEN BY A . E . R . E . LIBRARY *
1154 READING ROOM AND REPORT ROOMS /
1154 THE INFORMATION OFFICE /
1155 DIVISIONAL LIBRARIES /
1155 CHOICE OF SURVEY METHOD *
1156 THE DIARY CARD INVESTIGATION *
1157 RESULTS OF DIARY CARD SURVEY *
1158 TABLE 1 OVERALL RESULTS 63 DIARISTS /
1159 TABLE 2 USE OF A PARTICULAR SOURCE OF INFORMATION /
1160 TABLE 3 MOST EFFECTIVE RETRIEVAL METHODS /
1161 TABLE 4 SPREAD OF NUMBER OF ITEMS RECORDED PER DIARY KEEPER /
1162 BREAKDOWN OF RESULTS FOR DIFFERENT TYPES OF DIARIST /
1163 TABLE 5 COMPARISON OF SENIOR AND JUNIOR STAFFS /
1164 TABLE 6 COMPARISON OF PURE AND APPLIED RESEARCH WORKERS /
1165 THE INTERVIEW SURVEY *
1165 COMPARISON OF DIARY AND INTERVIEW SURVEYS *
1165 TABLE 7 PROPORTION OF SAMPLE MAKING USE OF PARTICULAR REFERENCE SOURCE /
1166 2 REFERENCES /
1168 APPENDIX 2 NOTES ON USE OF DIARY CARDS /
1169 APPENDIX 3 DISCUSSION OF DIARY CARD LAYOUT /
1169 APPENDIX 4 TOPICS DISCUSSED AT INTERVIEWS /

1171 DETERMINING REQUIREMENTS FOR ATOMIC ENERGY INFORMATION FROM REFERENCE -
1171 REFERENCE - QUESTIONS .
1171 SAUL HERNER AND MARY HERNER HERNER AND COMPANY , WASHINGTON , D . C . /
1172 METHOD OF STUDY *
1173 THE NATURE OF THE REFERENCE QUESTIONS *
1174 THE CONCEPTUAL STRUCTURE OF THE REFERENCE QUESTIONS *
1175 CONCEPTS CONTAINED IN THE 3851 QUESTIONS /
1175 LOGICAL RELATIONSHIP OF THE CONCEPTS CONTAINED IN A REFERENCE QUESTION /
1176 APPENDIX 1 ORGANIZATIONS AND INSTITUTIONS COOPERATING IN THE COLLECTION -
1176 COLLECTION - OF REFERENCE QUESTIONS /
1177 APPENDIX 2 CATEGORIES OF NON - TECHNICAL QUESTIONS COLLECTED /
1177 APPENDIX 3 CATEGORIES OF TECHNICAL QUESTIONS COLLECTED /

1179 SYSTEMATICALLY ASCERTAINING REQUIREMENTS OF SCIENTISTS FOR INFORMATION .
1179 PROVIDE INFORMATION ABOUT NEWLY PUBLISHED DOCUMENTS -
1179 DOCUMENTS - BY MEANS OF ABSTRACTS CARDS /
1179 THE THEMATICAL PLAN OF THE ESTABLISHMENT OR INSTITUTION /
1179 JIRI SPIRIT AND LADISLAV KOFNOVEC RESEARCH INSTITUTE FOR MATERIALS -
1179 MATERIALS - AND TECHNOLOGY , PRAGUE , CZECHOSLOVAKIA /
1180 REQUIREMENTS OF INDIVIDUAL RESEARCH WORKERS /
1180 INFORMATION REQUIREMENT SHEET /
1182 RESEARCH INSTITUTE FOR MATERIALS AND TECHNOLOGY STUDY AND INFORMATION /
1182 INFORMATION REQUIREMENT SHEET FOR 1958 /

1185 HOW SCIENTISTS ACTUALLY LEARN OF WORK IMPORTANT TO THEM .
1185 NEW METHODS OF RECORDING AND RETRIEVING INFORMATION /
1185 BENTLEY GLASS AND SHARON H . NORWOOD JOHNS HOPKINS UNIVERSITY , -
1185 UNIVERSITY , - BALTIMORE , MARYLAND /
1186 DIFFERENCES WERE APPARENT BETWEEN THE WORKERS IN DIFFERENT FIELDS /
1186 NO POSITIVE CONCLUSIONS SHOULD BE DRAWN /
1186 TABLE 1 RANKING OF METHODS WHEREBY 50 REPRESENTATIVE SCIENTISTS -
1186 SCIENTISTS - ACTUALLY LEARNED OF WORK CRUCIAL TO THEIR OWN /
1186 PHYSIOLOGISTS AND BIOPHYSICISTS NEVER REFERRED TO CROSS CITATIONS /
1186 THE BIOCHEMISTS NEVER REPORTED DISCOVERIES IN JOURNALS THEY REGULARLY -
1186 REGULARLY - SUBSCRIBE TO /
1186 THE BIOPHYSICISTS , EMBRYOLOGISTS , AND GENETICISTS ALMOST WITHOUT -
1186 WITHOUT - EXCEPTION ESCHEW ABSTRACTING SERVICES /
1186 BIOPHYSICISTS AND PSYCHOLOGISTS FAIL TO USE REPRINTS /
1187 INTERNATIONAL COMPARISON WOULD BE VERY INTERESTING /
1189 PLANNED AND UNPLANNED SCIENTIFIC COMMUNICATION .
1189 HERBERT MENZEL BUREAU OF APPLIED SOCIAL RESEARCH , COLUMBIA UNIVERSITY ,
1189 UNIVERSITY - NEW YORK /
1190 A SYSTEMIC VIEW /
1191 MULTIPLE FUNCTIONS OF THE SCIENTIFIC COMMUNICATION SYSTEM /
1192 KEEPING SCIENTISTS UP TO DATE *
1193 UNPLANNED MECHANISMS /
1195 TABLE 1 USEFUL INFORMATION OBTAINED ACCIDENTALLY , HOW OBTAINED /
1195 TABLE 2 USEFUL INFORMATION OBTAINED ACCIDENTALLY , CONTENT /
1196 INDIVIDUAL ACCIDENT - AGGREGATE REGULARITY /
1197 FURNISHING ANSWERS TO SPECIFIC QUESTIONS *
1198 SEARCHES OUTSIDE THE LITERATURE /
1199 PERSONS CONSULTED AS SOURCES /
1201 SOME QUESTIONS FOR FUTURE RESEARCH *
1202 DETAILED QUESTIONS ON THE FURNISHINGS OF ANSWERS TO SPECIFIC QUESTIONS *
1204 SHOULD INFORMAL CHANNELS BE MADE MORE WIDELY USABLE /
1204 SHOULD MORE INFORMATION BE MADE SECURABLE /
1205 WHAT MAKES PUBLISHED INFORMATION HARD TO LOCATE /
1205 WHY IS INFORMATION OF CERTAIN TYPES SELDOM PUBLISHED /
1206 DETAILED QUESTIONS ON KEEPING SCIENTISTS UP TO DATE *
1207 WHY WERE SPECIFIC PUBLISHED ITEMS OF SCIENTIFIC NEWS MISSED BY SCIENTIST /
1208 IN WHAT FIELDS IS PUBLISHED INFORMATION MOST LIKELY TO BE MISSED /
1209 WHAT ARE THE FORMS OF PERSONAL COMMUNICATION /
1210 RESEARCH QUESTIONS CONCERNING REDUCTIONS IN THE TIME AND LABOR -
1210 LABOR - OF SURVEYING CURRENT DEVELOPMENTS /
1211 WHAT IS THE NATURE OF THE INFORMATION WHICH SCIENTISTS SECURE -
1211 SECURE - IN THE MEDIA SCANNED WITH LEAST EFFICIENCY /
1212 RESEARCH QUESTIONS REGARDING SCIENTISTS MOTIVATION /

1235 THE USE OF TECHNICAL LITERATURE BY INDUSTRIAL TECHNOLOGISTS .
1235 CHRISTOPHER SCOTT CENTRAL OFFICE OF INFORMATION , THE SOCIAL SURVEY , -
1235 SURVEY , - LONDON /
1236 DESIGN OF THE SURVEY *
1238 A GENERAL ACTIVITY FACTOR AMONG TECHNOLOGISTS /
1238 CHARACTERISTICS OF TECHNOLOGISTS FOUND TO BE POSITIVELY INTERCORRELATED /
1239 THE AMOUNT AND NATURE OF TECHNICAL READING /
1240 TABLE 1 WHERE DO YOU DO MOST OF YOUR TECHNICAL READING /
1241 TABLE 3 HOW WAS YOUR ATTENTION DRAWN TO THIS ARTICLE /
1241 ABSTRACTS , MEETINGS , AND OTHER INFORMATION SOURCES /
1242 TABLE 5 USE MADE OF ABSTRACTS /
1244 THE ROLE OF THE LITERATURE IN THE TECHNOLOGISTS WORK /
1247 APPENDIX INTERVIEW SCHEDULE /
1252 FIRMS CLASSIFICATION SHEET /

