DOCUMENT RESUME

ED 060 875 LI 003 541

AUTHOR	Mikhailov, A. I.; Giljarevskij, R. S.
TITLE	An Introductory Course on Informatics/Documentation.
INSTITUTION	International Federation for Documentation, The Hague (Netherlands).
REPORT NO	FID-481-1971
PUB DATE	71
NOTE	201p.; (100 References), revised and enlarged edition
AVAILABLE FROM	International Federation for Documentation, 7 Hofweg, The Hague, Netherlands ($11.90)
EDRS PRICE	MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Courses; Developing Nations; *Documentation; Guides; *Information Science; *Information Scientists; *Information Seeking; *International Programs
IDENTIFIERS	*Scientific and Technical Information

ABSTRACT

The efficiency of scientific information activities depends on the availability of highly skilled and experienced personnel and on the familiarity of all scientists and engineers with the essentials of information science. In 1963, the International Federation for Documentation (FID) appealed to its member countries to promote the establishment of documentation and scientific information chairs at universities and other institutions of higher learning. Such a chair was set up at the Lomonosov University of Moscow in the academic year 1963/64. One of its purposes is to teach the students basic methods of scientific information work and to make them acquainted with the major tools used in this work. This Guide has been written on the basis of that course. The Guide contains: the text of the lectures, questions for self-checking, tests or examinations, lists of references suggested for further study, curriculum and syllabus of the lectures and practical lessons. The Guide is intended primarily for students in countries still lacking any regular instruction in the discipline, and the initial training of information officers in developing countries. (Author/NH)

AN INTRODUCTORY COURSE

ON

INFORMATICS/DOCUMENTATION

by

A. I. Mikhailov and R. S. Giljarevskij

Revised and enlarged edition

INTERNATIONAL FEDERATION FOR DOCUMENTATION
7 Hofweg, The Hague, Netherlands
FID 481, 1971

PREFACE

The scientific and technical revolution now under way is associated with a rapid growth of scientific information. The increasing number of scientific and other publications is the most noticeable feature of the phenomenon which is frequently described as "information explosion". Scientists and engineers naturally seek to obtain with minimum of effort as much information as possible in a form convenient for use. In response to this social need, a special kind of activity called scientific information, or documentation has emerged. Most countries have created ramified networks of information centres or documentation services which regularly provide the specialists with necessary information.

It is a specific feature of information activities that an overwhelming majority of specialists in science and technology take part in them. They record the results of their research and development work in papers, books, reports, projects, invention specifications, which are later used by other scientists and engineers as initial data for further research and development endeavours. There is a special category of scientific specialists - information officers - whose task is to regulate the flow of these documents and to mould the totality of scientific facts and data into forms convenient for use. The efficiency of scientific information activities depends on the one hand on the availability of highly skilled and experienced personnel and on the other hand on the familiarity of all scientists and engineers with the essentials of information science and on their ability to find and use scientific information accumulated through the ages by mankind's exploring endeavour.

In many countries training is provided for information officers and users through regular college and university courses. In 1963, the International Federation for Documentation (FID) appealed to all its member countries to promote the establishment of documentation and scientific information chairs at universities and other institutions of higher learning. Such a chair of scientific information was set up at the Lomonosov University of Moscow in the academic year 1963/64, one of its purposes being to teach the students basic methods of scientific information work and to make them acqainted with the major tools used in this work. A 24-hour optional course in information science is offered to the students in the natural science faculties; this Guide has been written on the basis of that course.

The Guide contains material for an introductory course in informatics/documentation: the text of the lectures, questions for self-checking, tests or examinations, lists of references suggested for further study, curriculum and syllabus of the lectures and practical lessons. The Guide is intended primarily for the students of institutions of higher learning in the countries still lacking any regular instruction in the discipline; it can also be used for the initial training of information officers in developing countries.

The Guide was written under a Unesco contract to FID. The authors wish to acknowledge their deep gratitude to the FID Secretariat and the staff of the Unesco Department of Documentation, Libraries and Archives for valuable advice and assistance rendered during the preparation of the manuscript. The principal ideas expressed in the Guide have been formulated by the authors in a monograph on the fundamentals of informatics ("Osnovy Informatiki", Moscow, Nauka Publishers, 1968, 766 pp.) jointly with A.I. Chernyi to whom they offer sincere thanks for the university lecture material placed at their disposal and the helpful ideas suggested by him in discussing those lectures. Thanks are also due to E. Azgaldov and K. Erastov who translated this work into English.

The authors hope that their Guide will prove useful both to the students being introduced into scientific information activities and to their teachers, and they would be grateful for any comments or suggestions.

CONTENTS

		Page
Preface		3
List of Figures		7
1.	Scientific Information Activities - Organic Part of Scientific Work	9
2.	Scientific Literature - Source and Means of Dissemination of Knowledge	22
3.	Information and Bibliographic Publications - Sources of Data on Literature	42
4.	Information Agencies and Special Libraries - Auxiliary Institutions of Science	63
5.	Essentials of Information Retrieval	77
6.	Conventional Information Retrieval Systems	86
7.	Descriptor Information Retrieval Systems	102
8.	Technical Facilities of Information Retrieval	117
9.	Reproduction of Scientific Documents	130
10.	Use of Scientific Information	143
Syllabus and Programme for the Introductory Course on Informatics/Documentation		155
Figures 1 to 40		159
Alphabetical subject index		199

LIST OF FIGURES

		See page
1.	Growth of world science periodicals	11
2.	Tasks and stages of scientific-information activities	13
3.	Types of scientific documents and publications	23
4.	Citation of science documents in later publications	36,59
5.	Distribution of new references to earlier publications in a hypothetical field	37,38
6.	Scatter of citations in an isolated special literature	38
7.	Half-life of scientific literature	39
8.	Age distribution of papers cited in publications of 1965	39
9.	Two different abstracts of the same paper as abstracted in the VINITI Abstract Journal "Chemistry" and the "Chemical Abstracts"	50
10.	Arrangement of entries for the same books in a classified, alphabetical subject and author order	51
11.	Basic types of bibliography	52
12.	Layout of entry in a permuted-title index (KWIC Index)	57
13.	Portion of a page in the Science Citation Index	58
14.	Channels of dissemination of scientific information from the generators to the users	63
15.	Distribution of scientific and technical information flows among the information agencies in the USSR	67
16.	Diagram of information retrieval	77,105
17.	Dependence of recall on precision	81
18.	General block-diagram of an information retrieval system	81
19.	Arrangement of bibliographical particulars on the title-page and in the entry	87
20.	Scheme of rules for description of the basic types of publications	89
21.	Specimen index entries	89
22.	Graph of an hierarchical classification and the relationship between its divisions	92
23.	Main divisions in the Universal Decimal Classification (UDC)	94
24.	Main divisions (continued), and synopsis of auxiliaries in the UDC	94
25.	Deciphering of a UDC number	95

		See page
26.	A hypothetical classificaton of objects according to attributes which do not form natural hierarchy	103
27.	Graphical representation of relations between sets	105
28.	Working principle of the Uniterm System	109
29.	Specimen layout of a subject class in a thesaurus	112
30.	Examples of codes for edge-punched cards	121
31.	The working principle of body-punched and superimposable cards	121,122
32.	Aperture cards	123
33.	Microfiches (specimens)	124
34.	Schematic operation and main units of Filmorex system	125
35.	Schematics of the basic copying processes	132, 134
36.	Relative advantages and disadvantages of the different copying processes	135,141
37.	Schematics of the different printing methods	136
38.	Block-diagrams of the different printing machine types	136
39.	Schematics of operation of the hectograph, rotary stencil duplicator and rotaprint	138,139,140
40.	Choosing means of document duplication	141

1. SCIENTIFIC INFORMATION ACTIVITIES ORGANIC PART OF SCIENTIFIC WORK

Science is a form of social consciousness that enables man to learn the objective laws of nature and society and to put his knowledge into practice. Science, too, is governed by its own laws of development which are necessary for successful work in any particular field of science.

Laws of the development of science

One of the principal laws is the differentiation of science and specialization of each scientific discipline. Learning has no limits, since the matter has an infinite variety of characteristics, types, forms of movement and their interrelations. The more subjects scientists study, the wider becomes their field of research. Scientists have to concentrate their efforts on studying ever narrower scientific fields. This inevitably leads to the division of science into separate disciplines.

Soviet Academician A.N. Nesmeyanov wrote in this connection: "Every field of science differentiates and gives rise to ever new disciplines. Chemistry, for instance, was divided already in the last century into organic and inorganic, analytical and, later, physical chemistry. In this century, physical, organic chemistry has emerged, and quite recently element-organic chemistry has become a discipline in its own right. In physics, the fields established as independent disciplines with their specific research methods are molecular physics, optics, acoustics, solid state physics, and - in the last decade - nuclear physics, etc." (1).

Differentiation and specialisation of science help the scientists penetrate ever deeper into the mysteries of nature and society, accumulate facts, and establish the interrelation of new phenomena. This process, however, conceals a danger of which scientists have long been aware. Nearly a century ago the famous German scientist R. Virchow wrote: "We who describe ourselves as explorers of nature have actually mastered only some particular field of it ... In all other fields we are half-ignoramuses." (2). In effect, differentiation leads to an ever greater isolation of scientists who gradually lose orientation in the general system of scientific knowledge and even within their own discipline find it increasingly harder to keep au courant with their colleagues' work.

At the same time, it is the interaction and interpenetration of sciences that are so particularly fruitful. Important trends of research often arise at the junctions of related, and sometimes even of remote sciences. This regularity of the evolution of science was pointed out by

---

1. A.N. Nesmeyanov. Na styke nauk (At the junctions of sciences). "Nauka i Zhizn", 1958, no. 3, p. 1 (The VINITI system of transliteration is used throughout this volume)
2. R. Virchow. Die Freiheit der Wissenschaft im modernen Staat. Berlin, 1877, p. 13

F. Engels who in his Dialectics of Nature rightly predicted that it was precisely at the points of contact of sciences that the greatest results were to be expected. (1) The emergence and rapid development of biochemistry, biogeochemistry, biophysics, physical chemistry, bionics, cybernetics, mathematical linguistics, engineering psychology and other scientific disciplines are sufficient proof of this statement. It is evident that to ensure interaction of sciences in their continuing differentiation it is necessary to establish reliable channels of communication between the scientists working in different disciplines.

Another important rule in the development of science is its historical continuity and international character. If every scientist and the scientists in every country and era had themselves alone to accumulate the necessary knowledge and to rediscover laws, science would have hardly made any progress. The accomplishments of scientists the world over and of the previous generations constitute the basis for the work of contemporary scientists. K. Marx wrote: "Universal labour is all scientific labour, all discovery and all invention. This labour depends partly on the cooperation of the living,and partly on the utilization of the labour of those who have gone before." (2)

It is natural, therefore, that every generation of scientists is engaged not only in originating new scientific data, but also in special activities of classification, evaluation and generalization of the data accumulated earlier by their colleagues to make these data accessible not only to their contemporaries but also to future generations of scientists. Obviously, this accessibility largely depends on the form in which scientific knowledge is disseminated among the contemporaries and handed down to posterity.

The theory and methodology of science are beyond the scope of this Guide; these problems are the concern of a new scientific discipline, the science of science. We are not going to consider all the rules of science, but shall only dwell on those relevant to our subject. However, there is one more rule that deserves to be pointed out here. It is the accelerated growth of science.

It has become a newspaper and radio cliché that the world is going through an era of scientific revolution, a powerful explosion of scientific research. Every schoolboy nowadays knows that science is developing at a stupendous pace and that before he has time to go over to the next grade in his school, developments will take place which have previously taken decades and centuries to accomplish. Not everybody realizes, however, that a similar acceleration of scientific development has also characterized the past several centuries.

It may be asked: how can one measure the growth of science? Aren't some major discoveries more valuable than hundreds of minor ones? Indeed,

---

1. F. Engels, Dialectics of nature. 3rd rev. ed. Moscow, Progress Publishers, 1964, p. 297
2. K. Marx, Capital. A critique of political economy. C. Vol.3, book 3. Ed. by F. Engels. 3rd imp. Moscow. Progress Publishers, 1966, p. 104

so far in making a quantitative assessment of the rate of scientific development we can only use such indirect indicators as allocations for scientific work, number of employees in the field of science, or the number of scientific publications. However, approximate the data upon which such measurements are based, it is sure evidence that each of these indicators doubles its numbers in equal periods of time.

Let us illustrate this by examples. In 1800, there were about 1,000 scientific workers in the United States; in 1850 their number rose to 10,000; in 1900 to 100,000; and presently their number is in excess of 1,000,000. In 1800, there were about 100 journals in the world; in 1850, there were 1,000 journals; in 1900, the number rose to 10,000; and by the 1950's, close to 100,000. In both cases a 10-fold increase took place after every fifty years. In mathematical terms such growth is described by an exponential dependence and can be graphically shown by an exponential curve.

The wellknown historian of science Derek de Solla Price has drawn up a table showing that while the world population doubles every 50 years, the percentage of new scientific workers per 1,000 inhabitants doubles every 20 years, the number of qualified scientists and of scientific journals doubles every 15 years and the literature on certain subjects, for example X-rays or experimental psychology, every 10 years. He made quite a clear picture of the rate of growth of science in the following words: "To put it another way, using any reasonable definition of a scientist, we can say that 80 to 90 percent of all the scientists, that ever lived are alive now. Alternatively, any young scientist, starting now and looking back at the end of his career upon a normal life span, will find that 80 to 90 percent of all scientific work achieved by the end of the period will have taken place before his very eyes, and that only 10 to 20 percent will antedate his experience". (1)

The accelerated growth of science accompanied by frequent doubling (every 15 to 20 years) of scientific literature has for a long time been causing apprehension among scientists. Written records - journal articles, books, and other publications in which the scientists describe the findings of their research - are the basic means for the conveyance of scientific knowledge in time and space. Until quite recently, it has been the system of written records and publications and the simultaneously developed system of library and bibliographic services provided for scientists, that have been ensuring the historic continuity, international nature and accelerated growth rate of science. This system, however, no longer satisfies the scientist.

The apprehension was visualized by the late President of the USSR Academy of Sciences S.I. Vavilov in the picture of a scientist standing perplexed before the "Himalayas" of libraries unable to extract from them

---

1. D.J. de Solla Price, Little science, big science, New York - London, Columbia University Press, 1963, p. 1-2, 6-7

the grains of gold he needs (1), while the English physicist J.D. Bernal has said that it is sometimes easier to rediscover a phenomenon than to learn from literature that it has been discovered earlier. (2) Scores of similar statements by other scientists can also be cited. Here is a cri-de-coeur from the well-known French physicist Louis de Broglie: "A scientist often feels himself buried under the heaps of articles and monographs appearing all over the world; for all the assistance rendered by the bibliographies, he as often as not cannot manage to read them through, to say nothing of thinking them over. Flooded by the unceasing stream of publications, he is under the constant risk of getting mazed in trifles and missing the important things" . (3)

Social demand for scientific information activities

Let us take a look at the situation from yet another standpoint. Since a scientist who wants to work efficiently must know what has been done earlier in his field and what other scientists presently working on the same or related subject are doing, he has to spend part of his time searching for this information. Surveys have shown that many research experts in the natural sciences spend on this one-third or even a half of their working time. This naturally means reduction of the time spent on actual research, as well as decrease in efficiency of scientific work. And even then a scientist can never be sure that he has looked through all the relevant material.

The results are deplorable: quite a number of research projects are carried out to no purpose since they duplicate what has been done earlier by other scientists. Huge amounts of materials and intellectual resources are wasted. Many examples of this can be given. Several research laboratories in the USA spent 5 years and $ 200,000 on a project, which had earlier been carried out in the USSR and the results of which had been published in a journal. It is estimated that unnecessary duplication costs approximately 10% of all money spent on research and development.

All these facts have been mentioned with the sole purpose of showing that, at the present stage of science and technological development, there has arisen a social demand for a special type of activity to provide scientists and engineers with the necessary knowledge, or, as we now call it, scientific information.

And such scientific information activities have evolved. There are thousands of large and small establishments in the world called information or documentation centres, services, institutes, bureaux, or departments of scientific and technical information. Vast networks of these institutions providing regular assistance to research and development groups and teams are functioning in the majority of developed countries. The basic principles of the organization of such networks and the activities of the main information centres in the Soviet Union will be described later on.

---

1. S.I. Vavilov, Neskol'ko zamechanij c knigakh (Some notes about books). Sovetskaja kniga, 1967, no. 1, p. 15.
2. J.D. Bernal, Science in history, London, Watts, 1954
3. L. de Broglie. Sur les sentiers de la science. Paris, Michel, 1960, p. 340

It is now necessary to understand some general principles pertaining to scientific information activities. First of all, it should be clearly understood that these activities have emerged as an altogether natural phenomenon in the course of the development of science. In effect, differentiation and specialization in science refer not only to the division of the fields and disciplines, but also to the division of the types of work performed by different scientists. It is not so very long since each scientist used to do everything required for his research himself. He conducted observations, proposed hypotheses, tested them in experiments, searched for the data required and then classified and generalized it.

As time passed, experimental techniques became so complex and the methods used came to require the application of such sophisticated mathematical tools that experimentation was established as a speciality in its own right. This process took place at different periods in different sciences, and it is not yet completed in some of them. As Soviet Academician P.L. Kapitsa said: "We know well from the history of physics that the division of the physicists into theoreticians and experimentalists has taken place quite recently. In the past not only Newton and Huygens, but even such theoreticians as Maxwell used to test their theoretical conclusions themselves. Nowadays it is only in exceptional cases that a theoretician has to stage experiments to test his theories". (1)

Now the turn has come for scientific information search and retrieval. Scientific information acitivites, along with theoretical and experimental work, became a discipline in the course of the social division of scientific work. A part of scientific workers now devote all their efforts to the gathering, critical analysis and generalization of all the available scientific knowledge pertaining to a certain subject field or a scientific discipline. Scientific information activities today require special knowledge on the part of scientists and application of complex techniques and equipment. The task has no longer to be shouldered by individual scientists, as there are specialized teams of scientists to cope with it.

Tasks and stages of scientific information activities

What are the tasks facing information workers and into what stages must these accordingly be broken up?

1.First, it is necessary to collect all the required scientific information as exhaustively as possible. This means that from the tremendous amount of scientific knowledge one must select and regularly update only the information that might be of use for scientific or practical work in one or several disciplines, or even a single R & D team or establishment or a group of these.

2.The second step is analytico-synthetic processing of scientific information. This means that every group of data recorded in a scientific document is to be analysed from the standpoint of its contents while the document is to be analysed from the standpoint of its form. The results

---

1. P.L. Kapitsa. Teorija, eksperiment, praktika (Theory, experiment, practice). Moscow, 1966, p. 16

of such analyses are briefly written down, synthetized with the aid of special notations developed for the purpose of scientific information and based on the words of natural language, letters, figures, formulae, tables, and diagrams. Information thus processed becomes suitable for subsequent storage, retrieval and use in scientific work.

3. The third task is long-term storage of scientific information. The objectives here are to ensure that the data gathered take as little storage space as possible, and that the media upon which it is recorded provide for a maximum storage period and are easily arrangeable into an orderly system enabling fast and easy retrieval of all relevant data.

4. In connection with the previous task, the scientific information worker can be faced with another more complex job - logical information processing of collected data. The results must answer questions which have not been foreseen during the collection of the data and which are not directly contained in the data collected. For example, if the collected data concerns physical properties of a specific substance in a given environment, automatic logical information processing of this data (without substantial participation of an expert) can give new information concerning physical properties of this substance in a different environment.

5. Information retrieval - which is a major aspect of information work - consists of a series of logical operations aimed at finding the information a scientist needs. Any procedure used to accomplish that task must be such as to allow at every point to retrieve from a collection of data all and only those items of information which provide a direct answer to the question posed by any member of a scientific team.

6. Since the results of each of the above mentioned processes (and especially of information retrieval) may be of interest to a great number of scientists and practitioners simultaneously, the task of scientific information dissemination ensues. Scientific documents (or their parts) containing the relevant information must be reproduced in sufficient numbers of identical copies for speedy dissemination among all those interested.

7. Finally, it is not unimportant to scientific information activities how their results are used. For that reason the objectives of these activities include the popularization of the accomplishments of science and technology and of the most efficient techniques of handling scientific and technical documents, the study of the efficiency of the utilization of these documents, and of the impact of information work upon science, technology, and the national economy as a whole.

The fulfilment of the tasks enumerated and the performance of the corresponding stages of work require the elaboration of the theory, methodology, organization and technical means of scientific information activities. These are the functions that constitute the subject of a new scientific discipline that we propose to call "information science". (1)

---

1. A.I. Mikhailov, A.I. Chernyi, R.S. Giljarevskij. Informatics - new name for the theory of scientific information. "FID News Bulletin", 1967, V. 17, no. 7, p. 70-74.

Basic concepts

Before proceeding further it is necessary to introduce and define some basic concepts. These are first of all the concepts of "information" and "scientific information" as well as the concepts derived from them: "scientific information activities", "informatics", "information officer" and "information scientist".

Information is "certain knowledge, a totality of some data, and known facts." (1) In a broader, philosophical sense information can be defined as the contents of the relation between interacting material objects which manifests itself in a change of state of these objects. Since all material objects can be divided into inorganic, organic and thinking objects (human), three kinds of information are accordingly distinguished: elementary, biological, and logical (semantic) information. Logical information is proper only to human society and its contents are ideas and images. Semantic information is realized through man's oral and written speech.

Scientific information is logical information obtained in the process of cognition adequately reflecting the laws of the objective world and used in socio-historical practice.

