

# SOME NEW TERMS IN DOCUMENTATION AND A GLOSSARY OF ACRONYMS

B GUHA

*Insdoc, New Delhi-12.*

J S GHOSH

*Directorate of Research Coordination and  
Industrial Liaison, New Delhi-1.*

The recent rapid growth of new terms in documentation is discussed. It is held that they are mainly due to the interdisciplinary nature of documentation and information retrieval field, and the current active phase of machine application in this field. A glossary of about 60 terms, mostly names of retrieval systems, are presented bringing out their salient features. References have also been added for each of the items.

## A Traditionalist's Choler

Reading is very often a hurdle race  
When one happens to start with GRACE

Soon stumbles on CCF, PCF and MEDLARS  
But he aspires to be amidst scholars

So he continues still, and if not dizzy  
Before long he has to be really busy

Getting to know what is ALP and ALPHA  
And that is not all as he comes on ASCA

Attempts to find needle in haystack  
Takes him only before a real HAYSTAK

Three sweet names ALGOL, COBOL, SYNTOL  
Are they three sisters, pretty and tall?

When he gets 'SURF uses MADAM' in a page  
It is as if he stops before a haze

Only he can pity the unknown madam  
Before deciding finally to abandon

God help him from such cruel wordiness  
Documentation literature is in utter mess.

[Inspired by 'Scholar's Choler' by Jane Apostol]

Indeed, documentation literature has, in recent years, been invested with such unfamiliar, at times quite intriguing, terms and expressions that for a traditionalist librarian it is often difficult to follow and understand current literature. This luxuriant growth of new terms are, we believe, mainly due to the interdisciplinary nature of documentation and information retrieval field, and the current

active phase of machine application in this field. Designing of systems for information storage and retrieval has become the common subject of investigation by many diverse groups during the last few years. Librarians, who were the earliest, have been joined in this field by documentalists, mathematicians, system designers, equipment manufacturers, operation researchers, computer programmers and scientists in increasing numbers. The presence of so many workers, in the same field, with different backgrounds, capabilities and views is bound to create some communication problems, as each worker is to some extent unfamiliar with the others' language. That is exactly why no research area has so soon needed a glossary of terminology in use. What is more important is that this glossary needs to be constantly revised augmented and kept up-to-date. This is one aspect of the picture.

The other aspect is the application of machines in information problems. Since the end of the last world war, computers, photographic processes, and magnetic recordings have been holding out, in a very tantalizing way, the prospect of solving many problems of information storage and retrieval. From a period of optimistic and fanciful claims and promises, when we had such magic name like MEMEX, we have now come to a phase of more careful and purposeful thinking. This has meant actually a close study of the characteristics of the data processing machines and reprographic equipment, on the one hand, and the nature and mechanics of the whole process of information storage and retrieval on the other.

Surely, two basic characteristics of these machines are that they are both unintelligent and expensive, and both have deep implications. Because a machine is unintelligent it has to be told not only what exactly is wanted but also what is to be done to get it. Again, all these will have to be told in a

language which will be intelligible to the machine. This complicated business has meant analysing the whole process of information retrieval i.e. classification, or grouping the documents on the basis of their information content and appropriately tagging with codes, on the one hand, and analysing the requests, on the other hand, and then matching the two to find the relevant documents. This is not at all a new process. Librarians too have been following a similar process. While adding a new publication to a collection, cataloguers and classifiers keep in mind the probable users and their ways of approach and accordingly tag it under a subject group (with Class number) and choose suitable headings for the catalogue. When this same process is analysed by a different group of people with different background and a system is designed for machine retrieval, we get such expressions as, hit, recall ratio, noise, users profile, selective dissemination of information or simply SDI and so on. In course of time we get such expression as 'in any SDI system, maintenance of accurate users' profile and facility to adjust it to alter the level of match are the most important prerequisites'.

The second characteristic of data processing machines, as we have mentioned, is that they are expensive. This too has one important implication. This means that unless the entire cycle of information generation, storage and retrieval is considered together, machine application will not be economically viable. In other words it means that a systems approach is necessary, as any piecemeal work will be too expensive. With the realisation of this fact many attempts are now being made to design and operate mechanised systems which will take care of the entire bibliographical needs of subject-communities. There is a marked tendency to christen such systems with affectionate acronyms or initonyms. So, we have today MEDLARS, INTREX, ABLE and many more.

There is still another source for the generation of new terms. Probably, very soon, we will be facing the cumulative effect of the new terminology, as a new term is coined, that itself may be used as a stem for further culture to generate a few variants or new terms. Thus using KWIC as a stem we can coin Kwicable, an expression which has already come into us.

All these explain, in brief, how and why new terms are coined and put into use, and it promises also many such new terms in the

future. The difficulties of a traditionalist facing such semantic hurdles and also his joy and delight, if he is able to get over these hurdles are very clearly expressed by William H. Carlson in the following passage:

"Uniterm, bits, symbols, KWIC, Zato coding, auto encoding, descriptors... CRAM, Minicard, Magnacard, ... FLIP, FOSDIC, Filmorex, ILAS... HAYSTAQ, Termatex, Verac, WALNUT — these and many, many other terms, for the most part newly coined, probably with some pride of authorship, have come crowding into the literature and discussion, to the confusion of traditionalists like myself. Through the strange miasma and fog and pounding surf of semantic noise steadily swirling and growing I still perceive, nevertheless, the old familiar coast, capes, headlands, rocky shoals, and light houses" [1]

The last sentence is very significant, as we think it assures all traditionalists, specially those like us, who are unfortunately far removed from the main arena of these new experimentations, that there are many things which are familiar to the conventionalists in these new experimentations. Carlson further assures that traditional indexing, cataloguing and classification systems have been improvised, improved and perfected by devoted men like Dewey, Ranganathan, Martel and Hanson. Machine control of literature will also need its Deweys, Ranganathans, Hansons and Martels. Perhaps they are already at work now. If that is so, every traditionalist, we are sure, would like to know and understand their work.