1257 REQUIREMENTS OF FOREST SCIENTISTS FOR LITERATURE AND REFERENCE SERVICES .
1257 STEPHEN H . SPURR PROFESSOR , SCHOOL OF NATURAL RESOURCES , UNIVERSITY -
1257 UNIVERSITY - OF MICHIGAN , ANN ARBOR , MICHIGAN /
1258 THE PLACE OF LITERATURE IN FOREST RESEARCH *
1259 THE FOREST LITERATURES IN GERMAN , RUSSIAN , SWEDISH , FINNISH /
1259 THE NATURE OF FORESTRY LITERATURE *
1259 RELATIVELY LITTLE OF IT IS ADEQUATELY SUMMARIZED IN BOOK FORM /
1259 ONE OR TWO AMERICAN TREE SPECIES HAVE BEEN ADEQUATELY MONOGRAPHED /
1260 BEST RESEARCH EVENTUALLY IS PUBLISHED IN BULLETIN OR JOURNAL FORM /
1260 SUBSTANTIAL FUNDAMENTAL RESEARCH IS AVAILABLE ONLY IN UNPUBLISHED THESES /
1260 FOREST AERIAL PHOTOGRAMMETRY /
1261 FOREST INVENTORY /
1262 SPECIFIC GRAVITY AND TREE GROWTH /
1262 THE NATURE OF THE FOREST COMMUNITY /
1263 EXISTING REFERENCE FACILITIES *
1264 BIOLOGICAL ABSTRACTS COVERS SOME FORESTRY LITERATURE /
1264 FORESTRY ABSTRACTS /
1265 THE OXFORD CLASSIFICATION WOULD BE USED /
1266 REPRODUCTION FACILITIES WOULD BE DEVELOPED FOR THE PRINTING -
1266 PRINTING - OF SUFFICIENT ADDITIONAL CARDS FOR CROSS - INDEXING /
1266 FACILITIES FOR PHOTOGRAPHIC REDUCTION AND REPRODUCTION /
1266 SINGLE CATALOG CONTAINING ABSTRACTS , CITATIONS , AND PHOTOGRAPHIC -
1266 PHOTOGRAPHIC - REDUCTIONS OF BRIEF ARTICLES /
1266 THE PROTOTYPE DEVELOPED BY THE AUTHOR /

1267 THE INFORMATION - GATHERING HABITS OF AMERICAN MEDICAL SCIENTISTS .
1267 METHODS BY WHICH SCIENTISTS OBTAIN INFORMATION /
1267 ECONOMIC AND POLITICAL SIGNIFICANCE OF SCIENTIFIC INFORMATION /
1267 SOVIET SCIENTIFIC INFORMATION MORE READILY AVAILABLE /
1267 500 AMERICAN SCIENTISTS MAKE USE OF SOVIET INFORMATION IN THEIR FIELDS /
1267 SAUL HERNER HERNER AND COMPANY , WASHINGTON , D . C . /
1268 A TRICHOTOMOUS STUDY OF AMERICAN MEDICAL SCIENTISTS /
1268 FAMILIAR CHANNELS OF INFORMATION /
1269 METHODS OF KEEPING ABREAST OF CURRENT DEVELOPMENTS *

1

FIG. 3. Sample bibliography page from J. Citron, L. Hart, & H. Ohlman, *A Permutation Index to the Preprints of the International Conference on Scientific Information.* Santa Monica, CA: System Development Corp., December 15, 1959 (SP-44), p. 1. The lines separate different papers presented at the ICSI Conference; the numbers to the left indicate the area of ICSI where the paper was presented, and the page number in the ICSI preprints volume. Note that even tables, figures, and appendices for each paper are detailed. Reproduced from the original publication.

DOI: 10.1002/asi

FIG. 4. Photograph of H.P. Luhn demonstrating automatic input-output equipment for the 9900 Special Index Analyzer to Richard Angell, Library of Congress, at the International Conference on Scientific Information, November 17, 1958. Reproduced, with permission of ASIST, from *H.P. Luhn, Pioneer of Information Science: Selected Works*, by C.K. Schultz (Ed.). New York: Spartan Books; copyright by American Documentation Institute, p. 32.

principle appears in Crestadoro's "Art of Making Catalogs of Libraries," 1856. (Citron et al., November 1958)

Following the conference, Ohlman received many requests for the ICSI index and instructions on how to prepare the program for others to use. The demand for additional copies caused a "Revised Edition" to be issued on December 15, 1959 (Citron, Hart, & Ohlman, 1959). In this edition, Ohlman and colleagues supplemented his earlier acknowledgment of other "closely related" systems by adding Charles L. Bernier, of Chemical Abstracts Service, and D.B. Netherwood at Wright Air Development Center.

After ICSI, Ohlman's colleagues Lewis Hart and G.R. Bach began to use permutation indexing to analyze verbatim transcripts of psychiatric patients (Ohlman, 1999, p. 188). In the April 1959 issue of *CRDSD*, Ohlman stated that the permutation indexing system was "tested" on the ICSI preprints and noted that the system can be ". . . applied to tabular data, using words, numbers or symbols" (Ohlman, Hart, & Doyle, 1959, p. 28). In early 1959, Ohlman and his SDC colleagues, a UCLA librarian, and the developers of the system at Rocketdyne Corporation formed a company called Permutation Indexing for the Literature of Technology, Inc. (PILOT) to attempt to commercialize the development. Ohlman (1999) recalled: "Unfortunately, despite almost one hundred subscriptions, the company was undercapitalized and did not survive long enough to distribute the first issue" (p. 190). Again, as Ohlman (1999) described it, ". . . I found little support for further development of permutation indexing and joined SDC's medical automation project" (p. 189).

Ohlman did, however, continue trying to understand the cost aspects of using his system for indexing. As preparation for his presentation on permutation indexing at the ADI meeting in October 1960, he sent several letters to IBM staff, including Luhn, asking about their costs for indexing using KWIC and requesting a copy of the program for producing KWIC. He received polite answers, some minimal cost information (with a note that he should discuss costs in more detail with Chemical Abstracts Service, which was beginning production of *Chemical Titles*), but apparently never received a copy of IBM's program for producing KWIC. He also tried without success to get the SDC medical automation project interested in permutation indexing and prepared a proposal to the U.S. Veterans Administration to use it in their internal documents control (Ohlman Papers, 1955–2002).

By October 1960, Ohlman had left the SDC in Santa Monica, CA and was now employed by Lockheed Electronics Systems Research Center in Bedminster, NJ. He never managed to return to permutation indexing work, despite a continuing interest in information retrieval, mechanical indexing, and automatic indexing. Instead, he had to watch from afar the increasing growth of KWIC indexing, research on automatic indexing, and the almost complete loss of any memory within the information community of his work in developing permutation indexing.

Hans Peter Luhn and the Origins and Development of KWIC Indexing

Hans Peter Luhn (Figure 4) was born in 1896 in Barmen, Germany, the first child of commercial printers. He was educated in Germany and in Switzerland, with an excellent education in technology, physics, and accounting, and served for a brief time as a printer's apprentice. From an early age, he was known for his inventiveness, including designing a punched-card system for controlling puppet show lighting. In the 1920s, he worked as an assistant manager of a textile

DOI: 10.1002/asi

mill in Italy and continued his freelance work as a textile designer. He came to the United States in 1924 as an agent of a German textile firm, but soon discovered that it had become bankrupt. He then became a bank clerk and quickly rose through the ranks to become a financial secretary for the International Acceptance Bank in New York City. In 1927, he became an assistant to the president of Textile Machine Works in Reading, Pennsylvania, where he continued to display his inventiveness, accumulating 10 patents by 1930. These included the Lunometer, a device for making instant thread counts on fabric, a punched paper tape method for controlling changes in the knitting process, and another to improve the process of knitting women's silk stockings. He established his own engineering consulting firm in 1930, H.P. Luhn & Associates, and continued to apply for and receive new patents, including the cocktail oracle recipe guide, which used optical coincidence searching (Schultz, 1968).

When Luhn joined IBM in 1941 as a development engineer, he already held 20 patents in his name, some of which he sold to IBM. He was hired as an "inventor," a practice that IBM occasionally employed. He was given a group to work with, an office, a laboratory, and considerable freedom to work on projects that interested him. Working first in New York City, then Armonk, and later in Poughkeepsie, his team first concentrated on relay adding and calculating. Later, after the end of World War II, he and his team (now numbering about 25) turned their attention to trying different ways to improve the amount of information that could be coded onto punched cards. When James W. Perry and Malcolm Dyson approached Thomas W. Watson, Sr., President of IBM, in 1948 for help in designing a system for improving the handling of chemical information, Watson asked Luhn to work with them. The result was the "Luhn scanner," a machine that used punched cards, photoelectric cells, and logical search circuitry that relied on Boolean logic. Experimental models were developed and tested, and they received considerable attention from the chemical information community; however, the machine was never put into production by IBM because of limited need and rapid changes in technology stemming from the newly developing electronic computer. (Moore, 1968; Williams, 2002, p. 21).