The attributes contained in this definition need to be explained. When speaking of the process of cognition by which scientific information is obtained, it should be understood that this process is based on practice, on all kinds of the active efforts of mankind to transform nature and society, and not merely on scientific research and development. However not all information obtained in the process of cognition of the world around us is scientific. Sensual cognition gives a man a notion of only particular external aspects of objects. He can comprehend their inner nature and interrelations only through logical thinking embodied in a verbal form.

The adequacy of reflection of the laws of the objective world is to be understood as the degree of correctness of this reflection which is determined by the level achieved by science at a given time. Scientific information also embraces hypotheses and theories that prove erroneous after a systematic study and practical testing. The use of scientific information in socio-historic practice is likewise an indispensable condition, since this is the only safeguard that its truth will be verified. The use-in-practice criterion precludes regarding as scientific information common truisms or obsolete facts and other non-scientific knowledge.

Thus, "scientific information" is a generic term, and the word "scientific" in this term does not mean that such information is the result of only pure scientific work. Its specific types relevant to the fields of science are - biological, physical, chemical information, etc., but relevant to the fields of practical work - technical, agricultural, medical, political, managerial, etc.

---

1. Philosophical dictionary. Moscow, 1963, p. 172

Scientific information work is a separate part of scientific work, which has become such in the course of the social division of this work. Its purpose is to provide the scientists and practitioners with all the information they need by means of collecting, analytico-synthetic and logical processing, storage, retrieval and dissemination of scientific information.

A specialist in a field of science or social practice, whose sole occupation is information work, is usually called an information specialist or officer.

As will be seen later, some functions and processes of scientific information work are akin to librarianship and bibliographic work, which have existed for thousand years. In effect libraries, too, collect literature, process and store it and serve their users. This fact has given rise to the widespread opinion that scientific information work has tended to replace library work, since libraries allegedly are no longer able to serve science in our days. We shall devote a special chapter to a comparison of the activities of information centres and libraries, but would here only warn that the opinion is incorrect.

Scientific information work constitutes that part of scientific work which was formerly done by every scholar on his own. Of course, scientists and practitioners have been making use of libraries and bibliographic aids in their research work since days of yore. As scientific information work is developing, this not only maintained its importance, but is likely even to expand in scope and become more complex. That is why we shall pay due attention to those aspects of library and bibliographic work that have retained their importance for meeting the information needs of scientists.

Informatics is a scientific discipline which studies the structure and characteristics (but not the specific contents) of scientific information, as well as the rules of scientific information work, its theory, history, methodology and organization.

The purpose of informatics consists in developing optimal methods and means of presentation (recording), collection, analytico-synthetic and logical processing, storage, retrieval and dissemination of scientific information. Information science deals with semantic information, but is not involved in qualitative evaluation of this information. Such an evaluation can be carried out only by specialists in particular fields of science.

It should be mentioned that the term "informatics" is quite new and has not yet received a common acknowledgement. It has been suggested by the authors of this Guide jointly with A.I. Chernyi to replace a number of existing terms that we deem unhappy. The oldest and most common of these is "documentation". It has been proposed for the designation of a specific activity of collection, processing, storage, retrieval and dissemination of documents, as distinct from librarianship and bibliography, back in 1905 by Paul Otlet, a well-known Belgian scientist and public figure. In 1931, it was incorporated in the name Institut International de Documentation which grew out of the Institut International de Bibliographie (established in 1895), and in 1937 in the name of its successor, the Fédération Internationale de Documentation.

In Soviet professional literature a related term has been used rather frequently - "documentalistics". For example, "Nauka" Publishing House published in 1966 a collection of articles entitled "Kibernetika i dokumentalistika" (Cybernetics and documentalistics).

We believe that both these terms - "documentation" and "documentalistics" - are inadequate, for they lay the stress on documents whereas the subject studied by the discipline in question is the structure and characteristics of scientific information. Nor can we call adequate the term "theory of scientific information", which is also sometimes used in that sense, for the discipline embraces not only the theory, but also the methodology, history and organization; moreover, what is meant by "scientific information" in that term is actually "scientific information work". All these reasons prompt us to urge the introduction of a new term - "informatics".

Specialists in this new discipline - informatics - likewise must have a name. Our suggestions is "information scientists".

Interrelationships between informatics and other disciplines

Informatics, being a new scientific discipline, is based on numerous other disciplines and makes use of their methods. We must say at least a few words on these related disciplines, which are quite diverse: mathematical information theory, cybernetics, semiotics, linguistics, psychology, library science, bibliography, book science, science of science, and several technical disciplines.

The concepts which are central to *mathematical information theory* are the concepts of information and of the measure of its quantity, while its major objective is to study the process of the transmission of information through a communication channel. This theory derives its definition of information from statistical reasoning and completely disregards the contents, or semantic value, of the messages transmitted. Nevertheless, the process of storage, retrieval and dissemination of information, considered in sufficiently general terms, may be called information transmission through a communication channel, which explains why informatics uses so many concepts of information theory.

Cybernetics has been defined as the science of "the modes of perception, storage, processing and use of information in machines, living organisms, and associations of these". (1) There is almost a verbal coincidence of the first part of this definition with the stages of scientific information work described above. Another factor linking informatics and cybernetics is that at present the efficient solution of the main problems of informatics - retrieval and logical processing of scientific information - is based on electronic digital computers, i.e. cybernetic devices.

---

1. A.N. Kolmogorov. Kibernetika (Cybernetics). In: Bol'shaja Sovetskaja Entsiklopedija, 2nd ed., vol. 51, Moscow, 1958, p. 149

Semiotics, or the general theory of symbolic systems, which is sometimes regarded as a part of cybernetics, is currently being developed at the interface of mathematical logic and structural linguistics. Even the little so far reported about analytical and synthetic information processing will suffice to make clear the importance for informatics of artificial formalized languages with special symbols or of programming-type languages. The wide scope of subjects studied by semiotics includes such aspects of importance for informatics as: the principles of construction of artificial languages, the procedures of translation from a natural language into an artifical one and vice versa, and the notation systems used to record scientific information used in different disciplines. Methods of mathematical logic are likewise used by informatics for formalizing the logical inference procedures applied in different sciences; this would allow the automation of a great number of functions of scientific information work.

Linguistics, which is concerned with the study of the rules of evolution of the natural languages, is closely connected with semiotics in regard to the machine translation problem - a problem of great importance for informatics. Other fields of linguistics, such as the general theory of translation, the principles of transcription and transliteration, and the principles of the construction of terms, also find certain applications in informatics.

Psychology, especially one of its recent trends - engineering psychology - places at the disposal of informatics very valuable methods of research: optimal forms of scientific information most convenient for perception and use by the user, efficient reading techniques, construction of man-machine systems and a great many other problems associated with the human aspects of information activities.

Library science, which studies the essence, organization and methods of public use of written and printed records, has historical ties with information. For thousands of years the techniques of information work individually carried on by every scientist as part of his work were his personal affair - never generalized and at best passed down directly from teacher to pupil. The only social institution where these techniques could find at least an indirect reflection was the library. Library science has managed over the last century and a half to generalize some of these techniques. Although as yet it remains a descriptive science and certainly lags behind in solving the pressing problems of service to science, some of its concepts are used in informatics.

Bibliography is "the field of knowledge and of practical activity which has as its aim the registration, description, classification and qualitative analyses of publications and the compilation of various aids which help to find the way in the current literature, popularize it and promote its efficient use". (1) The scope of bibliography is much broader than that of library science and encompasses virtually all the aspects of human activities associated with written and printed records. Alongside

---

1. E.I. Shamurin, Slovar' knigovedcheskikh terminov (Dictionary of library science). Moscow, 1961, p. 29

scientific and industrial purposes, it provides for educational and training objectives as well. Supplying information on printed and written records to scientists and practitioners, however, remains one of the most important tasks of bibliography, and in this respect it is close to informatics and information work.

Book science is a complex discipline that studies printed and written records from the theoretical and historical points of view. Since scientific information is at present disseminated in the form of publications and manuscripts, many aspects of book science are of significance for informatics, especially the history of books, book statistics, publishing (in particular its theoretical and practical aspects) and the graphic arts industry.

Science of science (organization of scientific work) is quite a recent scientific discipline which studies the rules of development of science and elaborates on this basis the principles of organization, planning, financing and management of science. Some of the problems explored by the science of science are also important for informatics: these include the problems of the optimal system of scientific publications, the improvement of the efficiency of scientific work, forecasting the development of science and the assessment of its level and development rate.

Finally, the *technical sciences* provide for the elaboration of new techniques which is a sine qua non for increasing the efficiency of information work. It is difficult even to enumerate here all the fields of technology that play a role in developing special information devices. Microphotographs and records on various magnetic carriers are used for long-term storage of scientific information. Photography and electrical engineering, optics and electronics contribute towards solving this problem. Considerable progress has been made in scientific document reproduction techniques. Some new processes, for instance electrography, considerably increase the efficiency of information dissemination. In information storage and retrieval, great hopes are placed on electronic digital computers. Engineers are presently striving to build special input and output devices for computers so that they may be used in automating the main routines of scientific information work.

So much for the range of sciences connected with informatics in its main aspects. This range is of course even broader, but this general outline is enough to indicate the numerous and close links of informatics with modern science and to delineate its scope of research.

On the contents of these lectures

We should like to conclude with a word on our lectures: they should not be taken as a regular course, expounding the fundamentals of informatics; their purpose is to serve only as an introduction to this discipline.

Both the structure and contents of this course are determined by that particular purpose. We shall first acquaint readers with the existing system of scientific publications, primarily those providing aid in searching for the pertinent scientific literature. We shall speak of the activity of those institutions - scientific information centres and special libraries - that function as intermediary between the scientist and the literature, besides storing scientific information.

The central place in the lectures is given to the principles of information retrieval: both conventional and non-conventional systems, as well as tools for their implementation, are considered. Major attention is given to the simplest means which can be available to any student.

The course ends with a lecture on the means and methods of utilizing scientific and technical information and the personal information work of the individual scientist.

Every lecture concludes with check questions and references for further study.

Questions for self-checking

1. What are the main rules of the development of science?
2. What are the indicators that characterize the accelerated development of science?
3. Which social demands have given rise to scientific information work?
4. What is scientific information work, what stages does it consist of, what are the problems it strives to solve?
5. What are the most important characteristics in defining scientific information?
6. What is informatics?
7. What scientific disciplines is informatics related to?

Literature

1. H. Borko. Information science: what is it? "American Documentation", Vol. 19, No. 1, Jan. 1968, p. 3-5.
2. Ch. P. Bourne. Methods of information handling. New York, a.o., Wiley, 1963, XIV, 241 p. , ill.
3. A.I. Mikhailov, A.I. Chernyi, R.S. Giljarevskij. Informatics: its scope and methods. In: On theoretical problems of informatics. FID 435. Moscow, All-Union Inst. for Scientific and Technical Information , 1969, p. 7-24.
4. D.J. de Solla Price. Little science, big science. New York - London, Columbia Univ. Press, 1963, 119 p.
5. J.M. Ziman. Public Knowledge. An Essay concerning the Social Dimension of Science. Cambridge, Univ. Press, 1968, XII, 154 p.

2. SCIENTIFIC LITERATURE - SOURCE AND MEANS OF DISSEMINATION OF KNOWLEDGE

"If the history of science and technology is considered from the standpoint of the level of means of information transmission and storage, three major eras can be more or less clearly distinguished thus far: the first, and the longest, when oral speech was the main, if not the only, means of information transmission; the second, which began with the emergence of a class society, ... characterized ... by the addition of writing to speech; finally, the third - approximately from the establishment of the bourgeois society up till now - with book printing as the major means of storage and transmission of scientific and technical information." (1)

Stages of development of written records

Since the methods of storage and transmission of semantic information have played a major part in the development of human society, it will be interesting to trace the evolution of these methods. Speech proved inadequate for the task of information transmission at a certain stage of development when the need arose to convey messages over large distances and to preserve them for prolonged periods of time. Written language made its appearance as a system of physically fixing speech sounds and words using conventional signs for their transmission in time and space.

Every civilization, as it perfects the oral and written language in the course of its development, works out certain types of documents which, along with oral communication, serve to convey knowledge about the surrounding world and the skills in this or that handicraft or industry. The oldest scientific documents that have been preserved until now are the tablets inscribed with cuneiform writing from Mesopotamia which date back to the 4th millennium B.C. Treatises in astronomy, geography, history, commerce and law have been preserved since the 4th millennium B.C. (the Sumerian and Assyrian-Babylonian cultures which flourished in the 2nd millennium B.C.). Along with small clay tablets 2.5 cm high were produced large plates 40 cm high which contained up to 400 lines written in 4 columns on each side. These books of clay served human cultural and scientific needs for thousands of years, but their major shortcomings were unwieldiness and insufficient capacity. Some texts required up to 10 tablets to record, and the collections numbering tens of thousands of tablets - remnants of such libraries have been preserved till our time - took enormous storage space to accomodate.

Papyrus scrolls were a more compact form of document which made it possible to accumulate vast collections of manuscripts. Beginning with the 3rd millennium B.C., this was the usual form for books in Egypt. In a scroll, the text was arranged transversely to the length in columns of 25 to 45 lines. The fragility and impermanence of papyrus are to blame for the scarcity of the specimens of ancient Egyptian documents preserved, the oldest

---

1. O.S. Kulagina, L.A. Kaluzhnin and V.V. Ivanov. O kul'turnom znachenii mashinnogo perevoda (On the cultural significance of machine translation). "Vestnik istorii mirovoj kultury", 1961, N 3(27), p. 23

of which dates from the 18th century B.C. Most of these are the ritual "books of the dead" unearthed by the archaeologists from the pyramids and other burial places. In Greece and Rome, scientific knowledge was for a long period disseminated through oral speech. It is known that manuscripts were not used for the purpose until after Aristotle. Not a single original text has been preserved since classic times, and often centuries separate an earliest preserved copy from the conjectured date of the original text. What is known about Greek and Roman papyri is based on Asia Minor and late Egyptian specimens. The average length of a scroll was not more than 10 m, and the width 30 cm.

Parchment made of calfskin has been known as a writing material since the 3rd century B.C. It allowed writing on both sides, and was preserved longer than papyrus. It determined the transition to the codex, a modern book form. Papyrus codices were manufactured till as late as the first centuries A.D., but by the 5th century A.D. they were superseded by parchment codices. The codex is a book form of a greater capacity than a scroll, better suited to the recording of long texts and to making references. Up till the 12th century, parchment codices were the only book form that existed in Europe. Their biggest drawback was the high cost, which prevented a wide spread of written documents. A whole herd of cattle had to be skinned to make a single copy of a parchment book.

Parchment was superseded by paper, which was invented in China in 105 A.D. and first manufactured in the Middle East in 751 A.D. In the 12th century use of paper penetrated Europe with the Arab conquerors of Spain and paper books, which were considerable cheaper than parchment books, came to be widely used for scientific and educational purposes. The several thousand years of the evolution of written documents show that the form of the latter has changed under the pressure of the needs of society: documents became an ever cheaper, more capacious and convenient medium for information recording. This resulted in the emergence of the book form, or codex, which has remained the main form of document till our day.

A widespread need for a new technique of document production began to be felt in the 15th century, when handwritten books were no longer sufficient for the cultural and scientific demands of society. There were two basic reasons for this: firstly, they were too costly and laborious to make; secondly, the manual copying of texts could not provide a sufficient number of identical copies, since the copiers frequently altered the text. It was at this time, between 1430 and 1440, that the German craftsman J. Gutenberg invented printing from movable metal type. The engraving and moulding techniques on which book printing was based had been known since antiquity, as also the wine press that Gutenberg modified to build his printing press. But is was only the urgent demand of society for rapid and exact mechanical reproduction of texts that called forth this invention, which still remains for us an object of unceasing admiration and the prototype of the techniques we continue to use for producing printed documents.

Scientific documents and their types

Documents, in a broad sense, include not only inscriptions, manuscripts, and printed publications, but also works of art, numismatic items, and museum exhibits of mineral, botanic, zoological and anthropological nature.

Any material object that records or confirms some knowledge and can be included in a certain collection may be regarded as a document. A scientific document is a material object carrying scientific information, serving as a record designed to transmit this information in time and space, and utilized in social practice.

Different types of scientific documents appeared in different eras, and during the last centuries - and even decades - they have been undergoing substantial changes. Books have existed for several thousand years, patent specifications have been known for some five hundred years, scientific periodicals are but 300 years old, while journal articles in their present form are even younger - 100 to 150 years old. Book science as yet lacks a well articulated typology of scientific documents. The distinction between published and unpublished documents until quite recently has been considered to be the basic division. Even a few decades ago, ideas and facts were recognized to have entered scientific usage only after their publication, which then meant a more or less extensive and official registration of the documents which contained them. This distinction is now of less importance for scientific information activities since, firstly, unpublished documents contain quite a lot of valuable scientific information in advance of published material and, secondly, present reproduction facilities make the division into published and unpublished documents rather arbitrary. Documents such as scientific and technical reports, dissertations, translations of scientific communications, which are usually considered unpublished, are now quite often disseminated in hundreds and thousands of copies.

Informatics has highlighted the division of documents into primary and secondary, a division which is also quite arbitrary and approximate as it largely refers not to documents but to the scientific information they contain. It is accepted that primary documents record the immediate results of scientific research and design and development, whilst secondary documents deal with the results of the analytico-synthetic processing of scientific information contained in the primary documents. In fact, however, the historically evolved system of scientific documents is such that many of these contain both the results of research and of the processing of information taken from earlier publications. Articles in scientific journals, monographs, textbooks, and especially, reference literature may be cited as examples.

This division is convenient, nonetheless, as it allows us to characterize the different flows of scientific documents in information work. We shall follow it further on, considering primary documents as those which mainly contain new scientific knowledge or new understanding of known facts or ideas and secondary documents as those which mainly serve to give information on primary documents. Taking into account the above reservations, one may consider as primary documents and publications the majority of books (excluding reference literature), periodical publications, special types of technical publications, scientific and technical reports, dissertations, translations, information cards, and as secondary documents - reference literature, surveys, abstract journals, library catalogues, bibliographic indexes and card files.

Leaving secondary documents and publications to be discussed in the next chapter, let us consider briefly the main types of primary documents and publications.

Book is an ambiguous term. With reference to its contents, one can speak of a book as a scientific, practical or literary (artistic) piece; if the form alone is taken into consideration, then any printed or written reproduction of text or graphic figures may be considered a book. Informatics defines 'book' as a nonperiodical multi-page printed item. "The Recommendations on the International Statistical Indicators Characterizing the Publishing of Books and Periodicals", approved by Unesco in 1964, provide a clearcut distinction between books and pamphlets. Book is a nonperiodical printed publication of no less than 49 pages, excluding the cover and the title page. Pamphlet is a nonperiodical printed publication of not less than 5 and not more than 48 pages.

As mentioned earlier, the major advantage of book printing over handwritten books is the capability of manufacturing any required number of identical copies exactly matching the original. In addition, in most countries a printed book is subject to special registration at the moment of publication, which ensures keeping an accurate record of all publications. These advantages have made printed books the most important type of scientific document playing a major role in the development of science and technology.

It was only in recent years that scientists began to voice their dissatisfaction with the existing system of dissemination of scientific knowledge in which books occupy the key position. This discontent stems from a number of circumstances, including a long time lag between the writing of an original and its publication, the inaccessibility of a great number of publications to many scientists and specialists, and the difficulty of orientation in the ever growing book flow. On the other hand, the principal features of printed records - black text against light-coloured background, the sizes of print, lines and pages, and the codex form - continue to satisfy the most exacting reader today just as they did centuries ago, and perfectly conform to the faculty of the contemporary reader for mastering knowledge through reading. This gives reason to suppose that publications in their traditional forms will survive for quite a long time to come. As to the scientists' discontent with books, which is certainly justified in part, it means that the system of publication and dissemination of books should be improved.

The many operations involved in manufacturing a printed book require a lot of time. Bearing in mind that publishing houses and printing offices usually work on a great number of publications simultaneously, each of which is to be printed in thousands of copies, it will be clear why books often remain for months in press. This is one of the most serious drawbacks of books as a means of disseminating scientific information. Information contained in books often dates back several years and is of faded interest for current scientific research.

For this reason books are less important in scientific information work than are other scientific documents: and yet a book of science is an important means of generalization of scientific knowledge. Books contain major generalizing theoretical studies and studies of crucial problems in science, national economy and culture. They are an unrivalled means of educating, teaching and training new generations of scientists and practitioners. The following book forms may be considered to be of prime importance for scientific information work: monographs, collections of papers, scientific conference proceedings, textbooks and manuals, and official publications.

A *monograph* is a work of science or popular science that exhaustively and comprehensively treats a specific problem, question, or subject. A monograph is usually written in accordance with a well elaborated plan covering all aspects of the subject or phenomenon under consideration; it may be written by one or several authors. The monograph was the predominant type of scientific book in the last century, succeeding the treatise which had prevailed in the literature of science during the previous era.

Presently, monographs have yielded priority to *collections of papers*, i.e. books consisting either of separate papers by the same or different authors, or of official, historical, archival or other documents and records. Collections of scientific papers written by different authors have acquired particular importance in the postwar years. In contrast to a monograph, a collection of papers, even compiled according to a definite plan, need not exhaust the entire subject, but may touch on only some of its aspects. Collections, which take less time to compile than monographs to write, have come to be the most wide-spread type of scientific book, and the importance of the data they carry is so great that no specialist can disregard them in his work.

Proceedings of conferences, congresses and symposia are as a rule published after the end of these meetings. Along with the resolutions adopted , they contain the abstracts and sometimes complete texts of the reports delivered, as well as summaries of discussions. More than half the communications published in such books do not appear in other publications, e.g. journals. Hence their great value for scientific information work.