This paper with short annotations on some of the major systems (specially with acronyms names) will, it is hoped, be helpful for a traditionalist to get some idea of the recent developments, and can be supplemented by reference to the bibliographical sources mentioned at the end of each item. We have selected such references which are easily available. We have not tried to prepare a comprehensive glossary of such terms, but our endeavour has been to pick out and explain only a selection of them and present them in a form which can be easily read through.

#### REFERENCES

1. Carlson, William H: The Holy Grail evades the search. Amer Doc, 1963, 14(3), 207-12.

2. Bourne, Charles P: Methods of information handling. New York, John Wiley, 1963.
3. Current Research and Development in Scientific Documentation. National Science Foundation. No. 13, 14, 1964, 1966, (NSF-64-17, NSF-66-17) (Abbreviated as CRDSD).
4. Nonconventional technical information systems in current use. No. 3, 4, 1962, 1966 (NSF-62-34, NSF-66-24).

## GLOSSARY

ABC (Approach By Concept): A storage and retrieval system recently developed by the Harry Diamond Laboratories, Washington to meet the specific information needs of the scientists and engineers of the organisation engaged in the production of new electronic and electrochemical systems, subsystems, and devices. In this system, all documents are evaluated, the selected documents are described in verbless sentences or strings of phrases, called descriptor sentences. Print-outs, arranging these descriptor sentences into a KWIC type dictionary with accession codes are obtained. This dictionary allows a user the facility of "browsability". The retrieval can be performed automatically or manually after consultation of the dictionary. If a computer output is preferred, the machine can be programmed to print out all the titles filed under the selected descriptors or their respective codes.

[Altonann, Berthold: A natural language storage and retrieval (ABC) method, its rationale, operation, and further development program. J chem Doc 1966, 6(3), 154-57. 3 ref.]

ABCD (Abbott Biological and Chemical Data): A system for storage and retrieval of biological and chemical data by means of machine readable tab cards and magnetic tape. It is in operation since 1959 at the Abbott Laboratories, Chicago. Documents are stored in the form of full-size copy or roll microfilm. Indexing characteristics used are subject or concept terms, headings or descriptors, molecular or structural formulae, biological, chemical or physical properties; project name or number, and accession number.

[Nonconventional scientific and technical information systems in current use, No. 4, p. 239]

ABLE (Agricultural Biological Literature Exploitation): A mechanized system under study and development at the National Agricultural Library (U.S., Dept. of Agriculture). Possible areas of automation are: Author Index for Bibliography of Agriculture (already started), addition of a monthly subject index, production of the annual index for the Bibliography of Agriculture, catalog card preparation including a machine readable record for producing other products, production of Titles Indexes in the Bibliography of Agriculture, update Subject Authority List, Selective Dissemination of Information system to include a customized title Service by Journal and permuted Title index according to user profile, complete production of Bibliography of Agriculture, producing records to be used later in an information retrieval system, Title received by NAL, Current serial records, and Research Project files. Some preliminary studies on some of these activities have already been done.

[U.S., National Agricultural Library: Agricultural Biological Literature Exploitation - Report of Task Force. 1965, p. 477]

ACORN (Associative Content Retrieval Network): The Arthur D. Little, Inc. has been investigating the degree of centralization which should be planned in designing documentation retrieval mechanisms best suited to the needs. As a part of this investigation, several experimental network devices have been built. One of them enables associative network identification of technical information centres in the area of Political Science, according to the subjects with which these information centres deal. This device is an ACORN, a simple analogue information processing device which performs a fairly complex highly-parallel information processing function.

[CRDSD, No. 13, p. 59]

ALP (Automated Library Program): The Datatrol Corporation has developed this system which is designed for the IBM1401 computer. The speciality of the system is that it can accept almost any vocabulary, record format, and degree of indexing sophistication. It permits simultaneous multiple searches; optional use of links, roles, subject categories, and indexer's or searcher's weighting

of terms; and the automatic arrangement of output in the order of probable relevance.

[CRDSD, No. 13, p. 83]

**ALPHA (Automated Literature Processing, Handling, and Analysis) System:** The system is under development at the Redstone Scientific Information Centre, Alabama. The project is aimed to develop a single automated system of performing the maximum number of scientific and technical library functions as an integrated whole. Under its perview comes advanced information retrieval, current awareness services, retrospective inquiry, circulation control, ordering, receiving, and general management processes.

[CRDSD, No. 14, p. 356]

**ASCA (Automatic Subject Citation Alert):** This a selective dissemination of information (SDI) service provided by the Institute for Scientific Information, Philadelphia. In this system, each scientist or organization submits a profile of interest by citing about 50 articles or keywords. The profile is used in the ASCA computer system to select, on a weekly basis, the pertinent literature currently appearing in over 1400 journals which publish about 250,000 items per year. A printout, in the style of Science Citation Index, indicating the articles and their sources where any of the articles from a client's profile is cited, is sent to the clients.

[Garfield, Eugene: ISI services in the design of small-user systems. J chem Doc 1966, 6(3), 164-68. 8 ref]

**BRAID (Bidirectional Reference Array, Internally Derived):** A technique for using large-scale random-access computers to search complete text, under investigation at the Datatrol Corporation. The technique also shows promise for use in chemical structure search. The speciality of BRAID lies in its capability to eliminate redundancy, which is made possible by the substitution of computer addresses for actual words or phrases, permitting the storage of a complete bidirectional concordance.

[CRDSD, No. 13, p. 84]

**ChemSEARCH (Chemicals Selected for Equal, Analogous, or Related Character):** Developed at the Colgate-Palmolive Research Center, New Brunswick. This is essentially a

mechanized system for handling documents relating to chemical and biological activity in relation to chemical structure. Hence, at the input stage, stress is given on the complete representation of the molecule, atom by atom, in a structural formula. This has its strength in the chemist's familiarity with structure. The input is through paper-tape-punching typewriter. When a search is made through the Question Reader program the question is made in the form of a 'question structure' and the computer makes a comparison of the question structure with the file structures.