The "Luhn scanner" was Luhn's first venture into documentation work, but it certainly was not his last. After 1948, Luhn devoted almost all of his time at IBM to information retrieval projects. When he retired from IBM in 1961, he had over 80 patents, many of them related to information retrieval. The best known of his developments in information science are the KWIC and SDI systems, which he considered complementary to each other. Luhn died of leukemia on August 19, 1964, in the middle of his year as President of ASIS.

According to Mary Elizabeth Stevens' (1968) brief biography of Luhn's career in information science, he was well prepared for the advances represented by KWIC and SDI: "It was indeed his challenging thesis that physical prototypes of meaningful messages, especially as available in machine-readable form, could be automatically manipulated to provide better tools for search, selection, and use of pertinent message contents" (p. 24). His work with Dyson and Perry on chemical information codes was the beginning of this interest and was continued by his work in 1956 on computer processing of full, natural-language text. By 1957, he was working on superimposed coding, statistical approaches to encoded searching of literary information, and collecting statistical information on keywords in texts (Luhn, 1957). By early 1958, he was experimenting with the automatic creation of literature abstracts (Luhn, 1958a). His 1958 paper (probably written in late 1957 or early 1958) "Indexing, Language, and Meaning" demonstrated a superb understanding of indexing fundamentals and shows that he decisively favored what he called a "native" approach, based on statistical analysis of a collection, declaring it to be the most effective one for information retrieval (Luhn, 1958b). In this same article, he also made a strong argument for machine-based automatic thesaurus construction.

The development of the KWIC automatic indexing technique, based on title word permutation, was a natural outgrowth of his earlier work. Luhn's work also was likely greatly aided by the work of his IBM colleague in California, Phyllis B. Baxendale, who was developing machine-made indexes to technical literature. Her emphasis was on identifying essential discrimination indices such as keywords, topic sentences, and coordinated indexes (Baxendale, 1958). Indeed, Stevens (1968) stated that "Baxendale in particular is generally credited with the first actual experiments in modified derivative indexing" (p. 73).

Stevens (1968), based on personal correspondence with Luhn, gives a brief chronology of the development of the KWIC idea:

> May, 1957: program for word isolation within 60 characters per card, written by H.C. Fallon [an IBM coworker];
>
> 1957–1958 creation of concordances of various scientific papers in the form of cards, each card showing a keyword centrally located within 60 letters worth of the associated phrase. Experimentation with these cards to arrive at thesauri for special fields . . . Idea of automatic indexing by means of significant or Keywords in context conceived by H.P. Luhn;
>
> May, 1958: Keyword-in-Context Index for titles only, initiated by H.P. Luhn and samples produced with Routine 1 program;
>
> June, 1958: started punching of titles for Keyword-in-Context for literature on information retrieval and machine translation . . .;
>
> August, 1958: simplified version of Routine 1, written by H.C. Fallon generating KWIC and delivered to Service Bureau Corporation;
>
> September, 1958: first edition of bibliography and KWIC on information retrieval and machine translation published by Service Bureau Corporation. (pp. 27–28) (see Figures 5 and 6)

According to Stevens, Luhn did not claim to have originated the idea of title permutation indexing, recognizing its similarities to concordances and similar keyword indexing. She quoted a letter that Luhn sent to Charles L. Bernier,

DOI: 10.1002/asi

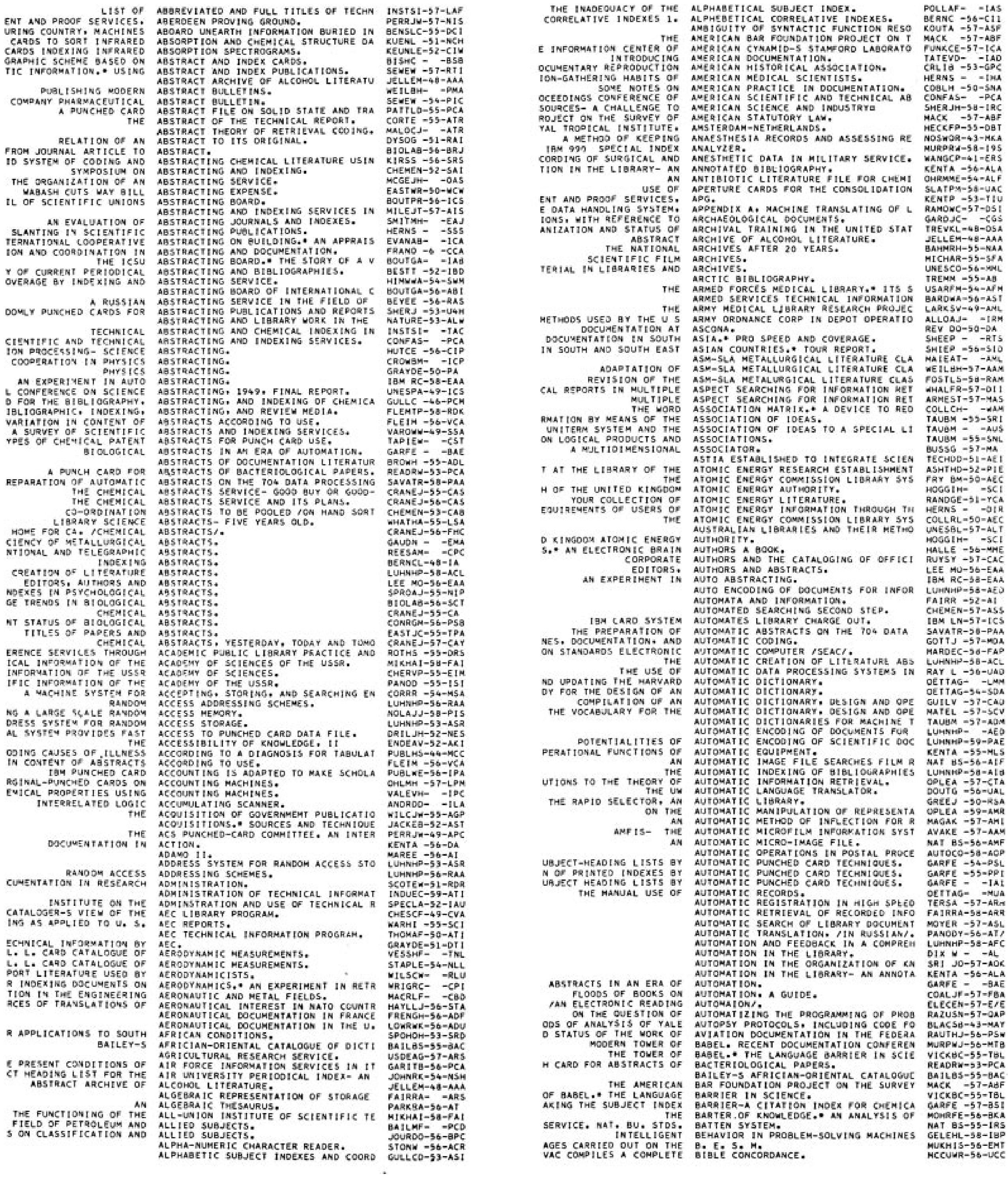

FIG. 5. Sample KWIC index page from H.P. Luhn & P. James (1959), *Literature on Information Retrieval and Machine Translation: Bibliography and Index; Titles Indexed by Key-Words-in-Context System* (2nd ed.). Yorktown Heights, NY: IBM Research Center, June 1959, p. 22. Abbreviations and numbers to the right of the keywords refer to items listed in the bibliography. Reproduced from a copy of the original.

a noted information science leader at the time, in 1960 that clarified what he did claim:

> I would say that the significance of the present KWIC Index is based on the fact that it is produced automatically by machine, affording speed of compilation, accuracy, and completeness. As far as the particular format of the Index is concerned, this is novel to my knowledge, in accordance with the information I have been able to ascertain from others. (as cited in Stevens, 1968, p. 28).