Textbooks and manuals hold a special place among other books used in scientific work. They describe only the principal facts and features of a topic and are mostly intended for instruction. However, the most original of them are also of interest to specialists, for their novel groupings and generalizations of known facts. Moreover, textbooks of allied disciplines can help a specialist to familiarize himself with the range of problems in a field not within his immediate interest, and also often serve to set standards for term usage.

Official publications, issued on behalf of organizations and institutions, are documents carrying information directly related to the activity of the institutions responsible. Many of them have no scientific value, but some contain important scientific information, e.g. reports, plans and reviews describing the work of scientific and other institutions, their resolutions, decisions and regulations, intra-agency norms (some of these are considered as special types of technical publications), industrial and engineering instructions, reference information, methodological and statistical documents, training curricula and syllabuses.

As mentioned, all these types of books need not necessarily be in single volumes; they may also be multivolume works or parts of a series. A "multivolume publication" is a set of printed items contained in two or more volumes, books, parts or issues that form a distinct unit. A "series" consists of a number of independent publications usually characterized by common or closely related subject content and issued by the same publisher under a common serial title. In contrast to the multivolume sets, the issues of a series may be unnumbered.

Over the past few decades major changes have been unfolding in book publishing that have also affected the sphere of scientific books. The so-called paperbacks - cheap pocket-size books printed in millions of copies - have included books of science, popular science and textbooks and this has prompted some bibliologists to speak of a revolution in the book industry. The picture would be incomplete without a word on the number of scientific books. Unfortunately, exact data is lacking on the number of books that have been published since the invention of book printing, but it may be roughly estimated at 30 million titles. More precise are the figures for the present annual book output of the world. According to Unesco, in 1952 it totalled 250,000 titles; in 1962, 385,000; in 1963, 400,000; in 1964, 415,000; and in 1967, 450,000 titles. Considering that the books printed for sale constitute but a part of the world book output, one may suppose that the true figures would be double those cited, though the figures are impressive enough as they are. An annual 450,000 titles means a daily batch of 1,200 titles, or about 50 titles every hour; in other words, a new title appears every minute round the clock.

According to some national statistics, scientific books account for 1/5 to 1/4 of the total book market. One can thus speak of an annual increase of close on 100,000 new scientific titles, at least a half of which are new editions, that is neither translations nor reprints. We should, however, be cautious in accepting all the complaints about the overabundance of books, recalling that this is a natural source of anxiety to scientists unaware of the law of the exponential growth of science.

A periodical publication is a published work appearing at regular or irregular intervals in separate issues of differing contents but under the same title, and generally meant to be published for an indefinite time. Usually the issues have identical format. Conventional forms of periodicals are journals and newspapers. In a broader sense periodicals are sometimes considered also to include continued publications (serials), which are an intermediate form between books and journals.

Continued publications (serials) are collections of scientific papers and other documents put out by various institutions, societies and other organizations, without any strict periodicity but in numbered issues and under the same common title ("Transactions", "Papers", "Proceedings", etc.) and in identical format. It is sometimes hard to distinguish a continued publication from a book series or a multivolume set, on the one hand, and a journal on the other. The feature of continued publications (serials) which helps to distinguish them from book series is an indication that, in general, they are produced by establishments other than commercial publishing houses. Serials are mostly published by academies, universities, research and education institutions, learned societies and the like.

Serials are distinguished from multivolume books in that they have no fixed timetable and number of issues (volumes). Apart from publications issued without strict periodicity, publications issued at regular intervals but for a period longer than a year are also considered serials (yearbooks and publications issued every two, three, four, five or more years). This distinguishes them from journals which are either published during the year at regular intervals or have a fixed number of issues per year.

Serials are an important source of scientific information; they often carry information which does not appear in the regular journals and printed reports, papers and other articles on narrow subject fields related to the trends of research pursued by the scientific establishment concerned.

A journal or a magazine is a periodical publication regularly appearing in weekly, monthly, bimonthly, quarterly or semiannual issues, in uniform format and containing articles or other material of scientific, technical or socio-political content, as well as works of fiction. Journals contain current information and describe recent advances in science and technology.

Scientific journals came into being 300 years ago, the precise date being recognized as 5 January 1665, when the first issue of the French weekly "Journal des sçavants" appeared -- which was to give the name to this type of periodical publication. The main purpose of the "Journal", which predetermined the nature of scientific journals for a century and a half ahead, was to report on all the disciplines of science, literature and arts, with special emphasis on the natural sciences and engineering. Direct quotations and excerpts were a regular method for rendering the contents of books at the time.

At first, science chronicles played a secondary part in the journal. Gradually, communications on the newly discovered natural phenomena and experiments in the natural sciences began to occupy more and more space. Original articles were rarely published in journals in the 18th century. For the most part these were by prominent scientists and took the form of letters written to their colleagues - the traditional form used to communicate scientific discoveries in preceding times.

Articles in scientific journals are presently the main source of scientific information and they hold pride of place among all types of scientific documents. A survey of the bibliographic requests submitted by several thousand American scientists has shown that 68 per cent of the requests were for journal papers, whilst professional journals account for 70 per cent of all sources used by British electrical engineers.

Journal papers in natural, exact and technical sciences can be divided into three categories: first, original research papers which enable a scientist to reproduce the experiments and observations of the author of a paper and to verify his conclusions without any reference to other sources; second, preliminary publications and notes which possess certain features of scientific novelty, but are inferior in respect of exhaustivity and accuracy of description; third, the so-called explanatory articles that hardly carry any new scientific information, their purpose being to analyse and discuss information published earlier.

On the whole, journal articles are the best elaborated existing system for communicating scientific information. This system, as we have seen, has taken shape relatively recently when it took over some of the functions previously fulfilled by scientific books. It is still developing and requires profound study. Recent observations have shown that gradually more and more articles in science journals are due to team effort. While 80 per cent of the articles at the turn of the century were by a single author, such articles now constitute less than a half. The number of articles by two, three, four and more authors has been growing steadily.

As compared with books, journal articles are published with much less time lag. But they contain less generalizations and therefore soon become outdated. The obsolescence rate of both journal articles and books depends on the development rate of the field of knowledge concerned.

The growth rate of the number of journal titles has been given in the preceding chapter - it is described by a tenfold increase every fifty years. While in the middle of this century there were 100,000 periodical titles, the number will reach 1 million by its end. It is rather difficult to isolate from this total the number of scientific journals; for this we shall make use of a bibliographic reference guide entitled "Word List of Scientific Periodicals". Its first edition (1924) recorded 24,000 titles; the third edition (1952) contained already 50,000, and the fourth edition (1965) over 60,000 titles of journals. (1) This latest edition records all scientific journals that came out between 1900 and 1960, including those no longer published. It has been roughly estimated that, since about half of the journals listed have ceased publication, close to 30,000 titles of scientific journals are regularly being issued at present, with an annual total of 3 to 4 million articles, the latter being distributed very unevenly.

The growing number of papers published in journals, the loosely defined profiles of these, the rapid obsolescence of the material they carry, and the inadequate announcement of their publication to scientists - all this has contributed to the fact that since the 1930s the journal as a source of information has been a target for criticism. In the past three decades, various projects have been advanced concerning the eventual replacement of journals by other media for dissemination of knowledge; these are based on the idea of using for the purpose separate papers not published in periodicals but deposited in special branch-oriented centres and announced by abstracting publications. One of the earliest projects of this kind was drafted by Soviet delegates to the 1933 International Geological Congress, and a similar proposal was made the same year by an American - Watson Davis. It was this type of proposal that J.D. Bernal used as the basis of his widely known plan put forward in 1939 in the book entitled "The Social Function of Science". The Bernal Plan was discussed at the Royal Society Scientific Information Conference in 1948. Elaborating on his plan, Bernal suggested at the 1958 Conference on Scientific Information in Washington that the very form of science journal article should be renounced as the means for communicating R & D results. He argued that the form of a scientific paper does not provide for a rapid and adequate description of the entire course of research. The Bernal Plan is unlikely ever to be implemen-

---

1. World List of Scientific Periodicals, published in the Years 1900-1960, 4th ed. Ed. by P. Brown and G.B. Stratton. Vol. 1-3. London, 1964-1965.

ted since it did not take into account some recently discovered patterns in the science publications system. Apart from dissemination of knowledge, journals have other functions to discharge, including upholding a critical atmosphere in science, priority protection, and assessment of scientists' competence. Still, the plan did contain useful ideas that are already being implemented in the depositing of unpublished scientific and technical documents.

Scientific books and journals remain the most important and time-honoured types of scientific publications, which is the reason why we have dwelt on them at such length.

Newspapers are defined as periodical publications that are generally issued daily - or once, twice or three times a week - and carry information on current developments, primarily of socio-political nature. Newspapers also regularly publish popular science articles and sci-tech news features. The freshness and novelty of such newspaper items invest them with definite informative value. Newspaper clipping services provide for current awareness on items of topical interest as treated by the majority of a country's newspapers. However, the advertising tone so typical of many papers and the often unreliable information they give detract from their usefulness as sources of scientific information.

Special types of technical publications include standards, branch and interbranch technical documents, patent literature, technical catalogues and price lists for materials and equipment, as well as publications put out by information agencies. The term "special types of technical publications" is sometimes also applied to unpublished scientific documents such as scientific and technical reports, information cards, and translations of papers. The term was defined by Soviet librarian A.N. Barabanov who in the mid-1930s identified the common features uniting these diverse publications in respect of contents, presentation, origin and purpose.

Standards are scientifically substantiated technical documents which specify the types and models of products as well as their quality, testing methods, packing, marking, transportation and storage. Standards also define general technical values, dimensions, terms, and symbols. Usually they give a comprehensive description of the product concerned, but sometimes they deal with particular aspects, for instance, the testing procedures.

Standards serve to ensure high quality of industrial and agricultural products and the unification of parts and assemblies, as well as separate items. They continually promote better use of materials, fuel and power, faster designing and lower costs of industrial products, and can serve as a source of information on technical accomplishments.

Interbranch and branch technical documents (specifications, norms, and other instructional technical documents) are closely related to standards. These have a more restricted scope of application and are established either in the absence of a corresponding standard or in case of some special requirements for a specific product associated with special manufacturing or operation conditions. Specifications are used on a wide scale side by side with standards. These are commercial documents which state the consumers' or buyers' technical requirements for products, materials, processes or services which the manufacturer is liable to ensure.

Patent literature is an important source of scientific and technical information, making it possible to trace the history of any invention or discovery and to get a clear picture of the present state-of-the-art and development trends in a specific field of science or technology. Patents provide an opportunity for comparing and assessing the technological levels in different countries and of ascertaining the feasibility of mechanisation and automation of various production processes. Timely acquaintance with patent literature prevents duplication of inventions or the development of techniques and machinery that already exist.

Author's certificates are the main form of protection of inventor's rights in the socialist countries. It guarantees the permanent preservation of an inventor's right, as well as his right to remuneration and all the privileges granted to inventors and innovators in those countries. In accordance with the structure of the socialist system under which all means of production are owned by the State, an author's certificate reserves for the State the exclusive rights of using the invention. In the capitalist countries, the inventors are granted patents whose holders are entitled to the utilization of the invention for 15 to 17 years.

The bulk of patent literature is made up of the specifications appended to patents and author's certificates. These consist of a brief description of the essence of an invention, indicating its scope of application, describing the existing solutions of the problem concerned, and enumerating the drawings of the individual parts of the invention. Thereupon follows a complete description of the invention specifying the peculiar features and the essence of the new device, machine, product or chemical substance, and enumerating its main parts and assemblies with an explanation of their functions and interactions and references to the corresponding markings in the drawings. A patent specification ends up with the so-called "Claims", that is a brief summary of the object of invention underscoring the essential features which are regarded as the invention proper and are subject to legal protection.

Patent literature is very ample. The first patent was granted in Britain in 1649 and since the middle of the 17th century patenting of inventions has been a regular practice. The overall number of patent specifications published since is estimated at 7.5 million items, to which some 300,000 new titles are added every year. Periodical publications of the patent offices are likewise treated as patent literature. These are bulletins of inventions, aids for classifying inventions, and other publications giving information on patents and author's certificates.

Technical catalogues (also called industrial, commercial and company registers) are lists of products manufactured by industrial enterprises and concerns or sold by commercial companies. These usually contain a specification of each product with drawings and pictures and sometimes also indicate the price, scope of application, and type or model. Technical catalogues are of three main types: first, those describing products of a particular kind or designation and manufactured by the different enterprises and companies; second, by contrast, those enumerating the different items produced by an enterprise or a company; and third, reference catalogues, which give comparative data on various kinds of similar products. The number of technical catalogues is so great that it is hard to keep account of them: their annual output is over half a million new titles.

They are widely used by engineers, economists and scientists to study up-to-date technical equipment with a view to its acquisition, economical utilization or improvement.

Information leaflets or newsletters are one example of the numerous publications issued by information agencies. They are circulated in series of several hundred copies, serving for information exchange between the research institutes and design bureaux of enterprises to prevent duplication and parallelism in their work. The leaflets, of 3 to 4 pages, briefly describe engineering calculation techniques, novel procedures used in production, regulation, adjustment, testing and control of instruments and apparatus. Information leaflets deal with particular technical solutions of a general significance and are published before the completion of an overall research and development project. They often contain electrical or mechanical diagrams of particular elements, assemblies and blocks, describe their design and give the characteristics and properties of new materials.

Unpublished scientific documents, as we have seen, constitute an arbitrary category which includes very different documents: scientific and technical reports, dissertations, information cards, and translations. They all may be issued in a great number of copies and circulated as fully fledged publications. Another group of unpublished documents - manuscripts, proof sheets and preprints - are products of intermediate stages of the printing process.

Scientific and technical reports (reports on the results of research and development projects) are a major source of the information necessary for the development of science and technology, application of their results to industrial practice, and improvement of engineering procedures. The techniques of writing reports are different in various branches of science and technology, in different countries and even organizations. Nonetheless, certain common features are distinguishable in all these documents.

A scientific or technical report usually opens with an abstract or summary, briefly stating the task of the research and the results obtained. This is followed by an introduction giving a state-of-the-art survey of the progress made in a given field both at home and abroad. The main text of the report contains the task of the project, the formulation of the technical assignment, an analysis of the existing methods of solution, a substantiation of the approach adopted, the necessary calculations and experimental results. The report closes up with conclusions that compare and analyse the theoretical and experimental data obtained during the research. The last section of the report contains an assessment of the results and their comparison with the technical assignment, as well as an outline of their prospective utilization. Usually a scientific or technical report is furnished with a review of the sources used, sometimes given in the first section of the report, in which case the latter is concluded by a subject index to the references.

This kind of document has become particularly widespread in the post-war years and is now competing with journals. The number of reports is also very great. In the USSR, over 13,000 research and development organizations carry out and describe in their reports up to 150,000 projects annually. In the USA, federal agencies alone issues over 100,000 reports every year. Since scientific and technical reports are not considered as publications and their circulation is restricted, many countries take steps to arrange for an official registration of these documents and for centralized information about them.

Dissertations are research papers submitted by those applying for a degree in science. Depending on their discipline and subject, dissertations may belong to different types of research work. In experimental research, the greatest attention is paid to the description of experiments carried out and the conclusions drawn. Methodological studies evaluate the existing methods, identify their hitherto unnoticed features, give new arguments favouring or opposing this or that method, or develop absolutely novel research or production methodologies. Descriptive studies give documental description of unknown scientific facts and phenomena, reveal their essence and interrelations with other facts and phenomena, and advance theoretical hypotheses proceeding from theoretical generalizations. Computational and analytic studies lay the emphasis on a mathematical interpretation of processes and phenomena. Finally, historical and bibliographical studies investigate the role and importance of the individual scientists, trends and schools in the development of science and technology. In practice most dissertations are studies of a mixed pattern, possessing most - if not all - of the above features.

Dissertations are normally not published and are available in only two or three copies. What is published, in 100-150 copies, are 15-20 page abstracts written by the author for the preliminary communication of his principal conclusions to the scientific community.

Information cards can serve as the main form of supply of scientific, technical, industrial and economic information by local enterprises and organizations to central agencies, that is as the main source of primary information. Information cards communicate the achievements of research and development centres, institutions of higher learning, enterprises, construction projects, etc. A card is made in one copy directly at a given establishment in conformity with a standard pattern and is signed by the head of the establishment.

Preprints are offprints of an article or other paper, printed before its official publication in the current issue of a periodical and intended to be sent out to a limited number of interested persons. Many periodicals and serials make offprints of particular papers along with the main run and deliver them to the authors who send them out to their colleagues. This is a time-honoured practice and it has paved the way for preprints. These have the advantage of being ahead of the official publication of science documents, sometimes by several months, and of providing faster communication of scientific facts and ideas to specialists.

Deposited manuscripts are a special category of unpublished primary documents. Manuscripts of monographs, collections of papers and separate papers of interest to a limited number of specialists can be transferred, by decision of the publishers, journal editors or research institutes, to a central information agency for deposition. This means that the manuscripts will be announced in special bulletins or catalogues and that their copies will be available on request from the information agency.

Unpublished documents are gaining importance as a means of communicating the latest results of scientific research.

The patterns of the distribution of scientific publications

Until quite recently, a descriptive approach was the only one applied in the study of the system of scientific publications. Bibliologists and bibliographers strove for a most detailed typology of scientific documents and publications and wanted to ascertain the part played by each type in scientific work. The development of informatics has ushered in a new stage in the study of scientific publications. Science revealed some of the general laws that characterize the inner connections existing between scientific publications and the growth of science, and derived quantitative relations between the number of publications and the indicators of the growth of science.

It has been ascertained that there exists a common and regular pattern of distribution for the number of authors who publish a given number of papers during their lifetime, the number of journals annually publishing a given number of papers, and the number of publications containing a given number of references to other publications. "They follow the same type of distribution as that of millionaires and peasants in a highly capitalistic economy. A large share of wealth is in the hands of a very small number of extremely wealthy individuals, and a small residual share in the hands of the large number of minimal producers. Whether the exact form of the distribution is lognormal, exponential, a Zipf Law, or an inverse square has been a matter of much conjecture in each of the cases".(1)

This means that the majority of authors produce only one or two papers in their lifetime, while a small group of writers are very prolific and publish dozens or even hundreds of papers. Periodicals that cease publication are for the most part those which have published only one or two issues, while a small number of periodicals issued since long ago account for the lion's share of all papers published. About half of all published literature stems from that number of authors or journals which is equal to the square root of their total number. "In short if there are for example 30,000 journals alive in the world, or 1,000,000 publishing scientists in a country then a selected list of about 175 journals or 1,000 scientists would account for half of the bulk of the literature but probably for 70-80% of the important content." (2) This is a discovery of a great importance:it helps to overcome the awe inspired by the tremendous amount of literature and provides theoretical ground for finding those publications of a genuine value among the vast multitude of others.

In this connection, the questions arise: what is the proportion of the articles dealing with a given problem to the total number of scientific periodicals; and what is the proportion of the publications of genuine value to their totality.

---

1. Price, D. de Solla. Regular patterns in the organization of science. "Organon", 1965, No. 2, p. 246
2. Price, D. op.cit., p. 248

Bradford's Law of Scattering

The former question was answered two decades ago by the British information scientist S.C. Bradford. He established that, taking as unity the totality of articles dealing with a subject, the specialized periodicals in the field concerned will contain but one-third; another third of the relevant articles will be found in a much larger number of journals of a different profile but having a relation to the given subject; and, finally, the remaining third of the articles will be scattered among a vast number of periodicals in which it would have been impossible to predict their appearance as these periodicals are either of a too general nature or absolutely alien to the subject in question. This appearance of articles belonging to any given field of knowledge in journals bearing no relation to that field is explained by the unity of science, by the fact that along with the differentiation of knowledge contrary integrative processes are unfolding.

Bradford has thus formulated the Law of Scattering of publications that he discovered: "If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as $1 : n : n^2 ...$" (1)

Another British information scientist, B. Vickery (2), refined the Law of Scattering, by deriving the following relation:

$$T_x : T_{2x} : T_{3x} = 1 : n : n^2,$$

where T_x is the number of journals containing x articles on a given subject; T_{2x} is the number of journals containing $2x$ such articles; T_{3x} is the total number of journals containing $3x$ articles; and n is any number depending on the value to which x has been set equal.

The scattering of publications over scientific periodicals, as has become clear, is a particular instance of the distribution characterized by the above-mentioned Zipf Law. The Zipf Law is an expression of a universal property inherent in very different languages. The property consists in the fact that the bulk of any text in any language is made up of a very small number of frequently used words while the rest of the words are used only rarely.

The law of scattering of publications is of great practical importance. It implies that the coverage of all the publications on a subject cannot be ensured by scanning only the periodicals of a given or related profile; for this it is necessary to look through practically all the scientific and technical periodicals. In addition, the law can be used to determine the number

---

1. Bradford, S.C., Documentation, 2nd ed. London, 1953, p. 154
2. Vickery, B.C., Bradford's law of scattering. "Journal of Documentation", 1948, vol. 4, No. 3, p. 200

of periodicals sufficient to provide for the coverage of a specified percentage of publications in a subject field or discipline. Later we shall describe how this law is taken into account in organizing information activities.

Bibliographic coupling of documents by citations

To answer the second question, it is necessary to know one of the new methods of the study of scientific publications - known as the method of bibliographic coupling. It has struck the information scientists that scholars regularly make references to the works of their predecessors who have explored the same, similar or related subjects. This practice, which first emerged a century ago, has been for the past fifty years an indispensable standard and criterion of the thoroughness of a scientific publication.

The study of bibliographic references contained in most publications has shown that they form a dense network of associations between scientific publications. Grouping of documents which cite the same publications reveals internal relations between the remotest subjects long before these relations can be otherwise noticed and explored. This is achieved by a purely automatic procedure since the writers of publications on most diverse subjects may cite the same document as their common point of departure.