[Gould, David, Gasser, Edward B and Rian, John F: ChemSEARCH - an operating computer system for retrieving chemicals selected for equal, analogous, or related character. J chem Doc 1965, 5(1), 24-32]

**CITIS (Centralized Integrated Technical Information System):** CITIS builds a model in which linear programming methodology is applied to the information sciences field. The CITIS contribution to the area of centralized integrated technical information systems is to provide the information center manager with the vehicle to utilize developments within the "Information Sciences" and to better manage the information center operation. The CITIS Model 1 illustrates the key summary resource allocation relationships that are necessary for the effective development of a "centralized-decentralization" operation for a technical or a scientific information center.

[Podell, H.J.: CITIS, a generalized mathematical model for a centralized integrated technical information system, from a resource allocation point of view. American Documentation Institute, Annual Meeting, 27th, 1964, Proceedings. 155-72]

**COMEINDORS (Composite Mechanized Information and Document Retrieval System):** Developed at U.S. Army Biological Laboratories, Fort Ketrick, Frederick, Maryland. A semi-mechanized information and document retrieval system. This system uses a "pure decimal" notation; in one or another manner, the classification resembles the categories of Roget, the ASTIA thesaurus, the codes of Dewey, the arrays and chains of Ranganathan, and the faceted classification of Vickery. Dewey's zero is extended to serve as a code marker for categories that are, or may be,



divided into more specific terms. One of the chief components of COMEINDORS is a series of lattice-type classification forms that simplify and expedite the development and control of the ordered classification of terms.

[Batchelor, H.W. & Maloney, C.J.: A Mechanized information and document retrieval system. *J chem Doc* 1964, 4(3), 181-5]

**COMIT:** This is the name given to the MIT computer programming language for mechanical translation. The programming language is quite different from others because of its different purpose. The language is designed to give the linguist direct access to the computer without his having to concern himself with details that are irrelevant to his problem. The language is being provided with a compiler and interpretative routine written for the IBM 704 computer.

[Yngve, V.H: The COMIT system for mechanical translation. (*In* Information Processing - Proc. int. Conf. Inform. Processing, Unesco, Paris, 1959. London, Butterworths, 1960. pp. 183-87)]

**CRIS (Command Retrieval Information System):** Developed in 1962 by a subsidiary of Information for Industry, is an outgrowth of some earlier work done by Avatcian\* in 1956. This system stores images photographically on a scroll of microfilm 400 feet long by 17 inches wide; each scroll contains over 500,000 page-size images or over 28,000 large drawings. A keyboard device is used to enter a CRIS address, and the image at that address is displayed to the operator or is provided in aperture card form. The average time to retrieve any desired image is under 20 seconds, and the retrieval time for sequential images is even faster. The scroll material is a Mylar base with a Kalfax emulsion and is usually prepared by contact printing from strips of microfilm that were obtained by conventional procedures.

[Larson, P.W.: CRIS (Command Retrieval Information System) Proc 11th Annual Convention of the National Microfilm Association, pp. 41-50 (National Microfilm Association, Annapolis, Maryland, 1962). CRDSD No. 13, p. 247]

**DACOM (Datascope Computer Output Microfilmex):** Translates computer language on magnetic tape to human language on microfilm. It includes five basic, interrelated sub-

systems: 1) The input by means of magnetic tapes; 2) The character generator, using a glass plate with a capacity of 64 characters; 3) a display tube; 4) on-demand 16 mm camera; and 5) the logic and control sub-system. DACOM can create pages of alpha-numeric data, and has also available three other modes: plot mode, axis generator mode, and vector generation mode. Information written on eight or more reels of 2,400 foot magnetic tape can be recorded on 100 feet of Dacom microfilm.

[Effros, A.L.: The Recordak Dacom system. New York, 1961, p. 5. (Paper presented at the Symposium on Information Processing of the Los Angeles Chapter, American Documentation Institute, September 29, 1961)]

**FAIR (Fast Access Information Retrieval):** This is a research project based at MRC Laboratories, Holly Hill, Hampstead, London and is being financed by the Medical Research Council. The aims of FAIR are: (i) to produce a formula for creating efficient indexes to feature-card (also known as 'peek-a-boo') information retrieval systems; (ii) to explore the possibilities of members of a learned society helping in the setting up of an information retrieval system for their own use, and (iii) to test the practicability of providing a whole library on the individual's desk. Subject to copyright arrangements the whole collection may be microfilmed and a copy made available to each satellite, where there would need to be a microfilm reader or reader-printer.

[Morton, Leslie T: Notes from London. *Bull med Lib Ass* 1966, 54(3), 255-56.  
Pickford, A.G.A: FAIR (Fast Access Information Retrieval) Project; aims and methods. *Aslib Proc* 1967, 19(3), 79-92]

**FLIP (Film Library Instantaneous Presentation):** This is a photographic film reading machine developed by Bensen-Lehner Corporation. The machine is able to search for a particular frame on 16 mm film at a speed of 300 to 600 frames per second, then present this frame for viewing to the operator. Each frame contains pictorial data and binary coded information. The equipment for the operator consists essentially of a keyboard interrogation device, and a large display screen. The keyboard positions correspond to the coding positions on the film so that the inquiry can be keyed into the system via the keyboard.

[Worsley, Peter K: Data retrieval with special application to use of Film Library Instantaneous Presentation (FLIP) in literature searching. (In Boaz, Martha, ed. Modern trends in documentation. Oxford, Pergamon Press, 1961, p. 70)]

GREMAS (Generic Retrieval by Magnetic Tape Storage): Work on the system is going on at the Farbwerke Hoechst AG, vormals Meister Lucius & Bruning, Frankfurt since 1955. It is adapted for literature on organic chemistry. Organic chemical compounds and their reactions are classified according to the principles of analytico-synthetic classification and the notations based on this classification are used by computer for storage and retrieval. At present more than 150,000 structural formulae of organic chemical compounds and of groups of compounds have been stored as well as several thousand organic chemical reactions. IBM 705 and IBM 1401 are used and 32 different machine programmes are in operation for storage and retrieval.