By the time of the ICSI conference in November 1958, Luhn already had produced the first edition of his KWIC index to the literature of information retrieval and machine

DOI: 10.1002/asi

```
ABA OR-54-SPC  A.B.A. OPERATIONS RESEARCH COMMITTEE
                 SUBCOMMITTEE ON PUNCHED CARDS AND LEGAL LITERATURE
                   SEARCHING.
                 NATURE AND SCOPE OF PROGRAM JUNE 1954
ADAIWC-55-CIS  ADAIR WC
                 CITATION INDEXES FOR SCIENTIFIC LITERATURES.
                 AMERICAN DOCUMENTATION. 6. /1/. 1955.
ADAMS -56-INR  ADAMS S
                 INFORMATION - A NATIONAL RESOURCE.
                 AMER DOC V. VII NO. 2 APR 1956
ADIAWC-55-CIS  ADIAR WC
                 CITATION INDEXES FOR SCIENCE.
                 AM. DOCUMENT. 6. 31 /1955/.
ADKIBW-  -DPL  ADKINSON BW
                 DATA PROCESSING AND LIBRARY OPERATIONAL PROBLEMS.
                 LIBRARY OF CONGRESS
ADKIBW-56-IUR  ADKINSON BW LIBRARY OF CONGRESS
                 INTERNATIONAL UTILIZATION OF RECORDED KNOWLEDGE.
                 CHAPT. VIII IN DOCUMENTATION IN ACTION REINHOLD PUB CORP
                   1956.
AEC RC-53-ECS  AEC REPORT CRO 102. UNIVERSITY OF TENNESSEE NOVEMBER 1953.
                 EDGE-PUNCHED CARDS FOR SCIENTIFIC LITERATURE REFERENCES.
                 AEC REPORT CRO 102. UNIVERSITY OF TENNESSEE NOVEMBER 1953.
AHLIJT-56-GUF  AHLIN JT
                 GENERAL USE OF FOUR-HOLE RANDOMLY PUNCHED CARDS IN FILE
                   SEARCHING APPLICATIONS.
                 PAM 1702 DEC. 1956 RESEARCH LIB. SAN JOSE CAL..
ALEXSN-57-DPI  ALEXANDER SN
               STEVENS ME
                 DATA PROCESSORS FOR INFORMATION RETRIEVAL PURPOSES.
                 PAPER PRESENTED AT THE 132TH MEETING OF THE AMER. CHEM.
                   SOCIETY. NEW YORK. SEPT. 10. 1957
ALLEEP-43-PCN  ALLEN EP
                 A PUNCHED CARD FOR NEOPLASTIC DISEASES.
                 NEW ZEALAND MED. J. 42 121 1943
ALLOAJ-  -IRM  ALLOT AJ
                 INFORMATION RETRIEVAL METHODS USED BY THE U S ARMY ORDNANCE
                   CORP IN DEPOT OPERATIONS.
                 DEPT. OF THE ARMY
AMDOC -53-CRM  ABSTRACT IN AM. DOC. APR. 1953.
                 CORRESPONDENCE REGARDING METALLURGICAL DOCUMENTATION OF THE
                   CORDONNIER-BATTEN SYSTEM OF PUNCHED CARDS.
                 ABSTRACT IN AM. DOC. APR. 1953.
AMDOC -53-FSE  ABSTRACT IN AM. DOC. JANUARY 1953.
                 FILMOREX SYSTEM FOR ELECTRONIC SELECTION OF MICROFILM CARDS.
                 ABSTRACT IN AM. DOC. JANUARY 1953.
AMERDI-56-RCR  AMER DOCUMENTATION INST. 1956-1957 1957 16 PP.
                 ROSTER OF CURRENT RESEARCH IN DOCUMENTATION AND
                   LIBRARIANSHIP.
                 AMER DOCUMENTATION INST. 1956-1957 1957 16 PP.
AMERDO-52-GCR  AMER. DOC. III 1952 91-94
                 THE GENESIS AND CHARACTERISTICS OF REPORT LITERATURE.
                 AMER. DOC. III 1952 91-94
AMERDO-52-PEM  AMER. DOC. III 1952 58-60
                 PRACTICAL EXPERIENCE IN MICROFACSIMILE PUBLICATION.
                 AMER. DOC. III 1952 58-60
AMERDO-54-PRM  AMER. DOC. JANUARY 1954.
                 PRELIMINARY REPORT ON THE MECHANICAL ANALYSIS OF INFORMATION
                   BY USE OF THE 101 STATISTICAL PUNCHED-CARD MACHINE.
                 AMER. DOC. JANUARY 1954.
AMERDO-57-ICS  AMERICAN DOCUMENTATION 8 /1/ 76-77 JAN.. 1957
                 INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION.
                 AMERICAN DOCUMENTATION 8 /1/ 76-77 JAN.. 1957
AMERDO-57-LSL  AMER. DOC. OCTOBER 1957.
                 LOGICAL STRUCTURES IN LANGUAGE.
                 AMER. DOC. OCTOBER 1957.
AMERDO-57-TLN  AMER. DOCUMENTATION V. VIII. NO. 1 JAN. 1957
                 THE TECHNICAL LIBRARY IN THE NEW RESEARCH LABORATORY OF THE
                   NATIONAL CARBON COMPANY.
                 AMER. DOCUMENTATION V. VIII. NO. 1 JAN. 1957
AMERFO-48-FSU  AMER. FOUNDRYMAN. P. 29 OCTOBER 1948.
                 A.F.S.USES PUNCHED CARD SYSTEM TO INDEX FOUNDRY LITERATURE.
                 AMER. FOUNDRYMAN. P. 29 OCTOBER 1948.
AMERIA-53-ISF  AMER. INST. OF ARCHITECTS. 1953 63P. WASHINGTON D.C.
                 A.I.A.STANDARD FILING SYSTEMS.
                 AMER. INST. OF ARCHITECTS. 1953 63P. WASHINGTON D.C.
AMERSC-54-CTB  AMER. SCI. 42. P. 248 APRIL 1954.
                 CAN TRANSLATION BE MECHANIZEDS.
                 AMER. SCI. 42. P. 248 APRIL 1954.
AMERSM-57-MIC  AMERICAN SOCIETY FOR METALS. 1957
                 THE METALS INFORMATION CENTER OF TOMORROW/16-MM COLOR SOUND
                   MOVIE/.
                 AMERICAN SOCIETY FOR METALS. 1957
AMES  -48-UPC  AMES
                 THE USE OF PUNCHED CARDS FOR INDEXING AND CLASSIFYING
                   BIO-CHEMICAL LITERATURE.
                 SPEC. LIB. 39. P. 233 1948.
AMSTAB-49-UPC  AMSTER AB ET AL
                 USE OF PUNCHED CARDS.
                 J. CHEM. EDUC. 26.* 304-6 1949
ANDRDD-  -AOO  ANDREWS DD
               NEWMAN SM
                 ACTIVITIES AND OBJECTIVES OF THE OFFICE OF RESEARCH AND
                   DEVELOPMENT IN THE U.S. PATENT OFFICE.
                 U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
ANDRDD-  -CIP  ANDREWS DD
                 COOPERATIVE INFORMATION PROCESSING - PATENTS.
                 U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
ANDRDD-  -ILA  ANDREWS DD
                 INTERRELATED LOGIC ACCUMULATING SCANNER.
                 U.S. PATENT OFF.
ANDRDD-  -LPO  ANDREWS DD
                 THE LAW AND PATENT OFFICE RESEARCH.
                 U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
ANDRDD-  -MFG  ANDREWS DD
                 MAJOR FACTORS GOVERNING THE SELECTION OF MACHINE DATA
                   HANDLING TECHNIQUE.
                 U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
ANDRDD-  -PEP  ANDREWS DD
                 PATENT EXAMINING PAST. PRESENT AND FUTURE.
                 U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
ANDRDD-  -PRU  ANDREWS DD
                 PROGRESS REPORT ON U.S. PATENT OFFICE MECHANIZED SEARCHING.
                 U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
ANDRJL-  -DEP  ANDRIOT JL
                 DOCUMENTS EXPEDITING PROJECT.
                 LIBRARY JOURNAL. LXXVII. 693-695.
APPLPL-52-SHL  APPLIED PHYSICS LAB. JOHNS HOPKINS UNIV. NOV. 1952
                 SUBJECT HEADING LIST FOR GUIDED MISSILES.
                 APPLIED PHYSICS LAB. JOHNS HOPKINS UNIV. NOV. 1952
ARCECJ-56-MPT  ARCENEAUX CJ /ETHYL CORP. BATON ROUGE/
                 MOVIES PRESENT TECHNICAL PAPERS.
                 CHEM. & ENG. NEWS 34. 627-9 FEB. 1956
ARMEFM-56-PBM  THE ARMED FORCES MEDICAL LIBRARY
                 PROPOSED BIBLIOGRAPHY OF MEDICAL CONGRESS PUBLICATIONS.
                 UNESCO BULL. FOR LIBRARIES. 10. 22. HAN JAN. 1956
ARMEST-57-MAS  ARMED SERVICES TECHNICAL INFORMATION AGENCY WASHINGTON FEB.
                 1957.
                 MULTIPLE ASPECT SEARCHING FOR INFORMATION RETRIEVAL.
                 ARMED SERVICES TECHNICAL INFORMATION AGENCY WASHINGTON FEB.
                   1957.
ARMYAF-55-N5V  ARMY-AIR FORCE REGISTER. 77. 8. 15 OCT. 1955
                 NEW 50-LANGUAGE VARI-TYPER DEVELOPED.
                 ARMY-AIR FORCE REGISTER. 77. 8. 15 OCT. 1955
ARNHAA-47-NCL  ARNHYM AA
                 A NEW CLASSIFICATION.* LIMITED DISTRIBUTION.
                 TECHNICAL DATA DIGEST XII SEPT 1. 1947 2-3
ARNHF -57-PCA  ARNHOLDT F
                 PUNCHED CARDS AS AN AID TO INDEXING UROLOGICAL REPORTS.
                 NACHRICHTEN DOKUMENTATION 8/2/ 82-85 JUNE 1957
ASBUWC-57-MVT  ASBURY WC
               REEVES ED
               KNOX WT
                 MANAGEMENTS VIEWS ON TECHNICAL - INFORMATION ACTIVITIES.
                 PRESENTED BEFORE THE AMERICAN CHEMICAL SOCIETY SEPT 10. 1957
ASHTHD-52-PIE  ASHTHORPE HD
                 THE PUNCHED-CARD INDEXING EXPERIMENT AT THE LIBRARY OF THE
                   ATOMIC ENERGY RESEARCH ESTABLISHMENT. HARWELL.
                 ASLIB PROCEEDINGS V. 4. NO. 2 MAY 1952
ASLIB -57-PIS  DORKING. ENGLAND. MAY 13-17. 1957 LONDON.* ASLIB. 1957
                 PROCESSINGS OF THE INTERNATIONAL STUDY CONFERENCE ON
                   CLASSIFICATION FOR INFORMATION RETRIEVAL.
                 DORKING. ENGLAND. MAY 13-17. 1957 LONDON.* ASLIB. 1957
ASLIPR-57-IMA  ASLIB PROCEEDINGS 9 /10/ 301-309.. DISCUSSION 309-314.
                 OCT. 1957
                 INFORMATION METHODS. ADAPTATION TO GROWTH.
                 ASLIB PROCEEDINGS 9 /10/ 301-309.. DISCUSSION 309-314.
                   OCT. 1957
ASTRMM-58-RLM  ASTRAHAN MM
                 THE ROLE OF LARGE MEMORIES IN SCIENTIFIC COMMUNICATIONS.
                 IBM JOURNAL. OCTOBER. 1958
ATANJV-36-APE  ATANASOFF JV
               BRANDT AE
                 APPLICATION OF PUNCHED-CARD EQUIPMENT TO THE ANALYSIS OF
                   COMPLEX SPECTRA.
                 J. OPT. SOC. AMER. V. 26. FEB. 1936. PP. 83-88
AUTOCO-56-PIS  AUTOMATIC CONTROL V. 5. JULY 1956 P. 4
                 PHOTOSCOPIC INFORMATION STORAGE UNIT.
                 AUTOMATIC CONTROL V. 5. JULY 1956 P. 4
AUTOCO-58-AOP  AUTOMATIC CONTROL. V. 7. NO. 1. JAN. 1958 6-10
                 AUTOMATIC OPERATIONS IN POSTAL PROCESSING.* CONTROL ELEMENTS
                   SPEED SORTING FUNCTION.
                 AUTOMATIC CONTROL. V. 7. NO. 1. JAN. 1958 6-10
AVAKE -57-AAM  AVAKIAN E
                 AMFIS--THE AUTOMATIC MICROFILM INFORMATION SYSTEM.
                 SPEC.LIB. /APR.1957/
AVIAJ -52-IOS  AVIAS J
                 INTERNATIONAL ORGANIZATION OF SCIENTIFIC DOCUMENTATION BASE
                   ON LEGISLATION.
                 SCIENCE. 115. 250 /1952/.
AYERJA-48-LTL  AYER JA
                 LANGUAGE. TRUTH AND LOGIC.
                 VICTOR GOLLANCZ. LONDON. 1948. PP. 33-44. 87-102.
BABAS -56-SRD  BABA S
                 SOME RECENT DEVELOPMENTS IN JAPANESE DOCUMENTATION.
                 UNESCO REF. NS 94/LX20 1956 13 P.
BABEP -56-OMT  BABEL P
                 THE OUTLOOK FOR MECHANICAL TRANSLATION.
                 BABEL. P. 99 OCTOBER 1956.
BACHH -55-NIA  BACH H
                 THE NEED FOR AN INSTITUTE FOR ADVANCED LIBRARY RESEARCH.
                 WILSON LIBRARY BULLETIN.29.709-711.MAY 1955
BACHRW-52-SVM  BACHELDER RW
                 THE SCOPE AND VALUE OF THE MICROCARD.
                 SPEC. LIB. XLIII 1952 157-61
BAGEP -51-EDM  BAGELEY P
                 ELECTRONIC DIGITAL MACHINES FOR HIGH SPEED INFORMATION
                   ON SEARCHING.
                 MASSACHUSETTS INSTITUTE OF TECHNOLOGY /NOV. 1951/ /MASTER-S
                   THESIS. AUG. 1951
BAGLPR-51-ANE  BAGLEY PR
               PERRY JW
                 APPLICATION OF NEWER ELECTRONIC TECHNIQUES TO INFORMATION
                   SEARCHING.
                 PAPER PRESENTED BEFORE DIVISION OF CHEMICAL LITERATURE OF
                   THE AMERICAN CHEMICAL SOCIETY AT THE DIAMOND JUBILEE. N.Y.
                   CITY. SEPT. 1951.
BAGLPR-52-ANE  BAGLEY PR
                 APPLICABILITY OF NEWER ELECTRONIC TECHNIQUES TO INFORMATION
                   SEARCHING.
                 MACHINE TECHNIQUES FOR INFORMATION SELECTION JUNE 1952
BAHMRH-55-NAA  BAHMER RH
                 THE NATIONAL ARCHIVES AFTER 20 YEARS.
                 AMERICAN ARCHIVIST.18.195-205. JULY 1955
BAILBS-55-BAC  BAILEY BROS. & SWINFEN LTD. LONDON GILES HIGH ST. 1955
                 BAILEY-S AFRICIAN-ORIENTAL CATALOGUE OF DICTIONARIES.
                   GRAMMARS AND PHRASEBOOKS.
                 BAILEY BROS. & SWINFEN LTD. LONDON GILES HIGH ST. 1955
BAILCF-46-PCI  BAILEY CF
               CASEY RS
               COX GJ
                 PUNCHED CARDS FOR INDEXING SCIENTIFIC DATA.
                 SCIENCE 104.* 181. 1946
BAILM -53-MSU  BAILEY M
                 MECHANIZED SEARCHING IN THE US.PATENT OFFICE.
                 JOUR.PATENT OFFICE SOCIETY. P. 566 AUG.1953
```