Citations of the materials used have long been serving scientists and practitioners as important clues in searching for the relevant information sources. This path, however, is long and laborious, and, what is worse, it does not ensure the detection of all the pertinent papers. For a better insight into the system of citations, let us consider some single subject of research.

The American biologist G. Allen compiled a map of the citations of earlier documents in fifteen articles on nucleic acids published between 1941 and 1960. This map, represented in Fig. 4, shows that for the complete coverage of all the 15 articles on the basis of citations they contain it is necessary to refer to six articles at least: 6, 7, 9, 13, 14, 5, published between 1956 and 1958. Some articles that are rarely or never cited - namely, 7, 8 and 9 - are hard to find by this technique. It is quite natural, too, that the latest publications (14 and 15), issued simultaneously, are not connected by reciprocal citations.

If it were possible to trace the bibliographic relations between publications on the basis of citation data not from bottom to top, i.e. from the recent works to the earlier ones, as has been done usually, but vice versa, from top to bottom, that is from earlier to the recent documents, this would make the finding of all the articles pertaining to a given subject much easier. References to article 2 alone would identify at once a half of the remaining articles - 3, 5, 6, 7, 11, 12, 15, while references to three articles 1, 2, 4, would provide information on almost all the publications, missing only two of them - 9 and 13. Recently this has become possible. The bibliographic coupling method has found its practical implementation in citation indexes compiled with the aid of electronic digital computers. These will be dealt with later on. Here we shall only dwell on the application of this method for a theoretical study of the system of science publications.

It might be asked, to what extent bibliographic coupling of documents provides for an adequate and exhaustive reflection of the links actually existing in science. To answer this question, E. Garfield, the author of the first modern citation index and head of the Institute for Scientific Information (Philadelphia), carried out the following experiment in 1964.(1) The history of one of the most important discoveries in modern biology - the mechanism of transfer of hereditary characteristics through desoxyribonucleic acids - was subjected to parallel studies by means of conventional techniques and by the bibliographic coupling method. The conventional method was based on a historical monograph by the well-known American scholar and writer Isaac Asimov entitled "The Genetic Code", while the bibliographic coupling used a special computer-produced citation index to genetics literature.

The comparison showed that 65 per cent of the publications identified by both methods as turning points on the way towards the discovery were identical. The conventional method proved more efficient only in identifying the publications of the last century when the practice of making references to the papers used was insufficiently developed. As to the past few decades, bibliographic coupling spotted a great many associations between publications missed in Asimov's book. The experiment yielded a number of important conclusions. The most remarkable one was that bibliographic coupling of documents on the basis of citation data does provide for a correct and full recognition of the actual relationships between research papers. The assumption was corroborated too, that the greater the attention paid by the scholars to a given publication - and, consequently, the greater the number of references they make to it - the greater its scientific value.

We have already mentioned that the distribution of citations over the citing articles follows the general pattern of distribution of mass-scale phenomena; it is analogous to the numbers of authors and journals publishing a given number of articles during their lifetime. To illustrate this by figures: 10 per cent of scientific publications make no references at all, 85 per cent of publications containing from 1 to 25 references account for a half of the total citations, while the remaining 5 per cent of publications with a very great number of references - surveys and reviews - account for the other half. Despite the fact that some publications lack references altogether, while some may contain a sizable number of references to earlier works, certain average figures can be derived. These are: 15 citations per publication, of which 12 or 13 are references to journal articles.

At this point it is possible to try to derive the ratio of publications of genuine value to the total number. For this we shall observe how the new papers published in a specific field during one year are linked by citations to all the publications existing in that field (Fig. 5). To make the picture clearer, suppose that by the beginning of the year there were 100 publications dealing with the field or subject under consideration. Assuming the regular exponential growth of literature (a 7-per cent annual increment), seven new papers will be published during the year, each containing - according to the above average data - 13 references to earlier journal articles. The total for the whole year will be 91 references (13 x 7) to earlier publications.

---

1. Garfield, E. et al. The use of citation data in writing the history of science. Philadelphia, 1964, V, 75 p.

As shown by a special analysis, the new citations will have the following pattern of distribution with respect to the earlier publications: 40 of the earlier papers will not be mentioned at all, 10 references will be to publications in other fields of knowledge, 50 will be to 50 papers in the field, that is each of these will be mentioned once during the year, and 31 references will be to 10 papers cited more than once. These latter 10 papers (their associations with the 7 new ones are shown on Fig. 5) are the most important for science, since the more often a paper is cited in later publications the greater its scientific value.

We see that the new publications tie together, as it were by their references only, a small part of the previously published literature, which may be regarded as the active front of research, while the rest of the literature is on the whole of minor significance.

It is interesting to analyse the dynamic characteristics of this active research front and assess its relative distance from the publications of the preceding year. For this purpose, 200 consecutively issued papers constituting the literature of a narrow topic in physics were studied. In the autumn of 1903, French Academician R. Blondleau announced his discovery of a new radiation which he called N-rays. This communication gave rise to wide discussions and many publications until at last the discovery was recognized to be a mistake and the N-rays to be a mere fancy of the famous physicist. The episode, well known to the historians of science, is convenient for building a model of the distribution of citations over publications as it is definitely localized and fixed in time.

Figure 6 shows a matrix where every dot marks a citation of a paper in another one. The horizontal axis shows the numbers designating the citing papers, the vertical axis the cited papers, designated by the same numbers. The papers are numbered chronologically, as published. The matrix is a graphic illustration that a great number of references fall in the band of the 50 papers immediately preceding each publication concerned. Thus the continuous and the dashed lines are the boundaries between which lies the active research front identified on the preceding diagram.

Remarkable are the vertical bands of dots corresponding to the publications that contain the maximum number of references and the horizontal bands that correspond to the most frequently cited papers. The former bands indicate that after every 30 or 40 publications the need arises for a survey that would serve as substitute for the older publications which drop out of sight behind the active front. The latter bands point to the classic works in a given field which are perpetually referred to with a roughly constant frequency and constitute only a minor percentage of the total publications.

This model can be considered in general as showing the pattern of the distribution of citations in publications dealing with any subject, problem or discipline, and hence, as a characterization of the distribution of the most frequently used publications over their totality.

The laws of obsolescence of scientific publications

Studies of bibliographic references made it possible long ago to establish certain laws of obsolescence of science publications. Several years back the American bibliographer Ch. Brown analysed some 40,000 citations in the leading journals published in English, French and German in mathematical, physical, chemical, geological and biological disciplines during the period 1952-1954. He took several volumes from each of the 56 journals and serials covered in his study to a total of 40,000 pages. Among the many conclusions he reached what now interests us is his data on the obsolescence of scientific books and periodicals, which show that the faster the growth rate of a scientific discipline the quicker its literature becomes obsolete. In the journals covered in Brown's study, references to books dated at most ten years back accounted for 64% of all references to books in physics and chemistry, 62% in physiology, 56% in geology, 38% in botany, and 27% in mathematics. The picture was much the same with references to journal articles. The percentage of references to papers less than ten years old, however, was conspicuously higher than that for books, and the obsolescence rate of the papers of the preceding decade was more rapid.

To describe the process of obsolescence of literature in a field of science, a concept borrowed from nuclear physics has recently been applied. By analogy with the life period of disintegration of radioactive substances an attempt has been made to measure the periods of half-life of scientific literature. Half-life is defined as the time during which a half of all currently used literature in a field has been published. It is exactly equal to the time during which a half of all the literature published in the field by the present moment will pass out of use. This fact is of major importance as it enables us to make reliable predictions of the future relative value of publications on the basis of an analysis of the current obsolescence rates of the literature in the field concerned.

Proceeding from the findings of Ch. Brown, which they supplemented with their own data, librarian R. Burton and physicist R. Kebler of the United States computed the following half-life periods of papers in the different branches of science:

Physics 4.6 years	Botany 10.0 years
Physiology 7.2 years	Mathematics 10.5 years
Chemistry 8.1 years	Geology 11.8 years

This shows that a great part of physical papers becomes obsolete very soon - being, so to say, ephemeral literature - that in physiology and chemistry there are approximately as many short-lived as there are lasting ('classic') works, while in mathematics the latter type is predominant. The age of publications that constitute the active research front depends on the field of knowledge. It is shorter in the fields characterized by shorter half-life, such as physics, physiology and chemistry, and longer in the fields with longer half-life, such as botany, mathematics and biology.

We are now able to derive the general averages. Fig. 8 shows the age distribution of papers cited in the publications of 1965. The data are based on an analysis of 3.3 million scientific papers issued within several hundred years and cited in 296,000 publications issued in 1965 in the exact, natural and applied sciences.

The greatest proportion of the cited papers (13% of the total) are dated 1964. Since newly issued publications have to wait before the scientists become aware of them, the percentage of references to the publications of 1965 is naturally much lower (5% of the total citations made that year). Then follows a decline, the middle of which (6.5% of publications) falls to 1960. Publications dated from 1960 to 1965 make 52.6% of the total citations made in 1965. This means that the average lifetime of publications constituting the active research front does not exceed 5 years.

To sum up, we have seen that among the many types of scientific literature the major means of dissemination of knowledge, which is of paramount though varying importance are journal articles, books, and scientific and technical reports. Recent research has shown that far from all the enormous number of scientific documents accumulated by mankind possess any real significance for further research. The analysis of the interrelations between scientific publications on the basis of bibliographic citation data has helped to establish that the core of the scientific papers of genuine value is made up of only a minor proportion of the publications of recent years, namely those most frequently cited in later books and papers. It is these papers which constitute the active research front, that should be sought first of all.

All the subsequent chapters will deal with the different aspects of the means and methods available to search for these publications and to use them most effectively.

Questions for self-checking

1. What are the development stages through which written records have passed from antiquity till our day?
2. What is a scientific document and what are the existing types of scientific documents?
3. What is a book and what kinds of books are important for scientific work?
4. What are the existing types of periodical publications and what are their characteristic features?
5. In what does the importance of journals and journal articles lie?
6. What functions are fulfilled by the special types of technical publications?
7. What are the contents of scientific and technical reports?
8. How are scientific publications interconnected by bibliographic references?
9. What regular patterns of distribution of scientific publications do you know?
10. What laws of obsolescence of scientific publications do you know?

Literature

1. Burton, R.E. and Kebler, R.E. The "half-life" of some scientific and technical literature. "American Documentation", 1960, vol. 11, 1, p. 18-22
2. Houghton, B. Technical Information Sources. A Guide to Patents, Standards and Technical Reports Literature. Melbourne (a.o.), Cheshire, 1967, 101 p.
3. Johnson, E.D. Communication. A concise introduction to the history of the alphabet, writing, printing, books and libraries. New Brunswick (N.J.), Scarecrow, 1955, 211 p.
4. Kronick, D.A. A history of scientific and technical periodicals: the origins and development of the scientific and technological press, 1665-1790. New York, Scarecrow, 1962, 274 p.
5. Merton, R.K. The Matthew effect in science. "Science", 1968, vol. 159, 3810, p. 56-63
6. Price, D.J. de Solla. Networks of scientific papers. "Science", 1965, vol. 149, 3683, p. 510-515

3. INFORMATION AND BIBLIOGRAPHIC PUBLICATIONS — SOURCES OF DATA ON LITERATURE

As stated in the previous chapter, scientists have always considered scientific literature to be too abundant and diversified to serve as a medium for dissemination of scientific knowledge. Therefore, along with those kinds of scientific literature which we have studied, special kinds of scientific documents were evolving whose primary mission was to serve as guides to the literature of science and to scientific documents.

Stages in the development of secondary documents

Lists of scientific books existed even in early antiquity. Among the clay tablets dating back to the 2nd millenium B.C. were found some on which there was no text and only titles were inscribed: they were prototypes of today's book catalogues. In ancient Greece and Rome, very large and highly sophisticated book lists were compiled. The "Tablets" of the scholar of antiquity Callimachus, which were compiled around 250 B.C., comprised 120 volumes and contained detailed and accurate data on authors, titles and sizes of books, and the time they were written. These lists or catalogues gave rise to a special kind of literature - bibliography. About the close of the 4th Century, St. Jerome wrote "De Viris Illustribus", which is believed to be the earliest of all known bibliographies.

Scientific periodicals made their appearance a century and a half after the first printed books; as you already know, they have for a long time served to announce to scientists new scientific publications. Another hundred and fifty years later, in mid-18th century, Diderot's and D'Alembert's "Encyclopédie, ou Dictionnaire universel des arts et des sciences" started publication, giving a powerful impetus to the development of reference literature. The first quarter of the 19th century saw the beginning of national bibliographic registration of printed material, regular publication of national bibliographies, and the appearance of the first abstract journals interpreting the contents of journal articles.

Each of these types of documents performs a specific useful function within the contemporary system of scientific literature, and their appearance was in response to a certain social need, often not realized at the time, which was surely generated by the growth in the quantity and variety of scientific literature. Therefore, although we often express our dissatisfaction with the historically evolved system of communication of scientific knowledge, we cannot deny that its progress is subject to certain rules.

According to the previously established typology of scientific documents, those which contain mainly information on scientific literature or results of its analytico-synthetic processing are regarded as secondary documents. Since the majority of these are published in the course of information or bibliographical work, we call them information and bibliographic publications.

Types of secondary documents

If the degree of information processing be taken as a criterion, different kinds of secondary publications can be conveniently placed in the following order: reference literature, reviews, abstracts and annotations, catalogues, card indexes, and bibliographies. Here the first kinds of documents are those in which information has been subjected to maximum processing, while the last ones are those where it was minimal. We shall follow this order in describing secondary documents and publications, which are the subject of this chapter. We shall speak only of the basic types and shall give relevant examples.

Reference literature is intended for rapidly obtaining data of scientific, applied or instructional character, including results of theoretical generalizations, basic scientific facts, mathematical, physical and chemical constants, industrial information accompanied by various tables, diagrams, drawings and formulas. The most significant of reference publications are encyclopaedias, industrial handbooks, glossaries, vocabularies of terms, biographical dictionaries, bilingual and multilingual dictionaries.

Encyclopaedias are reference publications which contain the most essential information (comprehensive or brief) on every field, or some particular field, of knowledge and practical activity. They are divided, according to content - into general (universal) and those devoted to a particular topic; according to structure - into alphabetical and systematic; according to size - into desk-type (1 to 4 volumes), smaller (5-12 volumes) and greater (several dozen volumes). Encyclopaedias are usually written on the basis of a topical plan (by which the bulk of the work is broken down into different scientific disciplines or their parts) and of a word-list giving all the words for which separate entries are prepared.

Entries in encyclopaedias vary in type and length. The largest are review entries covering broad topics. The majority of entries are references which provide a definition and some basic information on a subject. Short interpretative entries include only a definition and sometimes the origin of the word. Finally, encyclopaedias contain a great many cross-reference entries which indicate that the relevant information should be looked up under a different word. Of great help are the bibliographical lists appended to entries: they enable the reader to study at greater length a subject only superficially treated in the encyclopaedia. Many encyclopaedias are provided with subject indexes which increase their value as reference tools.

General encyclopaedias are represented in the USSR by all the three types: greater, smaller and desk-type. The "Great Soviet Encyclopaedia" (GSE) has so far appeared in two editions: the first, completed in 1947, comprises 65 main and one supplementary topical volume "U.S.S.R."; the second appeared from 1949 to 1958 and comprises 49 main and 2 supplementary volumes. New data are published regularly in GSE Yearbooks. There is a two-volume alphabetical, subject index. The "Small Soviet Encyclopaedia" has appeared in three editions: the first, in 10 volumes, was brought out from 1928 to 1931; the second, in 11 volumes, from 1933 to 1947; and the third, in 10 volumes, from 1958 to 1961. For the latest edition, containing some 48,000 entries, a 165,000-entry subject and name index was printed in a separate volume. Desk-type encyclopaedias are represented in the USSR by the Encyclopaedic Dictionary which is intended for the general public. It was brought out in three volumes in 1953-1955, and in two volumes, in 1963-1964.

Of the general encyclopaedias, the oldest and probably best known is the "Encyclopaedia Britannica", which has been appearing for nearly two centuries. Its first three-volume edition came off the press in Edinburgh in 1768-1771. It is traditionally considered to be one of the most authoritative encyclopaedias. This is particularly true of the second, 29-volume, edition published in 1910-1911. It was that edition which the eminent Russian scientist K.A. Timiryazev called the first genuine encyclopaedia of our century. A feature of the"Encyclopaedia Britannica" is its small word-list and most minute review entries; it contains some 43,000 entries, with the subject index running up to 500,000 entries. The latest completely revised edition of the "Encyclopaedia Britannica" was the 14th edition in 24 volumes, published in 1929. (1) It has been reprinted annually since then, with one-eighth of the entries undergoing changes. The contents are thus revised rather rarely and contain much outdated information. Besides, this type of encyclopaedia is no longer wholly satisfactory to the scholar; with the ever deepening differentiation of scientific knowledge the system of review-type entries reduces the value of the publication as a reference tool.

In sharp contrast to the "Encyclopaedia Britannica" stands the multivolume German encyclopaedia "Der Grosse Brockhaus" which was first published in 1796-1811. It contains a great many reference-entries and is intended for the broad reading public. Its latest, 16th edition comprises 12 main and 2 supplementary volumes totalling more than 150,000 entries. (2) "Der Grosse Brockhaus" is distinguished by a uniform coverage of the whole universe of knowledge, with a slight bias towards natural sciences. It can be said to be the most up-to-date, detailed and carefully prepared of all existing encyclopaedias.

Among French encyclopaedias, the greatest popularity is enjoyed by the numerous Larousse publications. Natural and exact sciences and technology feature rather prominently in the 10-volume "Grand Larousse encyclopédique", which provides the most up-to-date and detailed information on nuclear physics, astronautics, electronics and similar disciplines. (3) Particularly widely used is the "Petit Larousse", first published in 1906 and reissued annually ever since. (4)

---

1. Encyclopaedia Britannica. 14th ed. Vol. 1-24. Chicago-London, 1929.
2. Der Grosse Brockhaus. 16.Aufl. Bd. 1-12, Ergänzungsband, Atlasband. Wiesbaden, 1952-1958.
3. Grand Larousse encyclopédique. Vol. 1-10. Paris, 1959-1964.
4. Petit Larousse. Paris, 1967.

Encyclopaedias for special topics differ from general encyclopaedias not only in coverage but also in a more limited purpose, being as a rule designed for specialists and researchers. To cite a few examples, the following publications were published or reissued in recent years:

-"McGraw-Hill Encyclopaedia of Science and Technology". Vol. 1-15, Yearbooks. New York, 1960-1969.
-"International Dictionary of Applied Mathematics". New-York, Van Nostrand, 1960.
- Kirk,R.E. and Othmer,D.F. "Encyclopaedia of Chemical Technology". New ed. Vol. 1-18, Supplements. New York, Interscience, 1963-1969.
- Gray,P. (ed.). "The Encyclopaedia of the Biological Sciences". New York, Reinhold, 1960.

Encyclopaedias can serve as sources of scientific information. However, a critical approach is needed in using an encyclopaedia. The latest editions are to be preferred, and the date of publication printed on the title-page should be checked in order to verify whether it stands for a new edition or merely a reprint. The rule of thumb to follow with respect to encyclopaedias is summarized in the words of the distinguished historian of science G. Sarton who said that to turn to an encyclopaedia for a primary reference is clever, to ignore encyclopaedias is unreasonable, to rely on them too much is stupid.

Handbooks provide data which can be useful to different subject specialists. They vary broadly in subject, structure, and size. For instance, there are very compact handbooks, such as:

- American Institute of Physics. "Handbook". New York, McGraw-Hill, 1957.
- Lange,N.A. (ed). "Handbook of Chemistry". 10th ed. New York, McGraw-Hill, 1961.
- "Machinery's Handbook: A reference book for the mechanical engineer, draftsman, toolmaker, and machinist". 17th ed. New York, Industrial Press, 1964.
- Flugge,W. (ed.). "Handbook of Engineering Mechanics". New York, McGraw-Hill, 1962.

Such a classic work as "Beilsteins Handbuch der organischen Chemie" (1) comprises about 100 volumes, which have been brought out in three cycles since 1918. The same great size and elaborate structure are typical of the well-known inorganic chemistry handbook popularly known as "Gmelin" (2) The hundreds of its issues and their indexes call for a special study before a specialist can make full use of them.

Closely associated with handbooks are tables of constants, guides to minerals, plants and animals, drawing and map atlases, collections of formulas and diagrams, and similar publications, for example, "Handbook of Mathematical Tables. Supplement to Handbook of Chemistry and Physics", Cleveland, Chemical Rubber Co., 1962; Kaye,G.W. and Laby,T.H. "Tables of Physical and Chemical Constants and some Mathematical Functions", 12th ed., London, Longmans, 1959.

---

1. Beilstein's Handbuch der organischen Chemie. 4.Aufl. Bd. 1-27, Berlin, 1918-1940; Bd. 1-31, 1928-1938; Bd. 1-27, 1941-1955; Bd. 1-, 1963-.
2. Gmelin-Krauts Handbuch der anorganischen Chemie. 8. Aufl. Clausthal-Zellerfeld, 1924-1965.

Biographical (who's who) dictionaries supply data on life, activities and work of eminent scientists, engineers, physicians, agronomists, public figures, etc. Such dictionaries are particularly useful in studying the history of science, technology and industry. Examples are:

- "American Men of Science". Vol. 1-5. Arizona, Cattell, 1960.
- "Directory of British Scientists". London, Benn, 1963.
- "World Directory of Mathematicians". Bombay, Tata Institute of Fundamental Research, 1961.

Dictionaries may be a very useful starting point as an aid to establishing a clear definition and the precise meaning of a term. Any single-language dictionary is concerned essentially with information about spelling, pronunciation , derivation,meaning and usage, e.g. "The Concise Oxford Dictionary of Current English", ed. by H.W. Fowler and F.G. Fowler, 4th ed. rev. by E. McIntosh, London, Oxford Univ. Press, 1956.