[Fugmann, Robert: Experiences with a faceted classification in organic chemistry using computers. (In Classification Research, Proc Int Study Conf, 2nd, 1964, Elsinore 341-67)]

HAYSTAC (Have You Stated Answer to this Question): The system design and computer programs which make up the completed portion of the project to date are the result of collaborative efforts between the U.S. Patent Office and the National Bureau of Standards. The system is particularly designed to search chemical information for mechanizing patent search questions. The file to be searched is contained on a magnetic tape, and each entry in the file is an encoded chemical structure. Each structure is broken up into functional groups, with each such piece assigned a number known as its 'designation number'. HAYSTAC executes a topological tracing among functional groups. The Hayward rules for chemical structure notation are being revised and expanded in order to simplify their application and widen their domain.

[Marden, Ethel C. and Koller, Herbert R: Present status of Project Haystack p. 163 (In ICIREPAT, 3rd Annual Meeting, Vienna, 1963. Baltimore, Spartan Books, 1964). Koller, Herbert R, Ethel C. Marden, and Harold Pfeffer: The HAYSTAC System; past, present, and future. International Conference

on Scientific Information, Washington, D.C., 1958. Proceedings 1959, Vol 2. pp. 1143-79]

IDEA (Information and Data Exchange Experimental Activities): The objectives of the program, which is under the Department of Defense, are to establish an experimental test bed for the purpose of gathering live statistical data and the testing of advance data processing techniques. To effect maximum economy the existing militarized equipment, existing software packages, and the knowledge gained from several research projects in the military computer field are being utilized. The main equipment is the FADAC computer which is used as a computational module at several stations of the network. The next piece of equipment is a magnetic tape unit called "Tape Tub". The Gunnery Officers Console (GOC), a programmable switch, allows the computers to be interconnected to each other and to peripheral devices. Existing Data Transmission Link Equipment are used to give the network a wireless capability. All communications with the systems computers are made through the key-board of Army Chemical Typewriter (ACT) a three-shift punch-paper-type typewriter. In short, the IDEA program represents a 'supermarket' of techniques that can form the basis for system decisions in terms of procedures, equipment, software, and communications.

[Hoffman, Allen: The Information and Data Exchange Experimental Activities (IDEA) Program and its relation to the national interests. (In Towards a national information system - 2nd Annual National Colloquium on Information Retrieval, 1965. Washington, Spartan, 1965. p. 87-104)]

INTREX (Information Transfer Experiments): This is a four year programme at the Massachusetts Institute of Technology under Professor C.F.J. Overhage as Director. The objective of the experiments is to provide a design for evolution of a large university library into a new information transfer system that could become operational in the decade beginning in 1970. The project would establish a 'model library' on computer. This would consist of a collection (equivalent to about one million pages) of books, periodicals, abstracting and indexing services, items of non-book character, all falling broadly within one subject field. Fifty terminal stations or consoles are envisaged within a radius of about

a mile on the campus which would give remote-access to the 'model library'. These would permit the transmission of the content of the documents as stored in machine-readable form for print-out at the terminals or for display on a tube and facsimile copying therefrom. It is proposed to organize the contents of whole groups of basic reference works so as to provide factual responses to questions, rather than references to works containing such responses.

[INTREX: Report of a Planning Conference on Information Experiments, ed. by C. F. J. Overhage and R. Joyce Harman. Woods Hole, Mass. 2 Aug - 3 Sept, 1965. Cambridge, MIT Press, 1965. p.294]

KAPL (Knolls Atomic Power Laboratory): Services available from the system are: Retrospective searches, current awareness searches, notification abstracts, accession bulletins, computer-printed reference tools for manual search, and dissemination of microfiche to fill requests for permanent copies of any documents cited. Documents are indexed according to a classification schedule which may be described as a post-coordinated faceted classification scheme.

[Schmidt, C. J., R. P. Wesley and C. H. Hunter: KAPL's automated information retrieval system. Proc Amer Doc Inst Vol 1. Parameters of information science, Annual meeting, Philadelphia, 1964, pp.177-64]

KWIC (Key-word-in-context) index: Usually we mean a permutation index based on titles and produced by machine. Best examples of this type of indexing are to be found in the 'Chemical titles' and the 'BASIC Index'. Although indexing by key-words, with meaning clarified by context was not new, yet the credit for developing the idea for machine adaptation and also coining this expression goes to Dr. H. P. Luhn of the IBM. The underlying principle is that all essential words can be extracted from the title, abstract, or text and can be used effectively. The context about a keyword helps to define or explain its use, in order to lead the index user to the exact article, paper, or other bit of information he desires. The principal merit of the method is timeliness. That is why it is being used more and more in current awareness services or where time and resources are not available to prepare a conventional subject index. Some variations of KWIC are KWIT

(Keyword-in-title), KWOC (Keyword-out-of context), and WADEX (Word and author index).

[Fischer, Marguerite: The KWIC index concept, a retrospective view. Amer Doc 1966, 17(2), 57-70]

LINCO (Linearly Organized Chemical Code for use in Computer Systems): Developed at the Shell Internationale Research Maatschappij N. V., the Hague. The previous name LOCUS has been changed to LINCO for foreign language reasons. The system is based on the use of Chemical structural formulas, which are treated as networks containing branching points, giving to these points an arbitrary number and writing down the series of atoms between two points of that type together with series of atoms extending from such branching points but not meeting others. The LINCO notations can be easily typed on conventional tape-making typewriter and used as the input system. The notations make it possible to search for whole compounds by simple comparison.

[Bouman, H: Computer program for the LINCO system. J chem Doc 1965, 5(1), 14-24]

LITE (Legal Information Thru Electronics): A system established in 1964 at the U. S. Air Force Accounting & Finance Center, Denver, to store and supply legal and regulatory information to departmental users interested in legal, fiscal and financial management. It uses IBM 1410 computer with 1403 printer and 1402 card reader-punch.

