

AD734388



ONR LONDON CONFERENCE REPORT

C-28-71

OFFICE OF NAVAL RESEARCH

BRANCH
OFFICE
LONDON
ENGLAND

CRANFIELD CONFERENCE ON INFORMATION RETRIEVAL

FRANKLIN F. KUO

10 NOVEMBER 1971

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

D D C
RECEIVED
JAN 8 1972
D

UNITED STATES OF AMERICA

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author)		2a. REPORT SECURITY CLASSIFICATION
Office of Naval Research Branch Office London, England		UNCLASSIFIED
		2b. GROUP
3. REPORT TITLE		
CRANFIELD CONFERENCE ON INFORMATION RETRIEVAL		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)		
N.A.		
5. AUTHOR(S) (Last name, first name, initial)		
KUO, FRANKLIN F.		
6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
10 NOVEMBER 1971	8	None
8a. CONTRACT OR GRANT NO.		8b. ORIGINATOR'S REPORT NUMBER(S)
a. PROJECT NO. N.A.		ONRL-C-28-71
c. N.A.		8d. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.		N.A.
10. AVAILABILITY/LIMITATION NOTICES		
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED		
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY
N.A.		N.A.
13. ABSTRACT		
<p>The Third Cranfield Conference on Mechanised Information Storage and Retrieval Systems was held on 20-23 July 1971 in Cranfield, England. The report describes a number of the key papers presented at this Conference.</p>		

DD FORM 1473
1 JAN 64

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Information retrieval						
Library automation						
Documentation						
Citation index						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parentheses immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those

imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

UNCLASSIFIED

Security Classification

CRANFIELD CONFERENCE ON INFORMATION RETRIEVAL

Introduction

The Third Cranfield Conference on Mechanised Information Storage and Retrieval Systems was held on 20-23 July 1971 at the Cranfield Institute of Technology, in Bedfordshire, England. Over 200 information scientists, documentation specialists, cataloguers, and librarians attended the bi-annual conference which is perhaps the most important international meeting on information retrieval. The reason it is held in Cranfield is that the conference organizer, C. W. Cleverdon, an internationally-recognized pioneer in information processing, is the director of the Institute library. Largely through Cleverdon's efforts, the program included invited talks given by many of the world's leading specialists in information retrieval and documentation including Professor A. Lazarow of the University of Minnesota, Dr. E. Garfield of the Institute for Scientific Information, Dr. C. J. Bell of IBM (UK), and Dean B. M. Fry of the University of Indiana. Of the attendees, most came from Great Britain. Other countries represented were: Czechoslovakia, Denmark, Sweden, the Netherlands, Belgium, Australia, Canada, Luxembourg, West Germany, France, Japan, Ireland, Switzerland, Yugoslavia, and the United States. Most of the speakers were either British or American, and the official Conference language was English.

Most of the talks described on-going projects, some of which are quite specialized applications of information retrieval. Other papers described the use of magnetic tape services such as MEDLARS and COMPENDEX. A number of talks described on-going efforts for on-line information retrieval. At least forty-five minutes were allocated to each talk and parallel sessions were kept to a minimum.

On Wednesday, the professional association for library sciences and documentation specialists, ASLIB, held a one-day conference of their computer applications group. Registrants to the main conference were invited to attend the ASLIB meeting also. The papers presented at the ASLIB Conference tended to emphasize more pragmatic issues dealing with the use of computers in the overall operation of the library. To an outside observer, however, the two conferences were so similar in intent that they could easily have been integrated into one.

The Conference was well-planned and efficiently run. Cranfield was an excellent place to hold it. Accommodations were comfortable, the surroundings quiet and beautiful, and the food was excellent.

Contributed Papers

D. B. McCarn, "Networks, with Emphasis on Planning and On-line Bibliographic Access Systems."

McCarn, who is with the Lister Hill National Center for Biomedical Communications, gave a tutorial paper on computer networks. His talk covered the advantages of networks, different configurations, transmission speeds, and decentralization versus communications costs. As an example, he described the planning for a national on-line medical bibliographic service. The service, known as AIM TWX, provides on-line search of all items in the index of the Abridged Index Medicus for the past five years. Users access the central data bases in the IBM 360/67 computer at SDC, Santa Monica via TWX or teletype terminals. McCarn estimates that 60% of the users are physicians. Each search averages 10 sec vs 3-6 weeks for a conventional batch-processing search using the MEDLARS system of the National Library of Medicine. The cost of each search is approximately \$4.00 with \$3.00 for communications and \$1.00 for computer time. When fully operational in 1972, the system is designed to handle about 180,000 requests per year.

A. Lazarow, W. J. Schoener, E. H. Brekhus, J. L. Izzo, K. C. Goodman, and K. Kreuger "The Diabetes Literature Index and the Developing Minnesota Diabetes Information Center."