Bilingual and multilingual dictionaries contain words and terms of one language and their equivalents in other languages. They can be either general or special, i.e. explaining terms in a special subject field, for example:

- "Hoyer-Kreuter's Technological Dictionary in three Languages: English-French-German". New York, Ungar, 1944.
- Callaham,L.I. "Russian-English Chemical and Polytechnical Dictionary". 2nd. ed., New York, Wiley, 1962.
- Emin,I."Russian-English Physics Dictionary". New York, Wiley, 1962.
- "Russian-English Atomic Dictionary." New York, Technical Dictionaries, 1957.

Guides to general reference books contain bibliographical details of these. The most important guides are:

- Walford,A.J. "Guide to Reference Materials." 2nd ed., London, Library Assoc., 1966-1969.
- Winchell,C.M. "Guide to Reference Books." 7th ed. Supplements 1-4. Chicago, American Library Association, 1951-1963.

We may also mention: Maichel,K. "Guide to Russian Reference Books," V.I. General bibliographies and reference books. Stanford, Hoover Institution, 1962, 92 p.

Reviews of progress in different fields of science and technology have been in existence for quite a long time. In the last decade they have come to occupy an important position among secondary publications. Reviews summarize data contained in primary documents, such as articles, monographs, conference reports, technical reports, invention specifications; they generally cover publications in one subject field over a definite time period, and may be analytical or informative.

Analytical reviews are distinguished by deeper analysis and are based on appropriate data selected from primary scientific publications. They usually contain all basic facts concerning progress in a subject field, and permit assessing the advances in the study of a problem or in an area of research.

Informative reviews are very valuable sources of information although they are less analytical. They initiate an incipient researcher into his field of interest, familiarize him with progress in related fields of science and technology, help him to choose or define more precisely the direction and methodology of his own study or design work; they are great time-savers, sparing a specialist long hours of searching and scanning primary materials.

Reviews are published in different forms, such as an article in an ordinary journal, or in a special reviewing journal (e.g. "Advances in physics"), a separate issue of a periodical, or of an irregularly appearing book series. They are typified by the reviews of different disciplines published by the British quarterly "Science progress", which prints review papers on advances in astronomy, meteorology, physics, chemistry, geology, mineralogy, zoology and botany. Among others, the following reviews have been published in recent years:
- "Annual Report of Progress in Physics" (Inst. of Physics and Physical Society).
- "Chemical Reviews" (American Chemical Society).
- "Advances in Space Science". Vol. 1-11. New York, Academic Press, 1959-1969.
- "Progress in Aeronautical Sciences". Vol. 1-9. London, Pergamon, 1961-1969.

Abstract journals, nowadays the most important secondary information publications, appear regularly and reflect the contents of the most recent scientific documents in the form of abstracts or, to a lesser extent, of annotations.

An *abstract* is a summary of a document, specifying its nature as well as the techniques, findings, time and place of an investigation; it also indicates the objectives and main theoretical premises of a study, supplies the pertinent figures, formulae, tables, charts and diagrams, and concludes with the number of illustrations, tables and references in the abstracted publication, as well as the abstractor's name. All abstracts fall into two general types: informative (abstract-synopsis) and indicative (abstract-summary). An informative abstract contains in an abridged form all the main ideas of a primary publication, while an indicative abstract reports only the most important of them. An abstract, as a rule, does not reflect the abstractor's subjective views, and does not make any judgments on a document.

It is worthwhile citing here the Guide to the preparation of abstracts formulated by one of the leading abstracting services, "Biological Abstracts": "An abstract should be a non-critical, informative digest of the significant content and conclusions of the paper, not a mere description. It should be intelligible in itself, without reference to the original, but it is not intended to substitute for it. It should be brief (preferably less than 3% of the original), written in whole sentences, not telegraphic phrases". (1)

Naturally, the size of an abstract depends on the size and importance of the abstracted document. However, the abstract length in different abstracting services has stabilized at the 800-900 character mark.

---

1. "Biological Abstracts", 1969, vol. 50, 1, p. XXI.

The average length of an abstract in some leading abstract journals in 1965 was as follows: "Physical Abstracts", 500-550 characters; "Chemical Abstracts", 900-1000 characters; VINITI's "Referativnij Zhurnal", 1100 - 1200 characters.

Annotations, in contrast to abstracts, are not summaries of primary documents but only short communications on their theme or content, although they also indicate their nature. According to purpose, annotations are divided into descriptive annotations, which provide a factual description of the contents and some other details of a document and recommendatory (critical) annotations, which appraise a document from the point of view of its usefulness to a particular user category.

Information and bibliographical services usually prefer the descriptive type. These may be very short, consisting of a few words or short phrases, or more detailed, but in either case they only describe and do not expound the main ideas and conclusions of a primary document. This feature makes annotations particularly suitable for revealing the contents of books. The reason is that the multidimensional contents of books does not lend itself easily to abstracting. A descriptive annotation usually indicates the kind of book, its purpose, and structure, the scientific school to which the author belongs, the topic of the study and its conclusions, as well as the auxiliary and illustrative material, supplements, addenda, and the reference material, including indexes and bibliography. On the average, an annotation does not exceed 400 to 600 characters. A characteristic feature of the annotation is that it must be closely associated with the description contained in the entry, which in abstract journals immediately precedes each abstract and annotation, but may also be included as a separate entry, that is as a secondary document in its own right. In addition, each document indexed in an abstract journal is assigned a registration (serial) number and a classification number (document symbol). For foreign materials, a translation of the title is also given and the language of the original document is indicated.

Abstract journals have two main objectives: to keep specialists informed of all significant current developments in their respective fields, without recourse to primary publications, and to help them find material on a specific subject published over a fairly long period of time. The first objective is taken care of by the publication of monthly issues of specialized series of an abstract journal (in which the material is grouped by subjects) and by their distribution on a subscription basis among specialists and research libraries. The second objective is ensured by the regular printing of subject, author, formula and other special indexes to each issue or yearly set of the journal.

It has to be admitted that abstract journals are not up to the mark with respect to either of these objectives: in the former case, because of considerable publication delays (an abstract appears in an abstract journal at best four months after publication of the original paper); in the latter case, because of the inadequacies of reference tools and lack of cumulations (in searching for literature over a fairly long period of time, a large number of annual indexes have to be checked, and a portion of indexed material will always escape the search). However, abstract journals are steadily improving their quality and also growing in number;

at present they number several hundred titles.

In using different abstract publications, the quality, type and size of abstracts, their arrangement, the time lag between the appearance of a publication and of its abstract, the availability and quality of indexes, are not the only features to be considered. An equally important factor is the coverage of scientific publications, their languages and country of origin, as well as their categories. In other words, it should be known whether papers alone or also books, patent specifications, technical reports are abstracted, and whether these are only in European or in Oriental languages as well, etc. By way of illustration, we shall present data on several well-known abstract journals.

"Referativnij Zhurnal", the Abstract Journal of the All-Union Institute of Scientific and Technical Information (VINITI) has been published since 1952; in 19 years it has grown to be a unique service, the largest of its kind in the world.

The "Referativnij Žhurnal" comprises the following series: Automation and Radioelectronics, Astronomy and Geodesy, Biology, Biological Chemistry, Geography, Geology, Mining, Geophysics, Mathematics, Mechanical Engineering, Metallurgy, Mechanics, Physics, Chemistry, Industrial Economy, Electrical Engineering. In 1963, the separate series "Scientific and Technical Information" was started to cover material in the discipline which is the subject of our guide (since 1966 it appears also in English). Each series comprises annually 12 issues included in cumulative volumes, with the "Chemistry" series comprising 24 issues. The material is arranged systematically, i.e. according to the frame work within which different problems are treated in the respective areas of science and technology. This framework is presented in a special publication, "Subject Headings of VINITI Abstract Journal", which is published annually. Some large portions of the main series included in the cumulative volumes appear in separate series of issues for the convenience of narrow subject specialists.

The "Referativnij Zhurnal" surpasses all similar publications in the length of its abstracts which average 1100-1200 characters and approximate the informative type. The time lag between the appearance of a publication and of its abstract has in many instances been reduced to 4 months, which is quite a low figure. Each series is provided with indexes: an author index to each issue and an annual one, and an annual subject index; the "Chemistry" series has, in addition, annual classified, formula and patent indexes. The VINITI Abstract Journal covers articles and other material from over 20,000 periodicals in 70 languages printed in more than 100 countries. The most significant scientific books and patent specifications are also abstracted.

"Chemical Abstracts", published by the American Chemical Society since 1907, covers scientific and technical materials in theoretical and applied chemistry, chemical technology, metallurgy and metallography, chemistry of foodstuffs, fuels, textile and leather industries, etc. Its abstracts are somewhat shorter than those in the VINITI Abstract Journal, being an average of 900 to 1000 characters in length. The material is arranged systematically. The journal appears semi-monthly, resulting in two volumes a year comprising 13 issues each. Each issue is provided with author and subject indexes; each volume has an author, subject and formula index; there are also cumulative indexes for 10 years (the latest appeared in 1956), and a list of abstracted periodicals including 7,000 titles. The distribution of the periodicals by countries is approximately as follows: USA, 25%; USSR, 20%; Britain, 15%; Japan, 10%; Federal Republic of Germany and German Democratic Republic, 10%; France, 5%; and other countries, 15%. The weekly current awareness service "Chemical Titles" is produced by computer, and covers 350 (mainly English-language) chemical publications. Fig. 9 compares two abstracts of the same paper which appeared in the VINITI Abstract Journal "Chemistry" series and in "Chemical Abstracts".

Among abstracting publications the following can be particularly useful: "Engineering Index", "United States Government Research Reports", "Mathematical Reviews", "Science Abstracts", "Nuclear Science Abstracts", "Analytical Abstracts", "Biological Abstracts", "Building Science Abstracts", "Computer Abstracts", "International Aerospace Abstracts", "Metallurgical Abstracts", "Copper Abstracts","Zinc Abstracts."

Express-information service is a broad and rather vague term embracing a variety of information services, e.g. sets of printed book cards, lists of new articles, lists of current titles of periodicals. The only feature they have in common is that all of them were established to speed up information on new publications. The emergence of this type of service is directly related to the considerable delays in most of the existing abstracting services, which is reduced multifold in express-information publications.

Recently, the term "express-information" has come to be increasingly applied also to periodical bulletins containing detailed abstracts or digests of the most significant inventions, discoveries or investigations, which must be promptly made known to specialists. This kind of publication is typified by VINITI's Express-Information which appears (in Russian) in 50 series on problems of topical interest, e.g. "Astronautics and Missile Dynamics", "Computer Technology", "Automatic Control of Production Processes", "Synthetic High-Polymer Materials", or on major branches of economy and industry, e.g. "Radio Engineering and Electronics", "Oil Production", "Railway Transport", "Ferrous Metal Industry". A total of 48 issues, each comprising 5 to 8 digests of articles, reports, invention specifications, i.e. some 300 different papers, are printed yearly in each series. Digests in the VINITI Express-Information appear approximately in a month's time after the original publication.

Library catalogues are listings of books held in a library, designed for the benefit of the readers to serve as a key to the library's collection. Library catalogues are the principal means to reveal the contents of library collections to popularize literature and to guide the reader in self-education. Catalogues have much in common with bibliographical card indexes, both being listings of publications. The distinguishing feature of the catalogue is that it is confined to the contents of a particular library, whereas a card index includes material irrespective of its location. Catalogues differ in the arrangement of entries, in the categories of material included, in the coverage of library holdings, in the language of material covered, in their purpose, in physical form, and some other characteristics.

According to the arrangement of entries, catalogues are divided into classified, alphabetical subject, and author and name catalogues.

In order to facilitate the selection of literature by subject, and to meet the topical requests of readers, information agencies and libraries keep classified and alphabetical subject catalogues. A *classified catalogue* arranges subject entries for books according to the parts of knowledge with which these books are associated. Within a subject field the entries are first grouped in classes, then in subclasses, and so on. The structure of a classified catalogue is directly related to the library or bibliographical classification adopted by the library. Some of these classification schemes will be described later. Classified catalogues serve as a useful tool in scientific information and play an important role in library activities. Grouping library material by fields of knowledge greatly facilitates reading guidance and selection of books. Classified catalogues are also used for in-house operations, such as acquisition or bibliographic and reference work.

From what has been said above one may be led to believe that a classified catalogue will always be instrumental in establishing what materials on a particular subject are available in a reference or library collection. If a subject is quite specialized, however, and the reader is not an expert in the given subject field, he will be hard put to it to determine under what subdivision in the classification scheme used in the catalogue he should look up the question of his interest. To facilitate work with a classified catalogue, alphabetical subject indexes, also called "keys", are constructed, which list in an alphabetical order all the subjects contained in the catalogue together with their symbols (classification numbers). An index to a classification schedule may also include subject headings which provide different formulations or groupings of the topics reflected in the classification system used, as well as synonyms of terms appearing in that classification system.

But even the key to a classified catalogue will sometimes fail when all the material on a particular topic has to be found. Imagine yourself searching for the most recent publications pertaining to aviation. In a classified catalogue they will be scattered over the different parts of knowledge, according to the aspect in which a give topic is presented, e.g., aircraft design - in engineering; aircraft construction - in industry; aircraft operation - in transport; military aircraft - in military science; occupational diseases connected with flying - in medicine; etc. Similar situations arise when one is looking for documents treating any other composite problem, such as a disease (e.g., problem of cancer), manufacturing process, farm crop, literature pertaining to a particular locality, or to the life and work of a person who contributed greatly to science, technology or public affairs. Often in such cases a catalogue which groups literature not by field of knowledge (as a classified one) but by some other subject complexes may prove to be helpful.

An *alphabetical subject catalogue* also reveals the contents of documents, but the entries are filed alphabetically by the subjects dealt with in the documents. The word or phrase denoting any one subject (notion) and used to express the subject of a document or of its part is called a subject heading. The arrangement of subject headings in an alphabetical subject catalogue will be made clearer by drawing an analogy between this and the already familiar book types, notably, a textbook where the material is arranged systematically, and an encyclopaedia where the material is arranged by subjects. Or take any scientific book: there the table of contents will simulate a classified catalogue, while the index will stand for an alphabetical subject catalogue. Alphabetical subject catalogues are eminently suitable for reflecting publications in applied sciences - technical, agricultural or medical. This is the principal reason for their popularity chiefly with special libraries, which use them to cover both domestic and foreign materials, books and periodical articles.

The *author and name catalogue* lists publications, irrespective of their content, alphabetically under their authors, names of institutions, or, in case of anonymous publications or publications having more than three authors, under titles. Such an arrangement enables the author and name catalogue to answer two principles questions: 1) Has the library a specific publication whose author and title are known? and 2) What publications of a given author, which have appeared as separate published items, are available in the library? If an author and name catalogue includes a sufficient number of added entries under persons that took part in the preparation of certain books, it will provide answers to some secondary questions (e.g. in what publications did a given person participate as compiler, editor, illustrator, translator, etc.). The importance of the author catalogue lies in providing the facility for answering questions on the availability of this or that item in the library collection. The author catalogue is thus one of the principal reference tools of an information agency or a library. This is explained by the fact that most readers when they come to a library already know what books they want. They use the author catalogue to check on the books sought, and find the class marks by which they are filed in the stacks. All library operations - acquisition, processing and storage of the holdings, service to readers, bibliographic and reference work - are only made possible by the existence of an author catalogue. Besides, an author catalogue is the only means of indicating what authors are represented in library holdings. It provides for the popularization of the publications of domestic and foreign classics of science, of outstanding personalities of today in all walks of life. Therefore it would be wrong to treat the author catalogue as a formal one and to assign it a subordinate place among the several types of catalogues.

The types of catalogue that have been described are the principal ones as regards the method of arrangement of their entries. Some documents, e.g. standards and patent specifications, are more usually sought by their numbers than by titles. Therefore, in such cases numerical catalogues are used in which entries are ordered not alphabetically by titles but numerically by the numbers assigned to them. Combinations of the principal types of catalogues are occasionally encountered. For instance, some special libraries maintain alphabetico-classed catalogues in which the main divisions are arranged systematically, i.e. according to a classification scheme, while the subject headings under these divisions are listed alphabetically. In the United States, the dictionary catalogue, which combines an author catalogue and an alphabetical subject catalogue, is widely used.

The bibliography is a special kind of scientific literature, and occupies a position of importance for science, culture and national economy as a means of bringing together, recording and classifying the contents of published and unpublished material. Bibliography has its own object, purpose, techniques and form. Written works, duplicated by some means and designed for indefinitely broad sections of contemporary and future readers, are the object of bibliography, while an immediate purpose is to influence the dissemination of written and printed records. The tracing and recording of written works, their selection according to some criterion, their description, sometimes supplemented with a critical appraisal, and their classification, constitute the techniques of bibliography; listings and literature surveys are its forms. This definition of bibliography - which, in our opinion, is the best of all existing ones - stems from the distinguished Soviet bibliographer K.R.Simon (1887-1966). (1) This term, bibliography, as explained in the first chapter applies also to the discipline dealing with the elaboration of theory, methodology, and history of bibliography as a special kind of scientific literature.

Bibliography, considered here as the last - but not least - of the types of secondary documents, holds a prominent position among other secondary documents in virtue of its importance, diversity and complexity: the whole system of bibliography, even in brief outline, could hardly be described in several special courses of lectures and certainly not in a single chapter. The only goal, then, we can set ourselves here is to attempt a general outline of the most important types of bibliographies with a few typical examples.

According to the content and kinds of indexed material, bibliographies may be general or special: a general bibliography lists material irrespective of its content and kind, while a special bibliography lists only publications covering a particular subject area or topic, or of a particular kind.

According to purpose, bibliographies can be enumerative, scientific, and recommendatory: enumerative, or recording, bibliography registers, as exhaustively as possible, material in different areas of knowledge, subject fields, or of specific character; scientific bibliography aims to inform specialists of the publications in their field of knowledge, on the theme, subject, or problem under study; recommendatory bibliography has the objective of active reading guidance and of helping certain groups of readers to choose the literature they need.

According to the time of publication, bibliographies are divided into current, retrospective, and prospective: current bibliography records material as it is published; retrospective bibliography covers a certain time period prior to the compilation of the bibliography; and prospective bibliography covers publications that are in press.

According to the place of publication of the literature indexed, bibliographies may be international, national or regional: international bibliography lists material in different languages published in different countries; national bibliography covers material appearing in a certain country or in a particular language - a state bibliography is a variation of national bibliography, including material published in one country, and is brought out by special State agencies, e.g. by book chambers in the USSR;

---

1. Simon, K.R. Bibliografija: osnovnye ponjatija i terminy (Bibliography: basic concepts and terms). Moscow, Kniga Publishers, 1968, p. 26-35.

regional bibliography records local material, i.e. publications issued within a region, district or any other locality.

Bibliographers distinguish many more types of bibliography according to different characteristics. We shall mention only two. Bibliography reflecting primary documents is commonly considered to be primary bibliography and is simply called "bibliography". Bibliographies listing bibliographies are considered to be secondary bibliographies and called "bibliographies of bibliographies", or guides to bibliographies. Work has long been in progress on compiling a bibliography of such guides, which would be a tertiary bibliography. Finally, according to the method of arrangement of entries, bibliographies - like catalogues - may be classified as subject or author.

The types mentioned above by no means exhaust the existing types of bibliography; they are meant to give some general idea of the multiplicity of bibliographical sources and to help in the appraisal of specific bibliographical indexes, of which there exist an enormous number. In the Soviet Union alone, thousands of titles of bibliographies and more than 150 special bibliographical periodicals are published every year, while all bibliographies published throughout the world run well into tens or hundreds of thousands. Mastering all this wealth of information sources offers great opportunities to an incipient scholar. It is no chance that early European writers used a golden key as the symbol of bibliography and its motto read: "Omnium scientiarum clavis" ("Key to all sciences").

The most typical literature requirements of a specialist are the need to follow all current publications in his subject area and in related fields, and the need, which generally arises at the start and at the completion of a study, to familiarize himself with all published sources pertaining to the theme of the study. The first need is met primarily by general current bibliographies, the second by special retrospective bibliographies. Since any particular bibliographic publication can be classified according to any of the characteristics listed above, we shall describe, by way of illustration, a few typical publications.

General bibliographies are typified by current enumerative national bibliographies. The Soviet current State bibliography is published by the All-Union Book Chamber as "Annals". The weekly "Book Annals" provide details of current books and pamphlets, as well as standards, technical catalogues, abstracts of theses and other separately issued material. All data appearing during the year in the "Book Annals" are cumulated with some abridgments in the "Bibliographical Yearbook of the USSR", printed in two volumes. Entries in both publications are filed systematically. They are provided with a subject index, an index of names, and an index of places. Details of newspapers, journals and magazines, bulletins, and serials are listed in the "Annals of Periodicals", which appear every five years with yearly supplements. Details of journal and newspaper articles are given in the weekly "Annals of Journal Articles" and "Annals of Newspaper Articles"; details of reviews and critical surveys printed in journals, newspapers and collections of articles are given in the quarterly "Annals of Reviews".