[Nonconventional scientific and technical information systems in current use, No. 4, p. 432]

MAC (Machine Aided Cognition, 'Man And Computer') and a few detractors allege stands for MIT at Cambridge. This is a model project operating in a realistic test environment at the MIT. In its present configuration, a user may sit at an electric typewriter, scan a stated range of literature and perform a search based on keywords, keyword in context, citation index, bibliographic coupling, author, location, and various combinations of these. The response is printed back on the same typewriter within seconds of the request. There is no intermediaries between the user and the system, and the system uses a language very close to natural English.

[Kessler, M.M.: The MIT Technical Information Project, *Phys today* 1965, 18(3), 28-36]

**MARLIS (Multi-Aspect Relevance Linkage Information System):** Developed at Imperial Chemical Industries Ltd., Research Department, Hertfordshire, England. It basically comprises three distinct units with some ancillary equipment. The first unit, the Intermediate Store, consists of a store of records, on each of which is inscribed: (a) one or more addresses in the ultimate store; and (b) code symbols characterising the items at these addresses. These code symbols relate to only a few selected aspects (Characterising features) of the items in the ultimate store. For each set of symbols (descriptors authors, etc.) records are arranged in the ordinal sequence of that set. The records may be physically separate units, slacked serially in cabinets, or they may be grouped in sheets. Other forms of storage may also be used: photographic, magnetic, etc. The second unit of the MARLIS is the ultimate store. This may also take several physical forms, but it consists essentially of an assembly of items which is known as "volumes" because of the large amount of information each contains. On each of these volumes is inscribed a symbol which is its address, and by which it is identified in the Intermediate Store. The third unit of the MARLIS is known as the HOMO, so called because it is almost human, is the computing unit of the system. It includes input and output devices, and a memory, comparison and control organ.

[Vickery, B.C. (The) MARLIS: A Multi-Aspect Relevance Linkage Information System - Present position and future needs. *Amer Doc* 1960, 11(2), 97-99]

**MATICO (Machine Application to Technical Information Center Operations):** A system established in 1962 at the Lockheed Missiles and Space Co., for generation of catalogue cards, KWIC announcement bulletin and retrieval history tape. It adds about 85,000 documents per year of all types on all aerospace related fields like missiles, ocean systems, astronomy, chemistry, physics etc. Indexing terms are used from the Library of Congress List of Subject Headings and Thesaurus of ASTIA Descriptors. Indexing characteristics used are subject terms or descriptors, personal author, organizational or corporate author, project name or number,

report number, contact number, security classification, geographic location of originating organization, date of issuance and publisher. Equipment used are IBM 1401, 1410 and 7094 computers.

[Conventional scientific and technical information systems in current use, No. 4, p. 205]

**MEDIA (Magnavox Electronic Data Image Apparatus):** The system was first demonstrated by Magnavox in 1961. This photographic unit-record storage system uses a combination of manual and machines search techniques. The basic film chip is a 16 mm-by-32 mm card that contains upto two 9-by-15 inch page images at a reduction of 30:1, and an information field of up to 17 encoded and human-readable digits to identify the image. The documents are photographed on 100 foot rolls of microfilm using a special camera unit, processed, and cut into cards of the rate of 240 per minute. The cards are kept in capsules or cartridges containing upto 200 cards. These capsules are stored in regular file cabinets and are handled and selected manually. To retrieve a specific file item a clerk selects capsule number from the file cabinet mounts it in the selector unit, and keys in the last two digit. The cards are then scanned at the rate of 600 cards per minute to select the desired one. This unit can provide a page-size blowback of the card image if desired. A browsing unit is also available; with this, the user can view the cards in a capsule one at a time in a regular microfilm viewer-printer, and can transfer selected copies to hard copy form.

[Jenkins, D.D. "Magnetic Indexing, Microfilm Storage and Information Retrieval" in *Proc of the 11th Annual Convention of the National Microfilm Association*, pp. 205-12. (National Microfilm Association, Annapolis, Maryland, 1962).

Laurent, R.L. "Magnacard-Magnavue-MEDIA", paper presented at a special conference of the Los Angeles Chapter of the American Documentation Institute (Los Angeles, September 1961)]

**MEDLARS (Medical Literature Analysis and Retrieval System):** The project is in operation at the National Library of Medicine (NLM) since 1964. It is a multi-purpose system designed, (1) to manipulate citations for the composition of published indexes, including the monthly Index Medicus and its annual



cumulation, (2) to re-package citations in the store to create specialized recurring bibliographies (or group Selective Dissemination of Information Services), and (3) to perform mechanised retrospective searches on specific medical topics and prepare demand bibliographies. It is a publication system as well as a retrieval system.

In order to meet the speed and volume requirements of the system, the following items of data processing equipment are employed: punched-paper-tape typewriters for conversion of source data to machine readable form; a Minneapolis-Honeywell 800 computer for editing, sorting, compressing, merging, and formatting data for subsequent printing; an optical printer called GRACE (Graphic Arts Composing Equipment) for converting computer output into high-quality photo-copy; and an automatic film processor for developing the film from GRACE, thus producing a photomaster for printing plates.

The NLM plans to decentralise the MEDLARS search capabilities by distributing magnetic tapes to university and other medical research centres with adequate library and computer facilities. Two pilot centres have already been established under this plan, one at the University of California and the other at the University of Colorado Medical Center (UCMC).

[The MEDLARS story at the National Library of Medicine Washington, Dept. of Health, Education and Welfare 1963, p. 74.

Adams, Scott: MEDLARS, progress and prospects. (In *Toward a national information system - 2nd annual National Colloquium on Information Retrieval*. Washington, Spartan, 1965, p. 77-85).