FIG. 6. Sample bibliography page from H.P. Luhn & P. James (1959), *Literature on Information Retrieval and Machine Translation: Bibliography and Index; Titles Indexed by Key-Words-in-Context System* (2nd ed.). Yorktown Heights, NY: IBM Research Center, June 1959, p. 1. Combination of letters and numbers to the left of cited item refers to items in the bibliography. Reproduced from a copy of the original.

translation (Luhn & James, September 1958), which he distributed at the conference. Stevens attributed to Luhn two innovations at the ICSI conference:

> True to his promise and his typical intent of reduction-to-practice, he was responsible for not one but two experimental demonstrations in connection with the ICSI papers: first, the application of his auto-abstracting technique to the papers for one session [Area 5], and secondly, the production of a KWIC index to all the ICSI papers. (Stevens, 1968, p. 29)

Stevens used as her evidence for these two innovations a footnote in a 1962 paper by Luhn (p. 100); however, close inspection of that footnote shows that it discusses only the

DOI: 10.1002/asi

production of the auto-abstracts and not the production of a KWIC index to ICSI preprints. Indeed, it is likely that Luhn was much more interested in the possibilities of automatic abstracting in the period of about 1957 to 1960 than he was with automatic indexing. The chronological list of his publications in Schultz (1968) indicates this to some extent.

As a member of ICSI Program Committee, Luhn was directly involved in the production of both the preprints and the final proceedings and, according to Stevens, was purposefully intent on preparing punched tapes of the papers for future statistical studies (Stevens, 1968, p. 29). Thus, it makes sense that since Luhn already had produced a KWIC index for the literature of information retrieval and machine translation, that he also might produce a KWIC index to the titles of the ICSI preprints. Schultz (1968) implied that Luhn *intended* to produce such an index: "The new Luhn . . . KWIC indexing technique was to be introduced and demonstrated by means of an index to the titles of the many papers at the ICSI meeting" (p. 11). Unfortunately, if Luhn did indeed produce a KWIC index to the ICSI preprints, no copy has survived. Two attendees at the ICSI, Charles Bourne and Ben-Ami Lipetz, do not recall an ICSI preprints KWIC index being handed out by IBM or Luhn but do clearly remember the one distributed by Ohlman (Bourne, 2009, personal communication, July 1 and July 29, 2009; Lipetz, 2009, personal communication, June 15, 2009). Similarly, Becker and Hayes' (1966, p. 143) textbook on information storage and retrieval cited only the ICSI index by Ohlman and his colleagues.