In one of the finest papers presented at the Conference, Lazarow described the Diabetes Literature Index work of the National Institute of Arthritis and Metabolic Diseases (NIAMD). He described the process whereby diabetes-related literature is derived from the National Library of Medicine's MEDLARS magnetic tapes which is a regularly updated index of all current medical literature. Approximately two-thirds of the diabetes-related citations are automatically identified by computer via a key-word index. The remaining one-third is manually selected by an experienced MD. The final derived magnetic tape for the diabetes index contains the complete bibliographic citation indexed by key words. Lazarow strongly believes that the complete citation is much more useful than a fragmented title and saves valuable user's time at the cost of slightly greater computer processing time. Every month, the diabetes index tape is sent to the Government Printing Office where the Diabetes Literature Index is printed. The project is a prime example of effective utilization of computers in information retrieval.

W. M. Woods, "Design and Development of a Transdisciplinary Engineering Information Program."

B. Hisinger, "A Multidisciplinary and Computerized SDI Service for Industry and Research - Practical Experience and Costs."

M. E. Williams, "Experiences of IIT Research Institute in Operating Computerized Retrieval System for Searching a Variety of Data Bases."

J. Byrne, P. J. Currivan, and P. V. Mahon "A Current Awareness System Based upon INSPEC Tapes."

The above four papers all deal with information services on magnetic tape. These services send to their clients each month a magnetic tape with thousands of abstracted items containing subject headings, cross references, bibliographic citations and abstracts.

An information service that was discussed in three of the above papers was the COMPENDEX (Computerized Information Index) tape service which was established by Engineering Index, Inc. in 1969. COMPENDEX offers its subscribers a package consisting of a monthly magnetic tape service with more than 6000 items, and the monthly and the annual Engineering Index. B. M. Woods, managing director of Engineering Index described how COMPENDEX works.

B. Hisinger of the Technical Library of Denmark gave a talk on a computerized SDI (Selective Dissemination of Information) service based upon COMPENDEX and INSPEC (a tape service offered by IEE) tapes. In another paper that dealt with COMPENDEX, Miss M. E. Williams described a current awareness (SDI) service developed at the Illinois Institute of Technology Research Institute. In addition to COMPENDEX, the SDI service also reviews Chemical Abstracts Condensates and Biological Abstracts Previews. In operation since 1969, the service also offers retrospective searches tailored to a user's or organization's needs.

For COMPENDEX services, the yearly leasing cost is \$7300, broken down as follows: \$6500.- for the COMPENDEX data base, \$500 - for the monthly subscription to Engineering Index, and \$300 - for the magnetic tapes themselves. Although individuals probably can not afford the COMPENDEX service, institutions with a fairly substantial library budget may be able to.

D. E. Garfield and A. E. Cawkell, "The Relationship Between Large Multidisciplinary Systems and Specialized Information Services."

In one of the most hotly debated papers, D. E. Garfield, President of the Institute for Scientific Information, and a pioneer in Information Retrieval gave a talk on citation indexes. He showed how citation indexes give valuable feedback on the relative importance of a particular paper

or journal in terms of the number of times the paper or journal has been cited in other papers. In advance of his talk, Garfield distributed an alphabetic listing of the 100 most frequently cited scientific journals. This listing ranks scientific journals according to two standards - rank by total citations, and rank by impact factor. The term impact factor is defined as the ratio of total citations to total source items over a given period. Thus, impact factor is a normalized criterion in which the size of the journal, i.e., its total number of pages, does not enter in, and perhaps is a more relevant index than total citations. For the reader's information, the ten most frequently cited scientific journals are listed below. The complete list is to be published in Science (number seven in total citations). It should be noted that Garfield's complete list includes over a thousand journals, all ranked by total citation and impact factor.

<u>JOURNAL</u>	<u>Rank by Total Citation</u>	<u>Rank by Impact Factor</u>
J. AM CHEM SOC	1	6
PHYS REV	2	18
J BIOL CHEM	3	4
NATURE	4	48
J CHEM PHYS	5	26
J CHEM SOC	6	31
SCIENCE	7	34
BIOCHEM BIOPHYS ACTA	8	29
P NAT ACAM SCI USA	9	2
BIOCHEM J	10	30

The journal ranked first in impact factor, J Mol Biol, ranked eighteenth in total citations. From the listing of the 100 most frequently cited journals we can make the following observations:

- (1) All the journals listed relate to science or medicine. Not a single journal in engineering or technology appeared in the top 100. This does not mean that their impact factor is not high, but does mean that engineering journals are rather too specialized to be cited widely.
- (2) Most of the journals in the top 100 are in the English language, which gives some support to the thesis that English is the international scientific language, rather than French, Russian or German.
- (3) Review journals, such as the Rev Mod Phys which ranks 97 in total citations and 12 in impact factor, usually have high impact factors. This is because

most scientists do not have enough time to keep up with the literature, even in their own specialization, and must increasingly rely on review articles for information.