Austin, C.J.: The MEDLARS Project at the National Library of Medicine. *Lib Res tech Serv* 1965, 9(1), 94-99]

MELISS (Literature and Information Search Service) MELCOM-1101 computer: A mechanized retrieval system under development at the Mitsubishi Electric Corporation, Japan. MELISS will be able to produce lists of accession numbers, technical information cards (abstract cards) or copies of original papers. The system will use coordinate indexing and use English language descriptors instead of codes. The user types the request in the same language as that used in the tape

memory and the request is then fed into the computer. In the preliminary stages, average retrieval time was less than 48 seconds per 300 documents.

[Current Research and Development in Scientific Documentation. No. 13, NSF-64-17, p. 132]

MERGE (Mechanized Retrieval for Greater Efficiency): The system has been recently developed by IBM to provide the Space Guidance Center with complete information retrieval and dissemination capability based on IBM 7090 and IBM 1401 computers. The system uses a controlled vocabulary in the form of the Thesaurus of Armed Services Technical Information Agency (ASTIA) Descriptors. The output of the system includes, SDI notification-response/abstract cards; a standard KWIC index (weekly bulletin of accessions), literature searches based on Thesaurus selective KWIC indexes based on documents selected during an SDI run; specialised KWIC index based on preselected keywords, descriptors, corporate authors, personal authors, contact numbers, and statistics for acquisition direction and weeding of the collection. It is essentially a system linking the techniques of KWIC with SDI.

[CRDSD, No. 13, p. 248; No. 14, p. 317]

MIRACODE (Microfilm Retrieval Access CODE): This is the name given to a 16 mm. film based system designed and developed by the Eastman Kodak Company. Miracode stores graphic images together with digital retrieval code. The Recordak Lodestar Reader-Printer has been modified by the addition of photocells, keyboards, and logical circuitry to permit automatic location (and printing) of document images based on logical comparison between keyboard criteria and binary coded data fields recorded on the film. The search equipment may also be actuated manually for 'browsing', where a visual image is presented on a screen or may be set to produce prints of automatically selected documents.

[CRDSD, No. 13, p. 244]

PACIR (Practical Approach to Chemical Information Retrieval): The object of the PACIR System is to provide a flexible universal approach to specific compound retrieval that is not limited to the unique characteristics of a particular area of chemistry, that is adaptable to various types of hardware, that

employs a maximum of machine assistance in the analysis, that provides a file that may be organized according to the need of the user, and that is practical from the economic viewpoint. This system has been developed at the Office of Research and Development of the U.S. Patent Office through several years' experimentation. The PACIR approach has the following characteristics: (1) non-art limited; (2) capable of storing and retrieving compounds, processes, compositions, biological data, etc.; (3) both specific and generic potential; (4) compatibility with several machines, including both random access and serial computers; (5) segmentable in accordance with need; (6) unlimited dictionary (open ended); (7) machine assisted encoding; (8) economic practicality.

The PACIR approach incorporates two different types of analysis that allow for selective depth of coding for a given type of subject matter in the chemical art being analyzed. One of these approaches is an underlining and role-assigning technique that produces catalogs of the ancillary subject matter of the art. The other procedure provides for specific coding in depth of any selected type of compound that can be represented by the conventional chemical structural formula. Both of these procedures are universal in scope as they are not limited by the need for a unique nucleus. In the application of PACIR to any particular chemical art, the analysis procedures are allocated to the various types of chemical subject matter within the art depending on consideration of cost and search need.

This system attempts to provide a standard analysis and linear and semi-automatic encoding technique that allows flexible and varied use of the analyzed subject matter. The approach has eliminated the dependence of the analysis on a particular type of file organisation, coding, or machine.

[Frome, J., O'Day, P. T. PACIR: Practical Approach to Chemical Information Retrieval. *J chem Doc* 1962, 2(4), 248-55]

PHILSOM (Periodical Holdings in the Library of the School of Medicine): Strictly speaking this is a product of a mechanized system. It symbolises the successful application of mechanized techniques in the routine works in the serials division of a library, like accurate determination of receipt of journals, correct billing and accounting, record keeping for binding, compilation of statistical reports,

and preparation of disseminable lists of library serials holdings etc. At the Washington University School of Medicine Library all these have been computerised, so that checking of receipt of issues, updating of holdings record and claims for issues not received are done automatically. The system, also produces two semi-annual lists: 1) PHILSOM, which contains all titles, holdings, history of variant titles etc; 2) PHILSOMS arranges all the currently received titles into appropriate subject classifications.

[CRDSD, No. 13, p. 183.]

Pizer I H, Franz D R, Brodman E: Mechanization of library procedures in the medium sized medical library 1. the serial record. *Bull med Lib Ass* 1963, 51(3), 313-38]

PICS (Pharmaceutical Information Control System): Developed at the Merck Sharp & Dohme Research Laboratories, and in operation since April 1963, provides centralized control and methodology for the series of decentralized information areas. The system input is approximately 1,000 documents per day, composed of reports, memoranda, correspondence, laboratory notebooks, regulatory agency submissions, prepublication manuscripts, etc. Published information is excluded from this system. An eight-digit, dual-faceted classification scheme has been designed for Division-wide project and product identification. Search requests are normally initiated through a telephone call or a personal visit from the user to the center.

[Kolb, Margaret C, Jerome T. Maddock, and Barbara N. Weaver: PICS - The Pharmaceutical Information Control System of Merck Sharp and Dohme Research Laboratories. *Amer Doc* 1966, 17(4), 180-85]

RAPID (Retrieval through Automated Publication and Information Digest): This name is given to the conceptual model of a coordinated, semiautomated information system that can improve dissemination without increasing storage problems. The whole system is built around a special science newspaper. Responsive to the recommendations of the Weinberg Report the RAPID system makes it obligatory on all authors to help others use that information, by selecting appropriate indexing terms, and in other ways. RAPID system has dual operation mission: (1) to publish and disseminate information; and (2) to process information in order to respond to specific requests.

[Schultz, Louise: RAPID, a system for retrieval through automated publication and information digest. *Proc Amer Doc Inst Vol 1. Parameters of information science, annual meeting, Philadelphia 1964*, pp.79-87]

**SABIRS (Semi-Automatic Bibliographic Information Retrieval System):** The system is being successfully experimented at the U.S. Naval Postgraduate School. Descriptors and uniterms are used in the storage and searches of documents. The Control Data Corporation 1604 computer is used for the search and retrieval process and the IBM 1401 computer is utilized for printing search and file updating data.