We do know, however, that the ICSI papers preprints were prepared on Monotype tapes by Columbia University Press. The conversion of these tapes, via punched cards, to magnetic tapes and then, via preset computer programs, to automatic printouts of the Area 5 abstracts was demonstrated at the ICSI conference by Luhn and his IBM colleagues (Luhn, 1958c; Luhn, 1962, p. 100; Savage, 1958). Luhn presented a somewhat detailed paper at the 1959 meeting of the ADI on his KWIC indexing method, which was later published in *American Documentation* (Luhn, 1960). No acknowledgment is made in that article of Ohlman's work on permutation indexing.

As noted earlier, Luhn's KWIC system had an almost immediate impact. Chemical Abstracts Service (CAS) began using it to produce an index to new chemical publications. The first five monthly trial versions were published in 1960, and in 1961, biweekly issues began to be issued. *Chemical Titles* became the first periodical to be organized, indexed, and composed almost completely by computer. Robert R. Freeman and G. Malcolm Dyson, of CAS, described the implementation of KWIC, using an IBM 704 computer in a 1963 article (Freeman & Dyson, 1963). CAS followed this a year later with the KWIC-based *Chemical Biological Activities*, a biweekly publication. Biological Abstracts, Inc. soon issued their KWIC version of *BASIC*, a semimonthly publication. By 1965, more than 30 different indexing publications were using KWIC as their indexing tool (Stevens, 1965, pp. 49–53).

Almost simultaneously with his public notice about KWIC, Luhn announced the development of SDI, a method for selectively disseminating information that was relevant to one or more specific users. It was announced as a trial for the IBM Yorktown Heights Library in May 1958, with the first full description in a November 1959 report (Luhn, 1959). When SDI and KWIC were used in concert, it was possible for an information system, including an information specialist running an information center, to quickly provide users with an indexed list of new publications and, using a machine-based profile of their interests, disseminate individualized items of information. Luhn visualized the two systems working together to create what he called "intelligence systems" (Stevens, 1968, p. 28). SDI was quickly adopted by many different types of organizations, but especially by special libraries in science and industry (Bourne & Hahn, 2003).

Luhn and Ohlman and KWIC and Permuterm: Who Was First?

According to the Luhn chronology reported by Stevens (1968), Luhn had the beginnings of his KWIC approach by May 1957, and sometime between then and early 1958, he had developed the idea of KWIC. However, his first publication using KWIC was the September 1958 bibliography (Luhn & James, 1958) on information retrieval and machine translation, based on an earlier bibliography by Bourne (1958). Exactly when Ohlman began actively working on his permutation indexing approach is unknown, but sometime between June and September 1957, he had produced, as an internal SDC document, his first "Permutation Index No. 1" to the SAGE programming documents. By September 25, 1957, he had produced an SDC internal document titled "Introductory Manual for Permutation Indexing" (McCafferty & Ohlman, 1957, as cited in Ohlman Papers, 1955–2002). Ohlman also prepared, but apparently never published, a paper on permutation indexing by September 19, 1957 (Ohlman, 1957a) and a second paper on November 5, 1957, also apparently unpublished, on the use of electronic accounting machinery (EAM) for doing permutation indexing (Ohlman, 1957b).

Thus, in terms of the originality of the "invention idea," the two are almost completely simultaneous. Ohlman, however, produced the first actual automatic permutation index and an instruction manual for it in mid-to-late 1957. Luhn, however, has the first *publicly* available product with his September 1958 KWIC index to information retrieval and machine translation literature. Both Ohlman and Luhn used the keyword in context approach. In his first two internal permutation indexes to the SAGE documents, Ohlman used a left-justified keyword approach, but by the time of his ICSI reprints index, he had switched the keyword to the center of the page. Luhn's September 1958 index centered the keywords in two parallel columns. Both Ohlman and Luhn used an approach similar to an accession numbering system to reference the items being indexed and produce a bibliography of the citations. As noted earlier, Ohlman enriched his mostly title permutation index with a variety of nontitle words (e.g., authors, terms from abstracts, etc.), and Luhn did the same with some of his later products but generally resisted much additional information for the process to be rapidly keypunched without much analysis of content. Both used a "stop," or "kill," list, to eliminate common and uninformative words. Luhn is sometimes credited with coining the term "stop list" (Wikipedia, 2009?), but neither his first KWIC index (Luhn & James, 1958) nor his article in *American Documentation* (Luhn, 1960) on KWIC contains the phrase. In fact, both Luhn and Ohlman used very similar language in discussing the elimination of certain words in their automatic indexes. Luhn used "non-significant or common words" (Luhn, 1960, p. 289), and Ohlman and colleagues used "non-significant words" (Citron et al., 1958). The coiner(s) of the terms "stop" or "kill" list may never be known. Bourne (1963, pp. 14–15) implied that the practice itself was well-known in early information retrieval circles, though he also did not use either term in his book. Weinberg (2004) showed that the concept was used as early as the 1400s.

Doyle described Ohlman and colleagues in late 1958 as being in a race to "scoop IBM," but it is not clear what they knew about IBM's intentions at ICSI. Were they trying to "scoop" an ICSI papers index or were they trying to "scoop" the idea of KWIC? It is possible that Ohlman and colleagues knew about the September 1958 KWIC bibliography by Luhn and James and wanted to show what another system, their permutation index, could do. It also is possible that they had heard that Luhn and his colleagues were intending to produce a KWIC index to the ICSI preprints. Unfortunately, we may never know because no ICSI archives exist and because the IBM archives do not provide adequate information on Luhn's activities in this period.

Permuterm and KWIC: How Different When Implemented?

This is a difficult question to answer because of the dominance of KWIC indexing systems over permutation indexing systems in actual use in the years after their introduction. It becomes even more difficult to answer because of the variations in the systems as they were put into use by the several indexing services. As noted earlier, Stevens' 1965 report listed more than 30 different indexing publications using, as she titled her table, "KWIC type indexes and programs." Most of these were briefly described using the phrase "standard Luhn/IBM," but some indicated that the developers had modified their systems and cited other published reports, though none cited Ohlman (Stevens, 1965, pp. 49–53). By the time Stevens wrote her report in 1965, and certainly by 1973 when Hilda Feinberg published her excellent study of title derivative indexing techniques, index developers were not careful to distinguish among cycled, rotated, and permuted indexing systems and the differences produced by them. The four-volume, 3,114-item "Bibliography on Information Science and Technology for 1968" (Neeland & Sever, July 1968) performed under a contract with the ASIS by the SDC, where one might expect use of Ohlman's permutation indexing approach, made no mention of the intellectual origins of the extensive permutation indexes included in them. Similarly, Hawkins (1983) "Online Information Retrieval Bibliography, 1964–1982," which contains an extensive permutation index, only noted it was ". . . produced and formatted by the Bell Laboratories proprietary program BELDEX and printed on a laser printer" (Introduction).

Thus, direct comparisons or side-by-side evaluations between Ohlman's permutation indexing system and KWIC have not been made in the literature. The nearest discussion is Eugene Garfield's (1976) article on the origins of the Institute for Scientific Information's *Permuterm Subject Index*. As noted earlier, Garfield credited Ohlman with the idea of permutation indexing, but his critique of why he used his basic idea and did not use KWIC cannot be considered an evaluation of the two different systems. His principal objections to KWIC are: uneconomical for a printed index; indiscriminate use of stoplists; and, most important, KWIC does not permit the coordination of terms, a feature that Garfield says users really like.

Why did KWIC Succeed and Permutation Fail?

There is no denying that the phrase *keyword-in-context indexing* and its acronym are much more descriptive and catchy than is permutation or rotated indexing, even though both terms more accurately describe how this method of automatic indexing actually works. In addition, Luhn and IBM had at the time much greater name recognition than did Ohlman. Perhaps even more significantly, Luhn had the IBM publicity "machine" behind him. Not only did Luhn arrange for IBM to have special demonstrations at the ICSI Conference but several IBM press releases over the next year touted the development of KWIC and automatic abstracting. Bourne (2009, personal communication, July 1 and July 29, 2009) also noted that KWIC and other IBM products were repeatedly promoted by IBM sales agents, IBM seminars, and the many reports issued by the company. The national press, popular and scientific, was quick to pick up these and publicity doubled when Chemical Abstracts Service begins using KWIC to index its new *Chemical Titles*. Dale Baker's quote about KWIC being the "best thing to happen to chemistry since the test-tube" was used in his Congressional testimony and aroused even more publicity (Schulz, 1968, pp. 11–13; Stevens, 1968, pp. 28–29). Luhn also continued to have a very large research and development group to continue work on automatic abstracting, SDI, KWIC, and auto-encoding of literature.