Similar publications exist in many countries. The popular American weekly "Cumulative Book Index", appearing since 1898, covers the books published in English in most of the countries of the world. Its specific feature is that it cumulates, i.e. it includes material from previous issues in subsequent ones. Cumulative volumes appear as substitutes for the April, July, September and December issues of the index and cover data for a period of, respectively, three, six, nine and twelve months and several years. The material is arranged as in a dictionary, with each book entered under its author, title and subject headings. The main medium of recording current British publications is the weekly "British National Bibliography", appearing since 1950. Details of new publications, arranged systematically, are cumulated in quarterly issues and annual volumes. France has the oldest national bibliography - the weekly "Bibliographie de la France", published since 1811, which covers books, first issues of periodicals, theses, official and other publications. The material is arranged systematically; an author index and an index of titles of anonymous publications are provided, as well as cumulative volumes printed every three, six, nine and twelve months. A special feature of this bibliography is that it announces books still in press, which are subsequently included in the cumulative listings "Books for the Week", "Books for the Month", "Books for the Quarter", "Books for the Half-Year", "Books for the Year", each provided with indexes of authors and titles,the last in addition with a subject index. The German Democratic Republic publishes the "Deutsche Nationalbibliographie", founded as early as 1931, which covers all publications appearing in the GDR, FRG and West Berlin, as well as books in German published all over the world. The "Deutsche Bibliographie", published in the Federal Republic of Germany since 1947, has approximately the same scope. Both bibliographies appear in several series as classified weekly issues which are cumulated every six months and five years.

Apart from these publications, designed for bibliographers, a range of general current bibliographical aids keep the scientists abreast of current literature. Details of the most significant books in all parts of knowledge which are published in Britain are provided by the "Reference Catalogue of Current Literature" (Vol. 1-2, London, Whitaker, 1965). American books are listed by the annual "Books in print" (Vol. 1-2, New York, Bowker).

Details of periodicals can be found in the following catalogues and lists:

- "Ulrich's International periodicals directory". Ed. by M. Rohinsky. 13th ed. Vol. 1-2. New York, Bowker, 1969; Supplement 1970.
- "Irregular Serials and Annuals. An international directory". Ed. by E. Koltay. New York, Bowker, 1970, 668 p.
- Gregory,W. "Union list of Serials in Libraries of the U.S. and Canada". 3rd ed. Vol. 1-5. New York, Wilson, 1966.
- "British Union Catalogue of Periodicals, incorporating World List of Scientific Periodicals: new periodical titles". London, Butterworths, 1960 - (monthly with quarterly, annual and 5-year cumulations).

Special bibliographies can best be described by giving examples of international and national special retrospective bibliographies:

Maizel,R.E. and Siegel,F."The Periodical Literature of Physics: a handbook for graduate students". New York, 1961 (American Institute of Physics).

"World Nuclear Directory: an international reference book". 2nd ed. London, Harrap, 1963, 626 p.

"Burman,C.R. "How to Find out in Chemistry: a guide to sources of information". Oxford, Pergamon, 1965, 220 p.

Bottle,R.T. (ed.) " The use of Chemical Literature". London, Butterworths, 1969, XII, 294 p.

Bottle,R.T., Wyatt,H.V. (eds.) "The use of Biological Literature". London, Butterworths, 1966, IX, 2[illegible] p.

Mason,B. (comp.) "The Literature of Geology". New York, American Museum of Natural History , 1953, 155 p.

Kerker,A.E. and Schlundt,E.M. "Literature Sources in the Biological Sciences". Lafayette, Purdue University Library, 1961, 133 p.

Smith,R. "Guide to the Literature of the Zoological Sciences". 6th ed. Minneapolis, Burgess, 1962, 232 p.

Moore,C.K. and Spencer,K.J. "Electronics: a bibliographical guide". London, Macdonald, 1961, 411 p.

Fry,B.M. and Mohrhardt,F.E. (eds.) "A Guide to Information Sources in Space Science and Technology". New York, Interscience, 1963, 579 p.

Examples can be cited ad infinitum, but even these few can give some idea of the nature of special retrospective bibliographies. The problem is, how to be aware of the existence of such bibliographies on a discipline, theme or problem you are interested in.

Secondary bibliographies (bibliographies of bibliographies, or guides to bibliographies) serve exactly this purpose. They may supply information on the whole universe of knowledge or a part of it. Examples of such guides are:

Besterman,T.A. "World Bibliography of Bibliographies". 4th ed. Vol. 1-5. Lausanne. Societas bibliographica, 1965-1966.

Malcles,L.-N. "Manuel de bibliographie". Paris, Presses universitaires de France, 1963, 328 p.

Totok,W., Weitzel,R., Weimann,K.-H. "Handbuch der bibliographischen Nachschlagewerke". 3. Aufl. Frankfurt am/Main, Klostermann, 1966, 362 p.

Mention should also be made of Soviet bibliographies of bibliographies:

Krichevsky,G.G. "Obshchie bibliografii zarubezhnykh stran" (General Bibliographies of Foreign Countries). Moscow, Publishing Office of All Union Book Chamber, 1962, 290 p.

Simon,K.R. "Istorya inostrannoi bibliografii" (History of Foreign Bibliography). Moscow, Publishing Office of All Union Book Chamber, 1963, 736 p.

Gudovshchikova , I.V." Obshchaia mezhdunarodnaia bibliografija bibliografii" (General International Bibliography of Bibliographies). Leningrad, Krupskaya Institute of Culture, 1969, 105 p.

New types of indexes

The bibliographical indexes which have been considered above originated at different times and have long been in bibliographical practice. They are naturally used also in information work, although some of their shortcomings have only recently come to be felt very acutely. One of the main faults with conventional indexes is the relatively long time required for their preparation. This is due to the fact that they are prepared exclusively by manual means and operations which do not lend themselves easily to mechanization and automation. Meanwhile, researchers and practical specialists feel a strong need for a bibliography with a short time lag between the issuing of a publication and the appearance of systematic data on it, which would minimize one of the basic shortcomings of the abstract journal, its slowness: considering all we have learned about the obsolescence of scientific documents, we cannot but admit the fairness of this requirement. Bibliography of this kind has come to be called "current awareness service", because its purpose is to make scientists aware of all current publications in their respective subject fields.

Although it was evident that the necessary promptness here could be achieved only if modern technology - in particular electronic digital computers - were to be used, the biggest share of time and effort is spent in systematization, which as yet defies all efforts at automation. A solution was found in the development of new kinds of indexes in which systematization is performed automatically, without human intervention. The production of the new indexes will be made clearer if we go back to the methods of arrangement of index entries. We have seen that most indexes use classified arrangement, while some are arranged by subject. Filing by authors is used only in a few bibliographies, which are in fact printed catalogues reflecting the contents of several library collections, i.e. union catalogues. Like any library author catalogue they indicate whether a given publication is available, and in what libraries. Now it has been discovered that alphabetization is rich in possibilities never before utilized in library catalogues and conventional bibliographical indexes.

Each word in the title of a document, composed by its author, is a potential subject heading, and the name of the author himself is a key to those bibliographical groupings which have been treated previously and which are established by means of citations. At the same time, punched card equipment and electronic computers have tremendous capacities for sorting and filing any information collections by formal criteria, in particular for alphabetization. This fact has been utilized in new kinds of indexes among which permuted-title and citation indexes enjoy the widest use in information work.

Permuted-title indexes are by definition based on document titles, most of which, in the fields of natural, exact and applied sciences, experience shows that they are fairly accurate identifiers of document contents. In the preparation of a permuted-title index, a computer programme using a so-called stop-list, identifies in the tiles - and skips - all common words (syntactic words such as pronouns, numerals, conjunctions, prepositions, particles, auxiliary verbs) as well as general nouns and adjectives (e.g. problem, study, method). The remaining words and terms appearing in document titles are used as subject headings: they are called keywords and are displayed in the index in the context of their title, which is why such indexes have been called KWIC (Key-Word-In-Context) indexes.

All keywords contained in abstract titles are alphabetized vertically in the search column, with the rest of the title words displayed to the left and to the right of the keywords. The index-line generally does not exceed 80 characters in length, including spacing. Therefore, if the text following a keyword does not fit into the right-hand part of the index-line, while there is room in the left-hand part, then the line is "wrapped around" i.e. transferred to the left.

If the title is still too long, it is truncated to match the length of the index-line. The end of the title is marked by an asterisk, which helps the title to be read correctly. If the title ending is cut off, the whole of the title can be reproduced in other parts of the index, which are easily located through title keywords. A title usually comprises not less than 4 or 5 keywords and therefore is displayed in the index at least 4 or 5 times. Because title words appear in the index in different groupings the index is called a permuted index, although it would more aptly be called a rotated index. In using a permuted-title index it should be remembered that the same concepts can be physically separated in it if they are expressed by synonymous words. For example, when looking for material on transistor triodes, it is necessary to scan - in addition to titles containing the word "transistor" - all titles containing the synonymous term "semiconductor" and its derivatives. The strategy of searching in a permuted-title index is easy to learn and usually presents no difficulty in practice.

Permuted-title indexes may be used as current-awareness publications or as auxiliary indexes to abstract journals and other secondary publications. In the former case they are provided with a name and a bibliographical index, and the codes located in the right-hand column denote the title abbreviation, year, number and the first page of the periodical in which the respective article appears. In the latter case these codes stand for the ID numbers of the abstracts in the abstract journal. In either case, these codes will be used to locate the entry which provides full bibliographical description of the publication.

Permuted-title indexes were first produced in the United States in 1958. They owe their introduction in information work to the late H.P. Luhn, the American engineer who is credited with being the principal originator of the method. The Chemical Abstracts Service publishes (since 1961) the semi-monthly current-awareness index "Chemical Titles", which announces new articles from 600 major journals in chemistry and chemical technology. Each issue, comprising details of 3,000 papers and including author and bibliographical indexes, takes 5 hours of computer time to compile. Another example of an extensively used permuted-title index is the index to the abstract journal "Biological Abstracts". Called "B.A.S.I.C." (Biological Abstracts Subjects in Context), it has appeared semi-monthly since 1962. It replaces the former subject index and is appended to every issue of the abstract journal. Its yearly coverage is about 120,000 abstracts. Regularly produced permuted-title indexes are scheduled for the immediate future for individual series of the VINITI Abstract Journal.

Citation indexes constitute a basically new kind of index: they supply answers to such requests from researchers and specialists as cannot be met by conventional indexes. The main question which they answer is, in what new publications are the former publications of particular authors cited. By cited publications are meant those to which bibliographical references are made in the texts of other publications. These indexes embody the bibliographic coupling method which was discussed in the last chapter.

The first, largest and best known index of this kind is the "Science Citation Index" (SCI) regularly produced since 1964 by the Institute for

Scientific Information (USA). In 1966, it covered 12,500 issues of 1,573 journal titles in all exact, natural and applied sciences. It included 3,000,000 references to former publications of different years contained in the 300,000 publications which appeared in 1966. The index is composed of two parts: the index proper (Citation Index) and the Source Index. The indexes appear in four issues each - three quarterly issues and one annual issue. The cumulative part of the Citation Index for 1966 comprised 5 volumes, and the Source Index 3 volumes.

The material in the Citation Index proper is arranged alphabetically under the names of the authors of cited documents; after each of these the authors of the source (i.e. citing) documents are listed, also alphabetically. Details of publications by the same author are filed chronologically. The name of the author is followed by the source of the publication, including an abbreviated journal title, year of issue of the cited or source document, volume, and the number of the first page; preceding the year of issue, a code is printed describing the source document type, viz. abstract, review, bibliography, patent specification.

In the Source Index, the material is also arranged alphabetically under first authors. In contrast to the Citation Index proper, where the first author alone is entered, here all the joint authors are given, if there are not more than ten. A specific feature of this index is the availability of the full title of an article or patent specification. It also gives the abbreviated title of the journal in which the article is published, the volume and page number, year of publication, the code for the source document, the number of citations, the issue number, and the ISI accession number.

Document searching in this index is conducted in the following manner. A starting (or target) reference is first established, which is known to the searcher beforehand, or obtained from conventional sources (encyclopaedias, textbooks, handbooks, guides, bibliographical indexes, abstract journals, library catalogues). Under its author, details of all the documents related to this target reference are found in the Citation Index. Full entries for the source documents are found in the Source Index, and their relevance to the query is assessed. If these documents are not sufficient, any of the earliest documents citing the target reference can be used as a new target reference for further searching. This cycle is repeated until all the necessary documents are identified or until the search yields no more information.

The technique of searching in a citation index can be illustrated using as example a search for documents on nucleic acids, which has been mentioned previously (Fig. 4). The 1941 paper by E. Rabinowitch and L. Epstein which was taken as the target reference, cites five other publications. If the earliest of the references listed under the first publication (L. Michaelis 1947 and 1950, filed one after the other) are taken as target references for further search, they will provide details of four more publications on this subject. Finally, using V. Zanker's publication as the target reference will help to identify three more papers. Thus, by knowing only the name of the originator of this theory, 13 out of 15 papers on the subject published between 1941 and 1960 can be found in three searches.

The most important and most typical feature of a citation index is that it provides answers of a topical nature, although the entries are arranged alphabetically under authors. All conventional indexes designed for this purpose require content analysis of each indexed document. This is the reason why they cannot be automated as yet. The "Science Citation Index" is fully computerized, because formal rules are used for its compilation; automatic grouping of material by subject is effected, as has been said before, by the principle of bibliographic coupling of documents.

The idea of such an index was first put forward by E.Garfield, now director of the Institute for Scientific Information, as early as in 1955. It was suggested to him by the Shephard's Citation (Chicago, 1873-) published in the United States since 1873, and in part also by bibliographies of reviews which use the same principle of material grouping and have been a standard feature of bibliography since the 18th century (the first bibliography of reviews appeared in Paris in 1771-1772). The importance of citation indexes is by no means confined to topical search. These indexes have helped to discover many scientific publication patterns, which were discussed in the previous chapter. They permit tracing the evolution of a scientific idea as it is reflected in the publications of different authors and help to establish how a problem stated in some paper has been treated - whether an author's theory has been supported, whether it has been applied and whether it was original.

All this assumes special importance in the context of the interpenetration of sciences, which is a sign of our time. If a paper by an author in a given field of knowledge is cited by many specialists in other fields, it points with certainty to the fact that the ideas expressed in this paper have proved fruitful for a number of scientific disciplines. The important point is that the establishment of new links between disciplines and theories is detected and recorded by citation indexes with a promptness which cannot be achieved by any other means, irrespective of the quality of any existing classification schemes or the skills and subjective judgments of indexers. In a citation index, the relationships of a given idea expressed in a document are established by the scientists themselves, viz. authors of the publications citing this document. The number of references to one or another document provides an objective basis for the assessment of the significance and viability of the ideas it contains.

Current-awareness information

A form of rapid information, called current-awareness service, has recently been gaining popularity. It consists in timely alerting (a maximum of one month following publication) of information users to the appearance of relevant material and is accompanied by a duplicating service supplying prompt and not too costly copies of this material.

According to the organization of the search technique, two types of current-awareness services can be distinguished: information and bibliographic publications, and selective dissemination of information. Information and bibliographic publications present data in a definite order and in various aspects - by subject, by author, by language, by time of publication, etc. - using the appropriate indexes.

The most extensively used current-awareness publications are classified or alphabetical subject indexes and permuted-titles indexes, discussed above. A specific form of current-awareness service are contents lists, compiled from the contents of recent issues of the major journals in particular subject areas. The most popular of these publications, "Current Contents" (USA) includes an index of authors and their addresses and appears weekly in three series covering more than 2000 journal titles in the natural sciences.

With selective dissemination of information, the information user entrusts to an information service the search for information that might interest him. He specifies his profile of interest which may include class marks, subject headings, descriptors, names of authors of the papers sought or their references. The material received by an information service is checked for these characteristics and, if it matches the interest profile of the user, he is supplied with the specific data that interests him. The feedback from the user provides for constant adjustment of document selection, thereby improving the personalized service.

A typical example of selective dissemination of information is the ISI's weekly service "Automatic Subject Citation Alert" (ASCA). This service notifies individual subscribers of the details of all the publications contained in the 1500 journals received by the Institute which match the title keywords or names of the authors of cited works, i.e. the characteristics included in the profiles of the subscribers.

We have discussed the most important kinds of secondary documents, information and bibliographic publications. They remain the main sources of data on scientific literature. This chapter is only the first step in getting familiar with them. It is our wish that at all stages of scientific work these tools should receive the most considerate attention. Familiarity with these sources and their skillful use are major elements of a scientist's qualification. But they should be treated critically, for along with science as a whole, they too are undergoing revolutionary changes. New index forms bear witness to it. The next chapter will deal with the activities of information agencies and libraries, which are responsible for the preparation and publication of secondary documents. Further chapters will describe the principles and procedures of their preparation and use.

Questions for self-checking

1. What are the typical features of a modern encyclopaedia?
2. What is the role of reviews of science and technology?
3. What is the structure of an abstract journal and what are its functions?
4. What principal answers are provided by library catalogues?
5. What are the existing kinds of bibliography and what purpose do they serve?
6. What is a permuted-title index and how is it used?
7. What new opportunities are offered by citation indexes as compared with conventional indexes?

Literature

1. Carey, R.J.P. Finding and Using Technical Information. A handbook for students, teachers, and practising engineers. London, Arnold, 1966, VI, 153 p.
2. Chandler, G. How to find out about literature. Oxford (a.o.), Pergamon, 1968, XV, 224 p.
3. Davidson, D.E. Periodicals: a manual of practice for librarians. 2nd ed. London, Grafton, 1964, 200 p.
4. Fowler, M.J. Guides to Scientific Periodicals. Annotated bibliography. London, Library Association, 1966, XVI, 318 p.
5. Garfield, E. Science Citation Index - A new Dimension in Indexing. "Science", 1964, vol. 144, p. 649-654.
6. Katz, W.A. Introduction to Reference Work. Vol. 1. Basic Information Sources. New York (a.o.), McGraw-Hill, 1969, XI, 376 p.
7. Luhn, H.P. Keyword-In-Context Index for Technical Literature (KWIC Index). "American Documentation", 1960, vol. 11, N.4, p.288-295.
8. Walker, J.R.A. Information Bulletins in Special Libraries. Background, Problems and Practice. London, The Libr. Assoc., 1966, VII, 128 p.

4. INFORMATION AGENCIES AND SPECIAL LIBRARIES - AUXILIARY INSTITUTIONS OF SCIENCE

Any study or R & D work requires certain knowledge on the part of the specialist concerned, who has acquired a portion of this knowledge during training, the rest accumulating in the course of his research and practical work, through the continuous contact with life and production, and in the course of his life-long self-education and raising of his qualifications.This knowledge determines the skills and constitutes the experience of a scientist or a practical worker. But however high the qualifications and broad the experience, they are not sufficient in themselves. To solve any scientific problem, the latest data on what has been, and is being, done in that field are an absolute necessity. In other words, initial scientific information has to be obtained.

Channels of dissemination of scientific information

We have familiarized ourselves with the manner in which scientific data are disseminated in a modern society, and in the two previous chapters we have discussed the system of scientific publications as the main medium for the communication of scientific knowledge. Here we shall discuss this problem in a somewhat different aspect to find out how scientific information is channelled, i.e. what routes it takes from the scientists who generate it to those who use it. Particular attention will be paid to the activities of those auxiliary institutions of science which are positioned along these routes.

Personal contacts and correspondence have long been the principal information channel. It is for this reason that the conventional form of a letter from one scientist to another has endured for many years in scientific publications. Personal contacts between scientists remain to this day one of the ways of information transfer. They are supplemented by the exchange of offprints and preprints of journal papers, by participation in scientific congresses, conferences and symposia. Recent investigations in the science of science, in particular those using the method of bibliographic coupling, testify to the importance of these channels of information transfer. It has been found that the scientific research front divides into small stretches, each corresponding to the work on a particular problem conducted in parallel by a small team of leading researchers (100-150 persons) in different countries and establishments, who form a sort of corporation, or "invisible college", as it has been called. Personal contacts, direct exchanges of information are of great significance for the members of these invisible colleges.

At the present time, however, the bulk of scientific information is disseminated through scientific documents.

Fig. 14 is a block diagram of the main channels of distribution of scientific information from generators to users. The outward lines represent personal contacts among scientists, through correspondence, exchange of offprints and preprints, and participation in congresses, conferences and symposia. The inner lines represent the distribution of unpublished documents, which after having received approval in the scientific or technological institutions, where their originators work, are normally made known to users by the scientific information agencies. Dissertations, defended in research and academic institutions, are made available to users through special libraries.

The bulk of scientific documents, it will be recalled, consists of published records, for which the channels of diffusion are at the centre of the diagram. The generators of these records, namely the authors of books, periodical articles or other material ready for publication, pass them over to the book-publishing houses or editorial offices of journals, as the approving bodies. Frequently, an author hands his manuscript to the publisher after a discussion and on recommendation from the scientific or educational institution where he is employed, but this is not reflected in our diagram.

Publishing house and printing office operations

In order to understand how scientific documents are published, it is necessary to envisage in general terms the major steps in the publishing and printing routines. Manuscripts of papers handed over by their authors to the publishing office are reviewed to evaluate their theoretical and practical value; those accepted for publication are thoroughly edited from the scientific and literary points of view. The editor sees to it that the author's ideas are stated clearly, the terminology and formulations correct, and the style concise; he verifies as far as possible the reliability of the facts and figures cited. After editing the manuscript is retyped. It is then read by the proof-reader, who is reponsible for the spelling and punctuation, and by the copy preparer who decides all typographic features of the future publication: the type size and face, the method of illustration reproduction, the set of type and paper size, the quality of the cover and binding. This completes the editing routines, and the manuscript proceeds to the printing office.

The complex printing process can be broken down into three steps: preparation of the forme or composition, presswork, and binding. Composition may be performed by hand, but more often it is done mechanically, on line-setting machines, or linotypes and monotype composing machines, or monotypes. The composed lines are put together in galleys, from which a trial print is made and read by the printer's readers or sent to the publisher for reading. After making corrections in the galleys, the makeup follows, i.e. type pages of the same size are formed from the setup type. The proof sheets of the imposition are read again in the publishing office (and usually also by the author), and the composed matter is corrected accordingly. This procedure is repeated for the third time as the so-called revision, following which the forme is considered to be ready for the press. The forme is used in the presswork, i.e. in the preparation of the necessary number of prints. The work on a publication in the printing office completes with the binding of pages. They are folded into signatures, usually 16 pages large, collated into an unbacked book, and stitched. The unbacked book is trimmed on three sides, provided with a cover or inserted into the binding.