[CRDSD, No. 13, p. 168]

**SAID (Automated Information Disseminator):** The Datatrol Corporation has developed this system which is designed for the IBM 1401 computer. It uses similar contexting and weighting techniques, like the ALP system, in the selective dissemination of information.

[Current Research and Development in Scientific Documentation: No. 13, NSF-64-17, p. 83]

**SATIRE (Semi-Automatic Technical Information Retrieval):** The system has been operating in the System Development Corporation in California using electronic accounting machine equipment. From one typing (key-punching) operation five records (library cross-reference card catalogues, automatic charge-out cards, selected bibliographies, library indexes for publication, association lists) are produced. It can retrieve according to accession and report numbers, author, title, and subject. Subsequently, increased retrieval capabilities were obtained by moving to a computer-based system, and incorporating equipment to allow for 'self-service' and the provision of hard copy of retrieval information as final output.

[Roach, John P: SATIRE, the technical librarian's EAM application of Semi-Automatic Technical Information Retrieval. Santa Monica, California, System Development Corporation, 1961, p. 14 (SP-595)]

**SCAN (Short Current Abstracts and Notes):** A literature abstracting, publication and retrieval system in operation at the Wyeth Laboratories. The objective is to produce a means for acquainting both non-scientific and scientific personnel with current information

from the literature and also to find information on demand. The heart of the system is the IBM 870 Document-Writing System. The organization is made up of a master input system producing three subsystems-publication, machine language and search. Each document is abstracted and indexed by editors and then using the IBM 870 punched cards and a set of 5 x 3 cards are generated which include the abstract text and citation. From the 5 x 3 cards, with the help of the Kodak Ektalith Processor and Camera, a weekly current awareness list is produced. The punched cards are filed in the search subsystem. The output format may be title only, title and author or full citation.

[Elias, A W: SCAN - a flexible scientific literature system. *Proc Amer Doc Inst Vol I. Parameters of information science, Annual meeting, Philadelphia 1964*, pp. 121-29]

**SCISRS (Sigma Center Information Storage and Retrieval System):** A system established in 1964 at the Brookhaven National Laboratory under support from the U.S. Atomic Energy Commission. Its purpose is to store and supply experimental neutron cross section data to users culled from all types of documents like books, articles in periodicals, meeting proceedings, review publications, abstracts from internal or external abstracting publications or services, correspondence, theses, etc. Indexing characteristics used are, subject or concept terms, headings or descriptors, personal author, organizational or corporate author, report number, type or form of document and date of issuance of publication. The index term relationship is established by chemical element.

[Nonconventional scientific and technical information systems in current use, No. 4, p. 267]

**SHARP (Ships Analysis and Retrieval Project):** The objective of the project is to develop a computer-oriented information storage and retrieval system for the Bureau of Ships Technical Library (U.S.). A thesaurus of descriptors using links and roles has been developed and tested. The project will also computerise some of the library procedures and works.

[Current Research and Development in Scientific Documentation No. 13, 14, NSF-66-17, p. 69, 303]

**SHIRTDIF** (Storage Handling and Retrieval of Technical Data in Image Format): The speciality of the system is that it uses two storage media - digital storage for descriptive information and analog storage for copies of documents, and uses computer based search methods. This separation of descriptive information from document image is advantageous because any updating (e.g. adding new descriptions) is possible without necessitating the retaking of picture or in any other way disturbing the position and sequence of the document image. Images are stored on high-density film by photochromic process. This permits linear reductions of up to 200:1, corresponding to a 40,000:1 area reduction with high resolution, so that images can be stored in a very small amount of space. One 3 x 5 card can contain 2600 images. When a card is put into the system it is assigned a key which when typed on the keyboard the edge of the card is notched. To retrieve the card, the user types the key on the keyboard and a series of rods is set which finds the property notched card. Any page from a selected card can be viewed on a display screen through a viewer and a hard copy print obtained within 30 seconds, with multiple copies if desired.

[Dam, Andries van and Evans, David: SHIRTDIF - a system for the storage, handling and retrieval of technical data in image format]

**SIDAR** (Selective Information Dissemination and Retrieval): The system has been developed at Chemicals and Phosphates Ltd., Haifa, Israel. It is a modification of the SDI (Selective Dissemination of Information) first proposed by H.P. Luhn and implemented by A.J. Soward for the IBM 1401 DPS. The dissemination procedure is effected in two runs, SELECT and NOTIFY.

[Ofer, K.D. SIDAR: Selective Information Dissemination and Retrieval. J chem Doc 1964, 4(1), 54-55]

**SMART** (Socony Mobil Automatic Real Time): Computer System: The system is in operation since 1962. This computer-communications system has enabled Socony personnel at remote locations to transmit data to the computer centre at New York, direct the computer to perform certain computational and/or retrieval functions, and to receive 'answers' in a matter of seconds. Thus as many as 64 different company locations may

use the large scale computing capabilities of the headquarters. Computer facilities, at present, are IBM 1410-80K computer, IBM 1301 random access disk memory, and a Digitronics buffer and they are linked with conventional teletype machines.

[Wilson, H.B.: SMART (Socony Mobil Automatic Real Time) Computer System. J chem Doc 1966, 6(2), 89-92]

**SMART** (Salton's Magical Automatic Retriever of Texts): The system has been developed by Gerard Salton of the Computation Laboratory of the Harvard University. SMART operates on the IBM 7094. 'Unlike other automatic systems which rely on manually assigned keywords or index terms for the identification of documents and search requests, an attempt is made in the SMART system to go beyond simple word-matching procedures by using a variety of intellectual aids in the form of synonym dictionaries and thesauri hierarchical arrangements of subject identifiers, statistical and syntactic phrase-generating methods and the like, in order to obtain the content indentifications useful for the retrieval process.