Ohlman, however, did not have such advantages. His department at SDC was only minimally interested in information retrieval applications at this time and did not do any kind of promotion of permutation indexing or encourage expansion of the work done by Ohlman and his colleagues. In addition, Ohlman's efforts to create his own company to advance work on permutation indexing failed because of a lack of adequate capital. His team of developers was

KWIC KeyWord In Context Each word that is not a stop-word becomes an entry word (aka lead term). Entry words are aligned within the page:

for Croations. **Cataloging** and classification
Cataloging and **classification** for Croations
for **Croations**. Cataloging and classification

KWAC KeyWord Alongside Context. As with KWIC but with entry words justified at left:

Cataloging and classification for Croatians.
classification for Croatians. Cataloging and
Croatians. Cataloging and classification for

KWOC KeyWord Out of Context. Entry at left (or above). Context not wrapped around the entry word:

Cataloging Cataloging and classification for Croatians.
classification Cataloging and classification for Croatians.
Croatians. Cataloging and classification for Croatians

Source: M. Buckland, SIMS 245, Organization of Information, Spring, 2003. Retrieved Sept. 24, 2009 from: http://www2.sims.berkeley.edu/courses/is245/s03/verbal.html . Used by permission.

FIG. 7. Simple explanations and examples of KWIC, KWAC, and KWOC automatic indexing.

very small, probably never more than four. With no internal support at SDC for permutation indexing, he soon joined another group that was exploring medical automation (Doyle, 1966). Ohlman did continue thinking about permutation automatic indexing, presenting a paper at the 1960 ADI meeting (Ohlman, 1960; Ohlman Papers, 1955–2002) that was never published. He also expended considerable time exploring the costs, advantages, and disadvantages of automatic indexing, including corresponding with Luhn and other IBM staff regarding techniques used, equipment choices, and whether automatic indexing was superior to manual indexing (Ohlman Papers, 1960–1961). None of this led to any additional published work, and his various professional activities after this did not involve automatic indexing.

Is Automatic Indexing Really Indexing?

By the early 1970s, the excitement over KWIC indexing was beginning to wane. Computers were now much more powerful, much less expensive, and, especially, had developed much more memory capacity, permitting large chunks of literary textual information to be held in temporary or permanent storage. Improved indexing studies were beginning to raise serious questions about extractive indexing systems, such as KWIC and permutation indexing, as replacements for human, or manual, indexing. Even more significantly, large databases, some online (both in-house and publicly available), containing not only titles and abstracts but sometimes full text, were now available and could be searched using Boolean logic with enhanced features such as truncation and adjacency and an online thesaurus.

While KWIC had some useful features when used in online systems (Bourne & Hahn, 2003, pp. 44, 93, 135), it was beginning to be acknowledged that its best use was just what Luhn and Olhman had initially proposed: as an announcement vehicle. It was, as the common descriptive phrase of the time had it: "KWIC and dirty:" easy to create, whether with a computer or EAM equipment, inexpensive, and not to be confused with real indexing. Feinberg (1973, pp. 32–34) was a bit more optimistic, likely reflecting the time in which she was writing, and had a fairly comprehensive listing of the advantages and disadvantages of "permuted title indexes," (see Figure 7) a term she used to be comprehensive of all the automated title derivative indexed approaches. In summary, the advantages she listed were: promptness of production; little human involvement; computer generated for few human errors; more information, in context, may be displayed in an index; author supplied terms; many access points; multiterm coordination is easily accomplished; lots of programs available on how to do it; multiple products, such as an SDI service, are possible; cumulation is easy; any literature is amenable to the technique; and very inexpensive when compared to human indexing.

Feinberg's (1973) major disadvantages were: titles are not informative; little or no terminology control; loss of retrieval power in a collection where the same term is in frequent use; many irrelevant entries; line length restrictions limit understanding; printing programs often produce poor graphic quality; and double look-up is almost always necessary. Wellisch (1996), however, also likely reflecting the time in which he was writing, thought that the days of KWIC/permutation were done. Discussing the advantages and disadvantages of KWIC and permutation indexing as well as other types of automatic indexing, he stated that as retrieval aids they are ". . . rough and ready substitutes for the real thing because they do not need skilled personnel . . . ." (p. 252). "Many retrieval tests," he noted, "have shown that searching for keywords in the natural language text of a document, combined with the use of assigned keywords from a controlled indexing language, is the most effective way of information retrieval from bibliographic databases" (p. 253).

KWIC/Permutation Indexing Today

Despite Wellisch's (1996) estimation of the value of KWIC/permutation indexing as indexing, the basic idea continues to play a role in many aspects of information retrieval, thesaurus construction, and information extraction techniques, and even in Web-based search engines

DOI: 10.1002/asi

(Lancaster, 1998, 2003). The basic idea of extracting keywords from a mass of text (whether in titles or in full text) and linking them in some way to the entire document, or parts of the document, is simply too good and too inexpensive to go away. David L. Parnas (1972) is usually attributed with first suggesting the technique as a way of utilizing it to compare different approaches to dividing systems into modules in programming. His technique was later used by the many developers of UNIX for a variety of purposes, including as a sentence retrieval tool in any language and in designing software architecture (An excellent, though technical, explanation of the use of KWIC in software design appears in Garlan & Shaw, 1993.) KWIC applications for thesaurus construction, analyzing text in most word-processing programs or any Windows-based file or a downloaded Adobe Acrobat file, and a variety of other purposes are easily found on the Internet—using any of the search engines that use a KWIC formatting approach! Lexis-Nexis trademarked KWIC when used in connection with their search and document display system (Holmes, 2001; Lexis-Nexis, 2004). The work of Luhn or Ohlman is almost never mentioned when these applications are discussed or promoted.

Conclusions

There is little doubt that KWIC/permutation indexing had a dramatic impact on the early development of information retrieval research and implementation. It was a tremendous boon to the publishers of indexing journals who were trying to keep up with the growth of the scientific literature in the 1950s and early 1960s. Special librarians loved it because it was of tremendous help in their efforts to keep their users aware of current report literature. It also was of great value to the users of these journals because it significantly increased how rapidly they learned of new scientific work and because the use of keywords in context simplified their scanning of this work. It was not great indexing—some would say not indexing at all—but it was fast, and many scientific fields responded with the most sincere form of approval: imitation, adoption, innovation, and improvement. It also was an almost-perfect technological development for its time since it could use existing EAM technology if it had to and was superbly adaptable to the new digital computers, as soon as you could afford one. Researchers saw then, and still see today, all kinds of questions they wanted to study in this new development, and the system developers had lots of room for innovation.

For the two main "inventors" of permutation/KWIC indexing, Hans Peter Luhn and Herbert M. Ohlman, their stories of success were remarkably different even though their inventions were very similar and almost simultaneous. Luhn, with IBM solidly behind him on KWIC and his related developments, had all the advantages that led to success and fame. Ohlman, the energetic visionary, was not able to garner the kind of institutional and financial support required for a similar type of success and fame. Perhaps their personalities were a factor in what happened after 1958, but all reports from their coworkers have indicated that both were innovative, enthusiastic, personable, and had solid leadership abilities.

This historical analysis of the beginnings of automatic indexing was not done to "set the record straight," though I hope it does help a bit in that direction. It was written to illustrate the rich history of our field, particularly as it played out in the area of automatic indexing—and the lives of two of our important pioneers—in the 1950s and early 1960s, and to show how that history continues to influence information science today.

Acknowledgment

Thanks to the following persons who assisted in the development of the article, reviewed an early draft, and provided additional information: Charles P. Bourne, Michael Buckland, Colin Burke, Eugene Garfield, and Martha Jane Zachert. Thanks also to the referees who provided valuable suggestions for improvement.

References

Anderson, J.D. (1994). Standards for indexing: Revising the American National Standard Guidelines Z39.4. Journal of the American Society for Information Science, 45(8), 628–636.

Baxendale, P.B. (1958). Machine-made index for technical literature—An experiment. IBM Journal of Research and Development, 2(October), 354–361.

Becker, J., & Hayes, R.M. (1966). Information storage and retrieval: Tools, elements, theories. New York: Wiley.

Bourne, C.P. (1958). Bibliography on the mechanization of information retrieval. Menlo Park, CA: Stanford Research Institute.

Bourne, C.P. (1963). Methods of information handling. New York: Wiley.

Bourne, C.P., & Hahn, T.B. (2003). A history of online information services, 1963–1976. Cambridge, MA: MIT Press.

Burke, C. (1994). Information and secrecy: Vannevar Bush, Ultra, and the Other Memex. Metuchen, NJ: Scarecrow Press.