After a book has been produced, its run is forwarded to the bookshops, while journals are mailed to the subscribers. It is self-evident, however, that the buying of books for personal libraries and individual subscriptions to journals can satisfy only a small portion of scientists' and practitioners' information needs. To put the case, the limited resources of a researcher have only to be compared with the huge and ever-growing volume of scientific literature.

Interrelationship of information agencies and libraries

This explains why information agencies and special libraries serve as the major auxiliary institutions of science, carrying out the functions of intermediaries between generators and users of scientific information and providing for the normal operation of information channels through which scientific documents are disseminated.

The tasks of these institutions have much in common as they largely correspond to the tasks and stages of scientific information activities which were described in the first chapter. Their mission is to identify and collect scientific information, to analyse and process it into forms suitable for storage and searching, to provide for its prolonged storage and dissemination, including retrieval on demand. But each of these institutions has certain peculiarities: different particular tasks solved by different subject specialists employing different techniques. We have to reveal these differences in order to learn how to use better the opportunities offered by information agencies and libraries.

One of the differences related to the staff of these institutions, was well stated by a group of leading American scientists in a report to the President of the United States: "In short, knowledgeable scientific interpreters who can collect relevant data, review a field, and distill information in a manner that goes to the heart of a technical situation are more help to the overburdened specialist than is a mere pile of relevant documents. Such knowledgeable scientific middlemen who themselves contribute to science are the backbone of the information center; they make an information center a technical institute rather than a technical library". (1) Thus, an information specialist is a specialist in a certain field, a researcher participating in the solution of a scientific problem, who provides the research team with information. On the other hand, a librarian - who may also have a training or even qualifications in one of the broad fields of knowledge - is primarily an expert in reader psychology, proficient in teaching methods, and capable of seeing his way in the multifarious book production environment. An information specialist and a librarian are experts in different fields.

The ultimate goal of an information service is to notify new facts and ideas, to provide answers to questions about facts, e.g. "What are the

---

1. Science, Government, and Information. A report of the President's Science Advisory Committee. Washington, 1963, p. 33

properties of a given object?" "What objects possess given properties?" "What is known on this problem?" Since answers to these questions today are provided by scientific documents - books, journals, patent specifications, technical reports,etc. - information agencies handle these and will frequently supply only references, not data, e.g. they state what facts are to be found in certain documents or what documents contain the relevant information. Hence, working with documents is not the primary task of information agencies, but only a means of logically processing of scientific information contained in them in order to obtain new data. The library is an educational and auxiliary scientific institution providing for public use of the library material, designed to help the reader in choosing his books, and to guide him in his reading. Special libraries are an important link in scientific information services and one of the most important auxiliary means of elaboration of strategic problems of science, national economy and culture. Therefore, these auxiliary institutions of science have differing ultimate objectives, although now they work with largely the same scientific documents.

Another difference pertains to the individuals whom these institutions serve. While these are the same persons, namely, scientists and practical workers, they are called differently by the information scientists, who use the term "information users", and by the librarians, who use the term "readers". It should be admitted that the existence of different terms is justified to the extent that they express different definitions. An information user turns to an information centre or its publications in order to obtain specific data which is necessary for his research work; he has a right to request an accurate and complete answer to a specific query from information specialists who keep track of the advances in this specific field by means of the documentary sources. In the library, the same person appears in the capacity of a reader, who is to be provided both with the material pertaining to the topic of his immediate concern, and with books to broaden his world outlook, improve his skills, inform him of related fields of knowledge, and satisfy his general scientific and broad social interests.

And finally, the institutions under discussion operate in different modes; while this difference is not one of principle, it is nevertheless apparent in practice, and is probably not accidental. Most information agencies provide their services on a regular basis, which corresponds to the rate of advance of science, to the frequency of appearance of new publications and unpublished material. Information agencies attempt to cover as fully as possible new documents, to subject them to analytico-synthetic processing, and then to bring their contents to scientists in equal portions in the shape of separate issues of periodically appearing abstract journals, surveys, express-information bulletins or bibliographies. True, information agencies will also preserve documents for prolonged storage, and for searching and retrieval at the request of users, but this work occupies a secondary place in most information agencies. Libraries, on the other hand, operate in the "question - answer" mode, accumulating material, revealing its contents, and organizing it for storage in such a way as to be able to produce the material on request. It is a fact that almost every library regularly informs readers of its accessions, but it is also true that this work is far more limited in scope than their on-demand service to readers.

Organization of scientific information activities in the USSR

Information work in the Soviet Union is based on the principle of centralized processing of the world's scientific and technical information and its dissemination to the different branches of national economy. This means that all kinds of scientific documents are processed in centralized state agencies, with their output in the form of abstract journals, express-information bulletins, surveys, bibliographical indexes and printed cards distributed among the various research institutes, design offices, industrial enterprises, institutions of higher learning, as well as branch and regional information agencies. On requests from these institutions and offices, the centralized information agencies furnish copies of the original primary documents announced in the secondary publications enumerated above. This is usually called a descending flow of scientific and technical information. Another descending flow of information to the information centres of enterprises and institutions is generated by the branch-oriented information bodies in the form of newsletters or information leaflets, surveys and other documents in specific fields of science, engineering and national economy, as well as by the territorial (republican and regional) information bodies which publish state-of-the-art reports on problems of topical interest to the enterprises and institutions of a given region and other scientific documents in local (non-Russian) languages.

In contrast to this flow, there is an opposite one which is called the ascending flow of information. It consists of R & D reports and information cards with details of the studies conducted, which are directed by research institutions and educational establishments, design offices and industrial enterprises to the All-Union Scientific and Technical Information Centre and to the branch-oriented information centres for registration and storage.

All information agencies in the Soviet Union thus fall into four large groups: all-union (national), branch-oriented, regional, and in-house centres serving the needs of enterprises and institutions. The all-union information agencies include the All-Union Institute of Scientific and Technical Information (VINITI), the All-Union Scientific and Technical Information Centre (VNTIC), the Central Research Institute for Patent Information (CNIIPI), and the All-Union Research Institute for Technical Information, Classification and Coding (VNIIKI). The branch-oriented information agencies include a few dozen central institutes for scientific and technical information and technical and economic research (CINTI) in specific areas of science, technology and national economy, each with its own central scientific and technical library (CNTB). The regional information agencies include 15 republican institutes of scientific and technical information with their respective republican scientific and technical libraries (RNTB), as well as several dozen central bureaux of technical information (CBTI) in the regions, with their central scientific and technical libraries (CNTB) and houses of popularization of science and technology (DNTP). Finally, the information bodies at enterprises and institutions comprise scientific and technical information departments (ONTI) of scientific research institutes (NII), design offices (KB) with their own scientific and technical libraries (NTB), as well as technical information bureaux (BTI) with technical libraries at enterprises. There are about 5,000 of these ONTIs and BTIs. The same functions are performed also by tens of thousands of separate technical libraries functioning in enterprises, and academic libraries in institutions of higher learning.

General supervision over scientific information activities is exercised by the State Committee for Science and Technology of the Council of Ministers of the USSR. The ministries and departments organize and control the work of the corresponding branch-oriented information agencies.

The *All-Union Institute of Scientific and Technical Information* (VINITI) of the State Committee for Science and Technology of the USSR Council of Ministers and of the Academy of Sciences of the USSR was set up, as has been mentioned, in 1952. In the past 19 years, it has grown into the world's largest information centre, employing some 2,500 full-time staff and more than 22,000 part-time research specialists. The volume of its publications exceeds 30,000 publisher's sheet signatures a year, which ranks it with the largest publishing houses of the Soviet Union. VINITI has large printing resources which include its Publishing and Printing Department with a staff of over 1,700. The main functions of the Institute are as follows: study and processing of the world's scientific and technical literature in all fields of knowledge; preparation and publication of abstract, survey and bibliographical information; duplication of the original publications on requests from institutions, enterprises and individuals; organization of information and reference services; co-ordination of translation activities in the country; publication of collections of digests of articles of topical interest ("Express Information"); conducting and co-ordinating research in informatics; and organization of advanced training courses for specialists employed in information centres.

VINITI can serve as a good example to show the internal work of a large information centre, and the procedures of preparation of abstract publications. The daily batches of material coming to the Institute are forwarded to the Department of preliminary scientific processing, where the scientific value of each publication is assessed, and the relevant series of the Abstract Journal where it is to be indexed are indicated. The publications selected are marked with special symbols, and passed over to bibliographers who prepare them for description by underlining the necessary bibliographic data. A card is prepared for each publication which accompanies the publication throughout its processing cycle. Card copies are added to the catalogues and auxiliary indexes. Because most of the publications go simultaneously to several series of the Abstract Journal, the appropriate number of their copies is made, while the original is stored in the Department of scientific resources.

After the Department of preliminary scientific processing and the bibliographic group, copies of publications with cards are distributed among the subject departments where they are assigned to the staff and part-time abstractors, according to their type, content, and language. The abstractors return the copies of the publications together with an abstract, which is then thoroughly edited and checked for errors in a relevant subject department. Then the abstracts, classified according to a special list of subject headings, are grouped for the next Abstract Journal issues; following a discussion in the editorial board, the manuscript of the journal is sent for publication to the Publishing and Printing Department, where it is submitted to the usual publishing routines already discussed at the beginning of this chapter. Imagine the high level of discipline and co-ordination that must be maintained by the numerous persons participating in these processes in order to complete them within four months which, you will recall, is the optimal information delay. And these are the techniques and procedures of only one of the various functions of VINITI.

Now let us familiarize ourselves with the functions of the other national information agencies. The *All-Union Scientific and Technical Information Centre* (VNTIC) of the State Committee for Science and Technology of the Council of Ministers of the USSR at the present (initial) stage of its development has the following tasks: recording and bibliographical control of the R & D projects carried out in the country; maintenance of a file of R & D documentation; publication of regular information on technical reports accessions; reference and information services.

The *Central Research Institute for Patent Information* (CNIIPI) of the USSR State Committee for Inventions and Discoveries has the following functions: scientific processing of the world's patent materials; preparation and publication of current-awareness and survey information for the field (abstracting services are provided by VINITI); information provision for patent examination work; reference and information services on the domestic and foreign patent materials; publication of information material on problems of invention and patent work. The *All-Union Research Institute for Technical Information, Classification and Coding* (VNIIKI) of the USSR State Committee for Standards, Measures and Instrumentation performs the following tasks: scientific processing, preparation and publication of analytical information on standards; reference and information work on standards and interdepartmental norms; conducting and co-ordinating research projects on classification and coding systems; organization of a State Service of Standard Reference Data (GSSSD); publication of information material on standardization problems.

The *branch-oriented information agencies* are represented in the Soviet Union by the central institutes of scientific and technical information and technical and economic research (CINTI) under the different ministries and departments, as well as by the scientific and technical information departments (ONTI) of the leading scientific research institutes (NII). They have the following objectives: to ensure scientific processing of information on the advances made in specific organizations, institutions and enterprises in the various fields; to prepare current-awareness, abstract and survey information; to provide reference and library services over the range of specific subjects in a given field; to conduct analytical survey and comparative technical and economic studies of the problems of the development of the field; etc. They lay particular emphasis on a systematic study of the technical and economic performance indices of industrial enterprises, of the design and operative technical and economic data on products and technology, as well as of the methodology or organization of production, research and development work.

Regional information agencies, as already mentioned, are represented by different types of institutions. Republican institutes of scientific and technical information provide the following services: information services for the Party and government bodies of the Union Republics; reference, information and library services to the enterprises and institutions within the respective republic; dissemination of interdisciplinary information within the republic; preparation and publication of information material in the corresponding national languages; co-ordination and implementation of popularization of science and technology in the republic; etc. The central bureaux of technical information (CBTI) and houses of popularization of science and technology (DNTP) in regions, cities and other localities carry out similar functions within their respective administrative areas.

The *information bodies of enterprises and institutions* are represented by scientific and technical information departments (ONTI) or bureaux of technical information (BTI) in enterprises, research institutes (NII), and design offices (KB). They are entrusted with the following tasks: prompt and accurate information and library services to scientists, specialists and industrial workers; popularization of progressive scientific and technical experience; control over the utilization of information material; supply of scientific and technical information and documentation to the corresponding all-union and central branch-oriented information agencies. These then, presented in very general terms, constitute the framework of information services and the missions of the main information agencies in the Soviet Union.

Special (research) libraries

Now let us get a bird's eye view of the libraries, another important type of auxiliary institution of science. According to their function, all libraries are grouped into public and special (research) libraries. *Public libraries* provide their readers with the most significant literature which is necessary for the all-round development of man, for his education, and mastering of general and polytechnical knowledge. They are not oriented towards any particular reader category and engage in active popularization of books and in reader guidance.

The mission of the *special (re[illegible] [illegible]ries* is to assist in research and development activities, in i[illegible] [illegible]ricultural production, in the advancement of construction, [illegible] [illegible]rt, [illegible]ic health, communications, social and cultural life of the [illegible] [illegible] well as in the general and special training and upgrading of the personnel in these fields, by providing for public use library material, by guiding specialists in their reading with the aim of improving their ideological, political, scientific, professional and cultural levels. These libraries are of four principal types: 1) large independent scientific libraries of universal type; 2) libraries of research institutes; 3) college and university libraries; and 4) industry-oriented libraries.

Large independent universal libraries are a special type of library. As well as meeting the information needs of readers at large - specialists in science, industry or culture - they also discharge cultural and educational functions which are typical of public libraries.

Libraries of research institutes provide scientific literature to their research teams.

College and university libraries, apart from serving scientists, assist in the educational process and play a significant role in training specialists. University libraries hold a special position in that they help the students in keeping with the specifics of university education, they acquire a broad scientific background.

Industry-oriented libraries help practical specialists in resolving their production tasks. They form a majority of the special libraries and include technical, medical and agricultural libraries, as well as libraries of transport and communication institutions, construction and similar organizations. In those institutions and organizations which have scientific information centres of their own, libraries form subdivisions of these; where such information centres are lacking, their functions are performed by special libraries.

Library routines

With all the diversity of library types, their structure and operations are basically the same and can be described in a generalized form. The main library routines are acquisition of library material, its cataloguing (or scientific processing), organization of library holdings, and service to readers. To these four functions correspond the principal divisions of a special library which are, as a rule, functionally structured the acquisition department, the cataloguing (or processing) department, the shelf- or stack department, and the service department. Many large libraries also have a bibliographic (or bibliographic reference) department, research and methodological department, administrative and other departments. As the Indian library scientist S. Ranganathan has put it, "Library is a growing organism". Transfers of large masses of literature are taking place in it daily. Two flows are clearly seen: one consists of the new books and periodicals received by the library, and the other of the literature issued on requests from the readers and returned after having been read. The first flow has been called "the way of the new book", the second "the way of the reader's request".

The starting point in "the way of the book" is the acquisition department. There, on the basis of a special document stating in general terms the reader's demands, which is called "acquisition policy", it is decided which of the new publications should acquired by the library. Librarians learn of the appearance of these publications from current bibliographies. The sources of acquisition are bookshops and specialized library institutions - book supply agencies - carrying out centralized acquisition, as well as other libraries with which book exchange is maintained. The purchased books are accessioned, after which they are considered added to the collection. The accessioning consists in marking a new item with the library stamp and an accession number. The accessions records are used in the inventory of accessions.

Now the accessions go to the cataloguing (processing) department where a card is prepared for each item, containing its main bibliographical details, and its subject briefly identified by means of conventional symbols - classification marks or subject headings. These symbols are inscribed on the book and the card, following which the book proceeds to the shelves, while the card is reproduced in the necessary number of copies. The cards are put together to form catalogues of libraries, described in the previous chapter. The system of catalogues in a library, which may comprise up to several dozen catalogues - author, classified and alphabetical subject - introduces both the readers and the staff to the library holdings.

The literature is shelved in a certain order, varying according to the principle of arrangement adopted in the library. In case of a classified arrangement, it resembles the filing of entries in a classified catalogue, books being filed by fields of knowledge according to their subjects. However, unlike a catalogue entry which can be placed under several divisions corresponding to the different subjects dealt with in the book, the book itself can only occupy a single place on the shelf. Classified arrangement is advantageous for a small library where the book collection does not exceed a few tens of thousands volumes. The librarian is enabled then to give advice to a reader straight from the shelf; he can more easily memorize the location of each book and the general layout of the stacks. The library can provide open access to the shelves, allowing readers to make their own selection among the books, though even then catalogues offer a more effective means of interpretation of the book collections and recommending books, since they reflect the books in several aspects, as well as those books which are off-the-shelf at the moment. Most public libraries use the classified shelf arrangement of books.

In larger libraries, whose holdings may amount to several hundred thousands of volumes, the advantages of classified arrangement recede, while its shortcomings become more pronounced. Spare place for accessions has to be provided in each section, class and subclass, and when found lacking, has to be made through the labour-consuming shifting of the holdings. Because books of the same content vary widely in size, the vertical shelf spacing must be such as to accommodate the largest book. Hence shelf space is lost both vertically and horizontally, as a librarian would say. For this reason larger special libraries increasingly adopt one of the form arrangements, which take into consideration certain form characteristics, and not the content of the books. The most popular method is shelving by size, when the books are placed on the shelf according to their height (size) and the date of arrival in the library (accession number). Such organization of the collection saves much space in the stacks and ensures a prompter and more accurate issuing of books to readers. In this case a book's call number (or its location mark) will not be its classification number, as with classified arrangement, but its format mark and accession (or serial) number.

Filing the cards for accessions in catalogues, and the new items in the stacks completes "the way of the book". Now these publications are ready to be used. A reader, when he comes to the library, finds out what publications he wants. To do this he makes use of secondary documents: bibliographical indexes, card indexes, abstract journals, card catalogues. In the catalogues the reader finds out whether the publications he needs are in the library holdings and looks for their call numbers. He is assisted in his search by the bibliographic reference department of the library or by a bibliographer on duty. He then fills in his requests for the books, giving the main bibliographical details and the call numbers. This starts the second library flow, known as "the path of the reader's request". The request form, sometimes checked for possible errors in a special division, is then sent to the shelf or stack rooms, where the staff selects from the shelves the required books. These are forwarded to the delivery station from which the reader takes them for his use. The time lag between the moment of handing in the request to the moment of book issue to the reader may vary in different libraries from a few minutes to several hours.

Provision of library material to readers is effected in two ways: a reader can use it in the reading room, or take it home. In the former case the literature is issued for the period until the library's closing time, and reserved for this particular reader on special shelves for several consecutive days. In the latter case, the reader is issued his literature for a fixed period of time ranging from one week to one month in the lending (circulation) department. In some libraries the books issued in the reading room, and those loaned to the readers have different physical locations in the stacks and are registered in separate catalogues. As a rule, the delivery station or the circulation department will keep a file of reader's cards, in which the issued books are recorded, and a file of book cards in which the names or numbers of the readers who are issued a given book are marked. Thus, the librarians at any moment can tell what books are borrowed by a given reader, when he must return them and who is reading the book which is off-the-shelf and requested by another reader. On return of the loaned books, they are restored to their places in the stacks. This completes the second continuously circulating flow of library material. It should be added that if the books requested by a reader are not available in the library, they can be ordered from other libraries - within the country, or even abroad - using the so-called interlibrary loan service, which links together many special libraries.

Thus, special libraries are important auxiliary institutions of science charged with providing scientists and practical workers with the necessary scientific documents. Ideally, every such library should give prompt and complete answers to a request asking what documents exist on this or that subject or theme, and, after the reader has specified the type of the material wanted, date of publication, language and aspect of treatment of the corresponding subject, should submit these documents for the reader's use. Actually, special libraries by no means display such a happy state of affairs. Not one of them, nor even their totality in any one country, can hope to accumulate all the world's scientific literature, to say nothing of all scientific documents. Moreover, although libraries attempt to cover all bibliographical publications, the sheer bulk and diversity of this material will not permit a timely and complete coverage of all the data on the existing scientific literature in a given subject. Catalogues of libraries reveal only the main items of the contents of scientific documents and are not meant for multi-aspect literature search. The majority of basic library routines are performed manually, they are poorly mechanized and totally lack automation. The time it takes to locate, retrieve and use library material far surpasses the time that a researcher or practitioner is prepared to spend for his information work. Hence a certain dissatisfaction with the present-day special libraries and the persistent efforts to develop techniques and means of improving their operation, which are frequently labelled, not too felicitously either, as the techniques and means of creating "the library of the future". It would be more appropriate to speak of the necessity of creating a truly modern library.

A World Information System and the developing countries

It is a very important task to coordinate information activities throughout the world, to organize a wide exchange of primary and secondary documents, programmes for machine processing of texts in natural languages, uniform information media and so on. But this is a very difficult problem.

The organizational principles and the build-up of a national system of scientific and technical information are entirely determined by historical, economic, social and other conditions prevailing in the country concerned. Technical means employed in scientific information activities depend on the country's area and population, level of technological development, economic conditions and a great many other factors. For that reason any attempt to make general recommendations that would be valid for the majority of nations or for the world are doomed to failure.