[Salton, Gerard: The evaluation of automatic retrieval procedures-selected test results using the SMART system. Amer Doc 1965, 16(3), 209-22.

Current Research and Development in Scientific Documentation, No. 14, p.314]

**SWIFT-LASS** (Signal Word Index of Field and Title-Literature Abstract Specialized Search): A system established in 1966 at the Carborandum Co., N.Y., for listing and retrieval of abstracts of literature in the subject fields of applied statistics, precision analysis of tests, quality control, bulk material sampling and inter-lab testing. The medium of document storage is keysort cards. Indexing characteristics used are subject terms or descriptors, personal author, date of issuance of publication, and accession number. Descriptors used are words in title. Main equipment used are IBM 7090, 360 computers IBM reproducer, IBM interpreter, and UNIVAC SS II computer.

[Nonconventional scientific and technical information systems in current use, No. 4, p.269]

**SWIFT-SIR** (Signal Word Index of Field and Title-Scientific Information Retrieval): A system established in 1965 at the Carborandum



Co., N.Y., for storage and retrieval of letter reports, memoranda, and final reports of the mathematics branch covering the fields of applied mathematics and statistics, quality control, operations research and scientific computing. Media of index storage are standard punched cards, magnetic tape, and book or other page arrangement. Indexing characteristics used are subject terms or descriptors (which are words in title), date of issuance of publication, accession number, sequence number. Main equipment used are IBM 7090, 360, IBM reproducer, IBM interpreter, and UNIVAC SS II computer.

[Conventional scientific and technical information systems in current use, No. 4, p. 272]

#### TERMATREX and MINIMATREX system:

The two concepts have been developed and a line of equipment using the concepts have been produced by the Jonker Business Machines Inc. Termatrex systems are basically information-retrieval or 'search' systems, which works on the simple principle of super-imposition of characteristic cards over a light source (peek-a-boo system). Each 'item of information' is naturally identified by a serial number and on the Termatrex card it is represented by a hole which can be correctly interpreted by its position in relation to the X and Y coordinates. The basic limitation of the Termatrex system is that the capacity of a set of characteristic cards is limited to 10,000 items of information. For each additional 10,000 items of information, a complete additional set of cards is needed. However, when a collection becomes too large, the user can switch over to "Minimatrex line". This is done by photographic miniaturization which reduces Termatrex cards to an area 1/2" x 1/2" on strips of film.

[Jonker, Frederick: The new Termatrex line of I. R. systems - The Minimatrex line of I.R. systems. Amer Doc 1963, 14(4), 276-82]

**TABLEDEX index:** A coordinate index displayed in tabular form. Basically it consists of three parts - the first is a retrieval-word list of index terms used, the second is the TABLEDEX tables, or index proper, the third is the bibliography proper. The National Biomedical Research Foundation, Maryland has prepared computer programs to compile TABLEDEX index. A good example of TABLEDEX format is the 'Selected Bibliography of the International Geophysical Year'

published by the National Biomedical Research Foundation, 1962.

[CRDSD, No. 14, p. 135]

**UNIDEX:** A name given to a mechanized application of the UDC for index preparation. The system has been developed by the American Meteorological Society to speed up and improve indexing operations for their Meteorological and Geostrophysical Abstracts and the experimental Meteorological and Geostrophysical Titles.

[CRDSD, No. 13, p. 203.]

Rigby, Malcolm: Experiments in mechanized control of meteorological and geostrophysical literature and the UDC schedules in these fields. Rev int Doc 1964, 31(3), 103-106]

**WALNUT:** WALNUT is a code name for a complex, mechanized, micro-image storage and retrieval system (IBM 9603 Image File) developed for a federal agency by the IBM Corporation. This system consists primarily of a file of microfilm strips stored in bits and a mechanical selection device that can quickly go to a specified bit, mechanically select a strip of microfilm, and copy images from that strip into aperture cards. The basic image file unit contains a total of 990,000 page-size images (200 plastic cells of 50 strips each, with 99 images on each strip) stored at a 35:1 reduction. This is equivalent to about 3000 books or the contents of about 100 filing cabinets. Given an image number and page count, random-access selection equipment can locate the desired image in less than 5 seconds and transfer it to a blank frame of Kalfax film mounted in aperture card.

Another piece of equipment which accompanies the basic file unit is the IBM 9403 Image converter (This unit serves as the file input device by transferring images from conventional 35 mm sprocketed silver microfilm (at an original reduction of 15.8:1) to the kalfax file strips at a further reduction of 2.2:1. The file strips utilize kalfax film that has an image resolution on the order of 500 to 600 lines per mm. The Image converter operates at a maximum rate of 1500 frames per hour to transfer images to the file strips, assign addresses, and punch these addresses into a tab card.

A few of these WALNUT units have been developed, but to date only the Central Intelligence Agency is using them.

[Bradshaw, P.D. "The WALNUT system: A large capacity document storage and retrieval system. American Documentation 1962, 13(8), 270-5.

Veyette, J.H., Jr. "Photo-image storage: Its role in modern business" (includes a description of the IBM WALNUT system). Business Automation 1961, 6(4), 16-21]

## List of Terms Explained in the Glossary

ABC	MAC
ABCD	MARLIS
ABLE	MATICO
ACORN	MEDIA
ALP	MEDLARS
ALPHA	MELISS
ASCA	MERGE
	MINIMATREX [under TERMATREX]
BRALD	
ChemSEARCH	PACIR
CITIS	PHILSOM
COMEINDORS	PICS
COMIT	
CRIS	RAPID
DACOM	SABIRS
	SAID
FAIR	SATIRE
FLIP	SCAN
	SCISRS
GREMAS	SHARP
	SHIRTDIF
HAYSTAQ	SIDAR
	SMART
IDEEA	SWIFT-LASS
INTREX	SWIFT-SIR
KAPL	TABLEDEX
KWIC	TERMATREX
LINCO	UNIDEX
LOCUS [under LINCO]	
LITE	WALNUT