Chu, H. (2007). Information representation and retrieval in the Digital Age. Medford, NJ: Information Today, Inc. for the American Society for Information Science and Technology.

Citron, J., Hart, L., & Ohlman, H. (1958). A permutation index to the preprints of the International Conference on Scientific Information. Santa Monica, CA: System Development Corporation.

Citron, J., Hart, L., & Ohlman, H. (1959). A permutation index to the preprints of the International Conference on Scientific Information (Rev. ed., December 15). Santa Monica, CA: System Development Corporation, SP-44.

Doyle, L.B. (1966). The story of language processing and retrieval at SDC. Santa Monica, CA: System Development Corporation, Internal Document No. N-23059.

Feinberg, H. (1973). Title derivative indexing techniques: A comparative study. Metuchen, NJ: Scarecrow Press.

Fischer, M. (1966). The KWIC index concept: A retrospective view. American Documentation, 17(2), 57–70.

Freeman, R.R., & Dyson, G.M. (1963). Development and production of chemical titles, a current awareness index publication prepared with the aid of a computer. Journal of Chemical Documentation, 3, 16–20.

Garfield, E. (1976). The Permuterm Subject Index: An autobiographical review. Journal of the American Society for Information Science, 27(5), 288–291.

Garlan, D., & Shaw, M. (1993). An introduction to software architecture. In V. Ambriola & G. Tortora (Eds.), Advances in software engineering and knowledge engineering (Vol. 1, pp. 18–23). Camden, NJ: World Scientific.

DOI: 10.1002/asi

Available at: http://www.cc.gatech.edu/classes/AY2009/cs6300_fall/readings/arch.pdf
Hawkins, D.T. (1983). Online Information Retrieval Bibliography, 1964–1982. Medford, NJ: Learned Information, Inc.
Holmes, N. (2001). The KWIC and the dead: A lesson in computing history. Computer, 34(1), 144–147.
Lancaster, F.W. (1998). Indexing and abstracting in theory and practice (2nd ed.). Champaign: University of Illinois, Graduate School of Library and Information Science.
Lancaster, F.W. (2003). Indexing and abstracting in theory and practice (3rd ed.). London: Facet.
Lexis-Nexis, Inc. (2004). At a glance: How to print [Online Instruction Manual]. Available at: http://www.lexisnexis.com/firstyearassociates/usinglexisnexis/pubs/HowToPrint.pdf
Luhn, H.P. (1957). A statistical approach to mechanized encoding and searching of literary information. IBM Journal of Research and Development, 1(4), 309–317.
Luhn, H.P. (1958a). The automatic creation of literature abstracts. IBM Journal of Research and Development, 2(2), 159–165.
Luhn, H.P. (1958b). Indexing, language, and meaning (pp. 208–218). In M. Taube & H. Wooster (Eds.), Information storage and retrieval. New York: Columbia University Press.
Luhn, H.P. (1958c). An experiment in auto-abstracting: Auto-abstracts of Area 5 Conference Papers, International Conference on Scientific Information, Washington, DC. Yorktown Heights, NY: Information Retrieval Research Department, IBM Research Center. (Reprinted in C. K. Schultz (Ed.). (1968). H. P. Luhn: Pioneer of information science (pp. 145–163). New York: Spartan Books).
Luhn, H.P. (1959). Selective dissemination of new scientific information with the aid of electronic processing equipment. Yorktown Heights, NY: IBM Corporation, Advanced Systems Development Division.
Luhn, H.P. (1960). Keyword-in-Context Index for technical literature. American Documentation, 11(4), 288–295.
Luhn, H.P. (1962). Automated intelligence systems—some basic problems and prerequisites for their solution. In L.H. Hattery & E.M. McCormick (Eds.), Information retrieval management (pp. 92–100). Detroit, MI: American Data Processing, Inc.
Luhn, H.P., & James, P. (1958). Literature on information retrieval and machine translation: Bibliography and index; titles indexed by key-words-in-context system. Yorktown Heights, NY: IBM Research Center.
Mooers, C.N. (1951). Zatacoding applied to mechanical organization of knowledge. American Documentation, 2(1), 20–32.
Moore, H. (1968). H.P. Luhn, engineer. In C.K. Schultz (Ed.), H.P. Luhn: Pioneer of information science: Selected works (pp. 16–20). New York: Spartan Books.
Neeland, F., & Sever, E. (1968). Bibliography of information science and technology for 1968. Parts 1–4. Santa Monica, CA: System Development Corp. for the American Society for Information Science.
Ohlman, H. (1957a). Limitations of peek-a-book and a new retrieval system: Permutation indexing. Unpublished paper [September 19, 1957]. Copy in Ohlman Papers, Babbage Institute.
Ohlman, H. (1957b). Permutation indexing: Multiple-entry listing by electronic accounting machines. Unpublished paper [November 5, 1957]. Copy in Ohlman Papers, Babbage Institute.
Ohlman, H. (1960). Mechanical indexing: Historical development, techniques, and critique (Technical Rep. No. 6002158). Washington, DC: National Bureau of Standards. Copy in Ohlman Papers, Babbage Institute.
Ohlman, H. (1961?). The early history of mechanical indexing, 1856–1959: A chronologically extracted bibliography with emphasis on permutation indexing (Suppl. to Technical Rep. No. 6002158). Washington, DC: National Bureau of Standards. Copy in Ohlman Papers, Babbage Institute.
Ohlman, H. (1999). Mechanical indexing: A personal remembrance. In M.E. Bowden, T.B. Hahn, & R.V. Williams (Eds.), Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems (pp. 187–194). Medford, NJ: Information Today, Inc. for the American Society for Information Science and the Chemical Heritage Foundation.
Ohlman, H., Hart, L., & Doyle, L. (1958). System Development Corporation (p. 34). Current Research and Development in Scientific Documentation, No. 3. Washington, DC: National Science Foundation [October, 1958].
Ohlman, H., Hart, L., & Doyle, L. (1959). System Development Corporation (p. 28). Current Research and Development in Scientific Documentation, No. 4. Washington, DC: National Science Foundation [April, 1959].
Ohlman Papers (1955–2002). CBI 179. 1.6 cu. ft. (A detailed finding aid to the collection is available online at the Babbage Institute Web site.) Retrieved July 27, 2009, from http://discover.lib.umn.edu/cgi/f/findaid/findaid-idx?c=umfa;cc=umfa;rgn=main;view=text;didno=cbi00179
Parnas, D.L. (1972). On the criteria to be used in decomposing systems into modules. Communications of the ACM, 15(12), 1053–1058.
Rocketdyne Corp. (1958). Electronic permutation indexing exhibit. Internal document [not paged]. Copy in Ohlman Papers, Babbage Institute.
Salton, G., & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.
Savage, T.R. (1958). The preparation of auto-abstracts on the IBM 704 Data Processing System. Yorktown Heights, NY: IBM [November 17, 1958].
Schultz, C.K. (1968). H.P. Luhn, the man. In C.K. Schultz (Ed.), H.P. Luhn: Pioneer of information science: Selected works (pp. 3–15). New York: Spartan Books.
Stevens, M.E. (1965). Automatic indexing: A state-of-the-art report. Washington, DC: National Bureau of Standards, Monograph No. 91 [for sale by the Superintendent of Documents].
Stevens, M.E. (1968). H.P. Luhn, information scientist. In C.K. Schultz (Ed.), H.P. Luhn: Pioneer of information science, selected works (pp. 24–34). New York: Spartan Books.
Veilleux, M.P. (1963). Permuted title word indexing procedures for a man/machine system. In Machine indexing: Progress and problems: Papers presented at the 3rd meeting of the Institute on Information Storage and Retrieval (pp. 77–111). Washington, DC: American University, School of Government and Public Administration, Center for Technology and Administration.
Weinberg, B.H. (1981). Word frequency and automatic indexing. Unpublished doctoral dissertation, Columbia University.
Weinberg, B.H. (2001). Index structures in the early Hebrew Biblical word lists: Preludes to the first Latin concordances. The Indexer, 22(4), 178–186.
Weinberg, B.H. (2004). Predecessors of scientific indexing structures in the domain of religion. In W.B. Rayward & M.E. Bowden (Eds.), The history and heritage of scientific and technological information systems: Proceedings of the 2002 Conference (pp. 126–136). Medford, NJ: Information Today, Inc. for the American Society for Information Science and Technology and the Chemical Heritage Foundation.
Wellisch, H.H. (1996). Indexing from A to Z (2nd ed., revised and enlarged). New York: Wilson.
Wikipedia. (2009?). Information retrieval. Retrieved September 23, 2009, from http://en.wikipedia.org/wiki/Information_retrieval
Williams, R.V. (2002). The use of punched cards in US Libraries and Documentation Centers, 1936–1965. IEEE Annals of the History of Computing, 24(2), 16–33.