Garfield was challenged rather sharply in his total dedication to citation indexes as the only acceptable means for information analysis. He stood his ground, however, and insisted that his numbers are valid. His concluding remarks were that in the information explosion we are living in today in which the number of journals and number of journals pages are increasing almost exponentially, a small number of journals, say 200 account for a substantial percentage (30%) of the total citations. Therefore, even with new journals cropping up everyday, the old standards, such as Science, remain as important as ever.

Garfield's talk served to underline the importance of information retrieval in an age where too much information is disseminated for people to absorb. It also showed the importance of meetings such as the Cranfield Conference.

S. Adams "The Role of Reprocessing in UNISIST."

UNISIST is an acronym which stands for a feasibility study on the establishment of a world science information system. The study was jointly sponsored by the United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the International Council of Scientific Unions (ICSU). The study, reported in a monograph entitled UNISIT (UNESCO, Paris 1971), shows that it is both desirable and feasible for developing nations to have information services, and calls for the sharing of information costs and services between developed and developing nations. Other recommendations call for the establishment of governmental policies favoring such interchange, and the experimental interlinking of information systems between developing and industrialized countries. Adams suggests that with the implementation of information sharing programs, national reprocessing centers, i.e., centers for the re-packaging of information from data bases such as COMPENDEX, should be created. The talk was highly speculative and it was difficult to assess what specific actions Adams recommended.

It should be noted that a UNESCO conference is being held at this time (1 October 1971) in Paris to discuss and implement the recommendations given in UNISIST.

A list of papers appears in Appendix A.

APPENDIX A

LIST OF PAPERS

- D. B. McCann: Networks, with emphasis on planning an on-line bibliographic access system
- W. M. Woods: Design and development of a transdisciplinary engineering information program
- E. H. Brenner: Interactions between distributors of discipline-oriented information, industry information scientists, and the ultimate users
- J. R. Kelly and P. Popper: Anglo-American cooperation in a company information system
- S. Keenan: Recent developments in the production of computer based services and prospects for the 1970's
- J. Newton: Development of computer based retrieval services for the International Food Information Service
- M. Williams: Experiences of IIT Research Institute in operating a computerized retrieval system for searching a variety of data bases
- E. J. Scott, H. M. Townley and B. T. Stern: A technique for the evaluation of a commercial information service and some preliminary results from the Drugdoc service of Excerpta Medica Foundation
- F. H. Barker, D. C. Veal and B. K. Wyatt: Towards the automatic construction of search profiles.
- F. H. Barker, D. C. Veal and B. K. Wyatt: Comparative efficiency of searching titles, abstracts and index terms in a free-text data base
- A. Merta: State information policy in Czechoslovakia in the field of science and technology
- B. Hisinger: A multidisciplinary and computerized SDI service for industry and research - practical experience and costs
- J. Martyn: Serving interdisciplinary need-groups from basic abstracting sources

APPENDIX A

- A. Lazarow, W. J. Schoener, E. H. Brakhus, J. L. Izzo, K. C. Goodman, and K. Kreuger: The diabetes literature index and the developing Minnesota Diabetes Information Center
- M. Stutzman and E. de Hart: Computer-based humanities reference service and the university library
- P. G. Watson: Computerised information services for the University community
- C. E. Elwin: Some aspects on integration of information services in medicine, biology and chemistry at a multidisciplinary documentation center
- C. J. Bell and T. W. Rogers: Adaptability to change in large data base retrieval systems
- E. Garfield and A. E. Cavkell: The relationship between large multidisciplinary systems and specialized information service.
- A. E. Negus and J. L. Hall: Towards an effective on-line reference retrieval system
- L. D. Higgins, M. Carville and F. J. Smith: Interactive reference retrieval in large files
- L. H. Thiel and H. S. Heaps: Program design for retrospective searches on large data bases.
- C. R. Clough and K. M. Bramwell: A single computer based system for both current awareness and retrospective search: operating experience with 'ASSASSIN'
- T. M. Aitchison and M. D. Martin: Design and development of a tape service
- J. Byrne, P. J. Curriuan and F. V. Mahon: A current awareness system based on INSPEC tapes
- W. Simmler: A documentation system for inorganic and physical chemistry
- C. D. Green: Problems of indexing and retrieval for multidisciplinary information fields

APPENDIX A

- Scott Adams: The role of reprocessing in UNISIST
- B. M. Fry: New management concepts for interlibrary cooperation
- H. F. Dammers (Shell Research Ltd., Sittingbourne): Introduction: The work of the Aslib Computer Applications Group
- A. Evans (Loughborough University of Technology): Developments in serial control systems
- R. M. Duchesne (Bath University of Technology): Survey of automation in university libraries
- C. W. J. Wilson (AERE, HARWELL): Trends in automatic circulation control systems
- R. Coward (British National Bibliography): The B.N.B. West Sussex MARC-based acquisition system
- T. M. Aitchison (INSPEC): Criteria for computer output in information systems
- L. Corbett (AWRE, Aldermaston): Magnetic tape data bases: Availability and search programs