This is particularly true of developing countries. In this connection, J. Dean from the Institute of Librarianship, University of Ibadan (Nigeria) said in his review of L. Asheim's book "Librarianship in the developing countries" (Urbana, University of Illinois Press, 1966, 95 p.) "The main limitation of Dr. Asheim's book is implicit in its extremley broad approach. He deals neither with particular developing countries, nor with a group of developing countries, but generalizes about all developing countries as though they possess to a great extent a common identity. This is not so. These countries comprehend a vast area of the world; the whole of Latin America and most of Africa, Asia and the Middle East. They present a picture of infinite variety. Each has been subjected to different historical, economic and political influences and has its own individuality...". Dr. Asheim's list of characteristic defects in the library systems of the developing countries - he deplores the almost total absence of public libraries and of service to children, the existence of national libraries with seemingly no function, the lowly status of the librarian in the community and the fact that the book so frequently has primacy over the reader - is reasonably valid for some groups, but not for others. For example, if we consider the library situation in those English-speaking territories which have achieved independence during the past ten years or so, we shall find some at least of his criticisms very wide of the mark. Public library systems have been established in a number of these territories, university libraries are making significant progress, the primacy of the reader over the book is generally accepted and last, but vitally important, librarians everywhere are accorded considerable status in their various communities.(1)

A major obstacle on the way towards coordination of the information activities throughout the world is the growth of "language barriers". The solution of the language barrier problem becomes increasingly complex and urgent from year to year. Whereas early in this century the main languages used for scientific communication were English, German and French, nowadays 50 percent of scientific and technical literary output is in languages unknown to more than half of the scientists of the world. Over the past decade, the proportion of scientific and technical literature published in Latin American, African and Asian countries in Spanish, Portugese, Hindi, Urdu, Arabic, Swahili and other languages has been rapidly growing.

We believe that the only way of overcoming the language barriers is the solution of automatic translation problems on the basis of an intermediary language. Coordination of national programmes for the compilation of bilingual and multilingual dictionaries in science and technology would be very useful. An important prerequisite for a World Information System is the compatibility of national scientific and technical information systems available or being developed in different countries, which necessitates the elaboration of international standards for information media, codes and types, transcription and transliteration systems, abbreviations of periodical and serial titles, bibliographical references, etc. Also necessary are uniform international rules for the preparation and presentation of primary publications, with special emphasis on providing adequate titles for these publications. It seems to us that it would be most convenient, when

---

1. "Unesco Bulletin for Libraries", 1968, vol. 22, N.2, p. 91-92

publishing a document, to index it by the Universal Decimal Classification (UDC) and keywords (descriptors) and to provide it with a brief abstract.

These tasks have to be solved through the joint efforts of the international organizations concerned:

- United Nations Educational, Scientific and Cultural Organization (Unesco),
- International Council of Scientific Unions (ICSU),
- International Federation for Documentation (FID),
- International Federation for Information Processing (IFIP),
- International Federation of Library Associations (IFLA),
- International Organization for Standardization (ISO).

To ensure effective coverage of developing countries by the World Information System, wide-scale measures should be taken in these countries to train skilled information personnel. Among other things, this calls for the speedy provision of appropriate curricula, programmes, study aids, and scholarships for specialists from developing countries, as well as for the creation of a network of training centres both in the advanced and developing countries. This will enable some 80 developing countries to improve their information activities and so speed up their economic and cultural development.

Thus, we have considered in this chapter several themes which should broaden our view of the routes or, as we call them, channels of dissemination of scientific knowledge. You should now be familiar with the fundamentals of organization of information activities, the principal types and functions of the information agencies and special libraries and the stages and routines in the operation of these auxiliary institutions of science. This will help you in making use of these institutions in your work and will also facilitate your understanding of the basic principles of information retrieval, which is the subject of the next chapter.

Questions for self-checking

1. What are the channels by which scientific information is disseminated from generators to users?
2. What are the main stages in the publishing and printing processes?
3. What are the differences between scientific information agencies and special libraries?
4. What principles underlie scientific information activities and how are they implemented?
5. What are the functions of the major information centres in the USSR?
6. What are the missions of special (research) libraries?
7. What are the main library routines?

Literature

1. Asheim,L. Librarianship in the Developing Countries. Urbana, University of Illinois, 1966, VII, 95 p.
2. Brogan,G.E., Buck,J.T. Using Libraries Effectively. Blemont, Dickenson, 1969, X, 116 p.
3. Clapp,V. The future of the Research Library. Urbana, University of Illinois, 1964, X, 114 p.
4. Dembowska,M. Documentation and Scientific Information. Outline of problems and trends. (Translated from Polish). Warsaw, Central Institute for Scientific, Technical and Economic Information, 1968, 154 p.
5. Gelfand,M.A. University Libraries for Developing Countries. Paris, Unesco, 1968, 180 p.
6. Giljarevskij,R.S. International Distribution of Catalogue Cards. Present situation and future prospects. Paris, Unesco, 1969, 84 p. (Unesco Manuals for Libraries, 15)
7. Kent,A. Specialized Information Centres. Washington, Spartan Books; London, Macmillan, 1965, 290 p.
8. Shera,J.H. Of Librarianship, Documentation and Information Science. "Unesco Bulletin for Libraries", 1968, vol. 22, N.2, p. 58-65.

5. ESSENTIALS OF INFORMATION RETRIEVAL

The problem of information retrieval is central to informatics. By now you have learned how scientific data are communicated in the contemporary world, what the sources of scientific information are, what institutions participate in the flow of scientific information and provide for its transfer from the generators to the users and what difficulties arise in conducting information work by conventional means.

Further improvement of this work calls for mechanization and automation of information routines, which in turn requires that these routines be formalized. In order to understand the gist of these processes and to identify their common features, they are embraced by the concept of information retrieval. The information retrieval problem can be viewed as the constant accumulation of an ever-growing bulk of scientific information, on the one hand, and the growth and increasing complexity of information needs of the specialists, on the other hand.

In a generalized form the problem can be seen as the need to select for each user from the totality of the available information only the data he needs at the moment to conduct scientific research or practical work. It is quite evident that no specialist is able to read all existing scientific documents in order to select the essential ones, so they are stored in special collections, and the users wishing to extract from these the necessary documents use a certain procedure which is called information retrieval.

This procedure can be illustrated by a hypothetical example shown in Fig. 16. Let us assume that we have a collection of ten documents. To eliminate the need for scanning the whole file each time the documents containing the required data are to be located, their contents have to be analysed beforehand according to certain characteristics and with a view to possible search questions. If now all the documents are numbered and the characteristics coded by letters, and each document number is linked with its content characteristics, a certain system permitting formalization of the document retrieval process will evolve. Let us assume further that documents represented by combinations of characteristics C and D or characteristics C and F are to be retrieved. Successive scanning of the columns in which the availability of a characteristic is marked with a dot, quickly shows that documents 1, 3, 6 and 8 present the above combinations.

This example shows that to implement information retrieval, it is necessary first to select documents for the system which we call an information retrieval system, and to assign each document a number which will serve as its address in the system. Then a range of characteristics by which these documents will be searched has to be established. This multiplicity of characteristics, expressed - according to certain rules - by terms or conventional symbols (e.g. numbers), constitutes an information retrieval language.

The subject matter of each document is matched against the terms or numbers of this language in order to establish which of the characteristics they represent are contained in the document. The terms or numbers selected as the result of this matching procedure form the search pattern of a document; its formation is called indexing.

During a search, a user's query is also expressed in terms of the information retrieval language and the search request formulation thus obtained is matched against the documents' search patterns. If they completely or partly match, the decision is taken to furnish the appropriate documents to the user. The necessary level of matching is determined by a specified matching criterion.

Basic notions

We shall now define and interpret the basic notions of this section of informatics. Information retrieval is a multiplicity of consecutive operations performed to locate the necessary information or documents containing it, with subsequent retrieval of these documents or their copies, and is effected by means of information retrieval systems.

An information retrieval (IR) system is generally formed by an information retrieval language and a matching criterion designed for information searching in a given information collection. Specific IR systems are realized by means of certain technical facilities, which are nevertheless not included in the abstract notion of an IR system, because the same system can be realized by using different means. These technical facilities, various equipment and machinery will be considered in a separate chapter.

The systems divide into document retrieval and data (fact) retrieval systems. Document retrieval systems produce, in response to a query, documents containing the information sought, their copies, or their addresses in the story. Data retrieval systems are designed to produce facts, e.g. the properties of a particular substance, the structural or molecular formula of a chemical compound, the characteristics of a particular biological species, or the names of those species which possess certain characteristics. The common feature of these systems is that they can retrieve only the information that has been introduced into them beforehand.

At the present time, investigations are under way aimed at creating logical information systems, in which logical processing of information will be possible. Such systems will permit obtaining in answer to a request, not only the previously entered information but also new information which has not been explicitly introduced into the system. Before such systems can be made operational, however, complex problems must be solved, many of which are of a general scientific consequence. The IR systems under review will form part of more elaborate logical information systems.

An information retrieval language is a major component of an IR system. It is a specialized artificial language designed to express the subject-matter of documents and information requests, in order to locate in an information collection those documents which provide answers to certain questions. Every IR language must meet certain requirements. First of all, each notion in it must be expressed by one and only one word (a certain sequence of symbols) and, vice versa, each word must express a single notion. This requirement can be referred to as uniqueness of the vocabulary. Naturally, the IR language vocabulary should not be biassed that is, should not reflect the attitudes of the information generator and user to the given notion. Another requirement, that of a formalized grammar, stipulates that any statement formulated in terms of an IR language must allow a single interpretation. Only when these requirements are satisfied will it be possible to formally match document search patterns and search formulations. It is self-evident that natural languages, in which the same word may have different meanings (homonymy) and the same notion may be expressed in different words (synonymy), where the meaning of a sentence is modified by grammatical means or by context, are unsuitable for information retrieval purposes.

That is why specialized IR languages have long been a standard feature of information retrieval. Conventional kinds of such languages include library and bibliographical classifications (such as those used in a classified catalogue) and alphabetical subject classifications (such as those used in an alphabetical subject catalogue), which we have come across while discussing secondary documents. A feature of these conventional information languages is that all existing knowledge has been distributed in them among the words and phrases in conformity with possible information requests.

However, such a structure is a source of a certain weakness on the part of the conventional IR languages, as the document contents and user requests depend on the progress of science and technology, on the varying interests and personalities of specialists, i.e. to all intents and purpose, they escape accurate accounting and prediction. The setting up of an IR system, the indexing of documents and their retrieval, are operations separated in time, sometimes by very large intervals. For this reason a good portion of information contained in documents that were indexed by means of a conventional IR language will be lost in searching . To obviate this disadvantage, new types of IR languages are created in which the words are not grouped in previously established statements (search patterns). This process takes place either during the indexing stage or even during retrieval. Particular IR languages will be the subject of a large part of the two chapters that follow.

The second component of an IR system, as we have seen, is the matching criterion. A matching criterion is a set of rules according to which the degree of semantic similarity between a document search pattern and a search request formulation is established in an IR system and the decision is taken as to whether to retrieve or not to retrieve a given document in response to a query. Special attention should be paid to the notion of semantic similarity contained in this definition; it has been termed relevance . At the present time this semantic similarity (relevance) cannot be yet established formally and hence automatically. Therefore the relevance judgments, normally made by the user himself, will be necessarily subjective.

The situation is further complicated by the fact that requests are not adequate representations of the specialists' information needs, which are dynamic in character and will continually change with the receipt of new information. When formulating a search request a specialist is not always aware of this actual information need. Getting to know he contents of the documents retrieved may change his idea of the actual need, which will immediately affect his request formulations. Thus, a sort of user-system dialogue takes place, in which the operation of the whole system is optimized through a succession of iterations.

Something like this is observed in everyday life. When a person applies to a consultant, the latter will not supply an answer before he identifies, by posing a number of questions, the actual information need of the applicant. The latter may sometimes gain full realization of his needs only following a talk with the consultant. A strong side of the conventional IR systems, e.g. catalogues of libraries or bibliographic indexes, is that the feed-back to the user is a built-in feature of their structure. Automated document retrieval systems should also be designed in such a way as to enable the searcher to investigate the search file, modifying his search request formulations in conformity with intermediate retrieval results. In other words, they must be designed as systems of the "man-machine" class.

Another way of overcoming this difficulty is the perfection of IR languages and matching criteria. For instance, the matching criterion in a Soviet IR system, named "Pusto-Nepusto" (or "Empty-Non-Empty ") , provides retrieval of documents not only when a match of a term or terms occurs in a document search pattern and a search request; if desired by the user, the system can retrieve a document if the terms in its search pattern occupy a higher or lower position in a logical hierarchy than at least one term in a request formulation. Thus, in response to a query concerning documents of which the subject is the properties of liquids, the system can provide documents dealing with the boiling point of water, or vice versa.

Efficiency of an IR system

It has been noted before that a document search pattern recognizes only the main subject treated in the document. Therefore, such a method of document identification cannot ensure the complete recall of all pertinent information in each particular case. On the other hand, part of the retrieved documents may not contain pertinent information at all. These documents constitute the so-called false drops, or noise, and reduce the precision of information searching. The incompleteness and imprecision in information retrieval are a kind of price that has to be paid for facilitation of the search procedure. The price is lower with the more advanced IR languages and matching criteria and, consequently, with the more elaborate searching strategies.

This reasoning brings us to an understanding of the methods of evaluation of document retrieval systems. Operational efficiency of an IR system (economic efficiency will not be discussed here) can be defined as the measure of the system's ability to discharge those functions for which it has been designed. The function of a document retrieval system is to pull out, in answer to a request, relevant documents from an information file, i.e. documents that are semantically related to the query. As a rule, in experiments comparing the efficiency of several IR systems, the documents are assessed for relevance by a team of judges. In practical operation of IR systems, the relevance judgments are made by the user himself, who assesses a document content in terms of his actual information needs. It follows that all evaluation studies are formal in nature and their significance lies mainly with comparisons of different types of IR systems.

The most widely used measures of IR system effectiveness are recall and precision. Recall may be defined as the ratio of the number of relevant documents retrieved to the total number of relevant documents in the information file. Precision refers to the ratio of the number of relevant documents retrieved to the total number of (both relevant and non-relevant) documents retrieved. Both measures are usually expressed as percentages. The upper part of Fig. 17 illustrates these definitions by a contingency table of relevance and recall.

As demonstrated in the experiments of the British documentalist C. Cleverdon, popularly known as the Cranfield Project, there is an inverse relationship, although not strictly formal, between recall and precision values: an increase in precision will reduce recall. This relationship can be traced in the diagram also contained in Fig. 17. Therefore, increasing the complexity of an IR language and matching criterion, and thereby the searching procedure generally, with the aim of reducing noise, will result in the loss of many relevant documents. If a 100% precision is sought, i.e. only relevant documents are wanted in the search output, the recall will approximate zero. And, conversely if the aim is a 100% recall, i.e. retrieval of all relevant documents in the collection, the precision will approximate zero on account of the increased noise.

The choice of the recall-precision ratio is of major importance in developing specific IR systems. Most systems have a recall of 70% to 90%, and precision 8% to 20%. These values can be held to be satisfactory for meeting information requests of scientists and engineers, because with a certain information redundancy of the contents of primary scientific documents, the loss of a small proportion of relevant documents will not be felt, and the possibility of iterative searching will help to overcome the adverse effects of noise. When the recall reaches 90%, its further increase will be attended by a sharp drop in precision. It would not be an exaggeration to say that the provision in the search output of the last 10% of relevant documents will demand a higher price than all the previous output. The data cited indicate that there are limits to the efficiency of any IR system, and that both information losses and noise have to be accepted in their operation.

General framework of an IR system

Although the notions of information retrieval and an information retrieval system had their origins several decades ago, the corresponding processes and facilities have been in use in scientific work for hundreds, even thousands of years. It was the achievement of informatics that the operations of a multitude of empirically evolved facilities for information retrieval (reference books, catalogues, bibliographical indexes, libraries etc.) could be fitted into a common framework. The general layout of an IR system is shown in Fig. 18.

We shall analyse its structure in terms of a document retrieval system. You will recall that such a system is designed to retrieve scientific documents in answer to the information requests of the users. Therefore, when an IR system is set up, it should be provided with an initial input of documents, and newly issued documents should be added to it in the future. This process is shown symbolically in the left-hand part of the scheme. The documents come to the Input Converter (IC) where they are subjected to a number of operations. Each document undergoes an analytico-synthetic treatment yielding its search pattern. The search pattern contains a brief formal description of a given document which distinguishes it from all other documents (bibliographical entry), as well as a concise statement of its subject matter (notation, terms, annotation, abstract).

In addition, each document is assigned an address (usually a number), by which it can be located later. The documents themselves can be converted into a different form (miniaturized or machine-readable). This completes the conversion of the documents at input into the system, and the documents (in the original or converted form) with their storage addresses proceed to the Passive Store (PS). They are kept there until requested by the information users.

The search patterns recorded on some material medium (catalogue cards, punched cards, microfilm, magnetic tape) together with the relevant document addressed, are seperated from the documents and forwarded into the Active Store (AS), There they are arranged in an order facilitating their matching against every incoming request.

The input of information requests into the system is symbolically shown in the right-hand portion of the scheme. First of all, they go into the Input Device (ID) where they are converted from the searcher's natural language into the information retrieval language used in this system . After conversion they emerge as search request formulations, suitable for matching against the search pattern stated in the same information retrieval language. The request formulation proceeds into the Resolver (R) through which are passed the search patterns of the documents from the AS (that is why it is called "active"). There, in the Resolver , the search request formulation is matched against the search patterns. In case of a match (complete or partial, depending on the matching criterion applied), the R gives an instruction to the PS to retrieve the corresponding document by its address, Between the PS and the user is usually an Output Device (OD) which provides for the return of the documents into the PS, and also converts into the original form the documents which have been stored as microreproductions or as machine-readable records. The furnishing of the wanted documents or of their hard copies completes the operation cycle which is started by a user request. Such are in very broad outline the structure and operations of any information retrieval system.

Examples of specific IR systems operation

This can be illustrated using the already familiar example of library operation. The Input Converter (IC) in a library is the cataloguing (processing) department. It is there that a record is made on catalogue cards of the search pattern of a document (bibliographical entry, classification numbers, subject headings) and its address (call number) which is also inscribed on special labels pasted on the documents (books or journals). The PS is the stacks where the books are shelved according to their call numbers, and the AS is the catalogues, where they are searched. The bibliographers in the bibliographic reference department and the catalogue-room attendants serve as the Input Device (ID), and the reading-room or counter attendants as the Output Device (OD). The Resolver (R) in this case is simulated by the intellectual efforts of the reader who scans the cards in the library catalogue, deducing from the entries which books may be of use to him for solving the problems of his concern.

This example can help to spot some faults of the conventional library as an information retrieval system, in addition to its general shortcomings which were discussed in the previous chapter. They consist primarily in the principles of organization of the Passive and Active Stores.

The PS in the library takes the form of tier stacks, where the document originals are arranged on the shelves. These documents occupy a great amount of space; their issue, shelving arrangement and rearrangement are very labour-consuming and do not lend themselves readily to mechanization. In consequence, the greatest difficulties in the work of modern libraries are those presented by the lack of suitable storage space for the stacks. Besides, the lending of literature can lead to situations where requests for literature in constant demand cannot be met because the books sought are not in the stacks, having been issued to other readers. These difficulties can be overcome only if the documents in the PS are stored as microreproductions or machine-readable records and the readers are issued only true-size hard copies of these documents.

The active storage, realized in the library in the shape of card catalogues, also suffers from serious disadvantages. Because the cards in a catalogue are filed in a linear sequence, any one catalogue can interpret a book collection in one aspect only. We have learned already that the librarians limit themselves to three such aspects: author, subject and classified. Within an aspect a separate card is made for each characteristic for which a search may be conducted.

Thus, increasing the number of aspects and characteristics by which a reader can search his literature will result in more complicated and cumbersome systems of catalogues. This, in turn, increases the labour costs of catalogue maintenance and hampers their use by the readers. For this reason the librarians are obliged to restrict the number of catalogues in libraries and the amount of duplication of bibliographical entries. This, of course, detracts from the search capabilities of the library as an information retrieval system.

This disadvantage of the library catalogues as active stores springs from their inherent organization, which requires that each document must be provided with a search pattern on a seperate information carrier – a catalogue card. A card duplicate with the given search pattern must be reserved for each search characteristic. A card duplication ratio of 1.5 for each catalogue has come to be accepted in library practice. This means that within each aspect a document will have, on the average, one or two search characteristics. But even with this restriction, a library reflecting its stock in three catalogues will have to keep four or five cards per document.

Such an organization of AS is called serial. It presents another inconvenience, namely, that separate searches are necessary for each aspect or characteristic and [illegible]-aspect searching is difficult. Most mechanized and automated [illegible] systems use an inverted organization of the AS. Essentially it means that a separate information carrier (e.g. a card) is assigned a certain search characteristic or aspect. The addresses of all documents whose search pattern contains this characteristic or aspect are recorded on this carrier.

The inverted scheme has several advantages in comparison with the serial one: it is more compact, it permits multi-aspect searching, and by the feed-back to the user it permits adjusting the search strategy. Adding or deleting certain search characteristics in the request formulation will reduce or expand the search output. These problems will be treated in more detail later on, while discussing descriptor IR systems and technical means of their implementation.

It should be mentioned in conclusion that one important advantage of the conventional library IR systems over mechanized and automated systems is that in the former the functions of the Resolver are performed by the reader himself, who during his searches in catalogues can modify his requests and the intuitively applied matching criterion. This to some extent compensates for the shortcomings of the organization of the AS and makes for higher efficiency of the library IR system as a whole.

In this chapter we have considered the essentials of information retrieval, defined the main concepts of this part of informatics, discussed the methods of assessment of IR system efficiency, the schematic diagram of an IR system and the interactions of its component parts, and the advantages and shortcomings of the different methods of organization of active storage. All this information forms the necessary background for a deeper understanding of the structure and principles of utilization of the existing conventional and automatic IR systems, which will be dealt with in the next chapter.

Questions for self-checking

1. What is information retrieval and what are its basic principles?
2. What component parts constitute an information retrieval system and what are their functions?
3. What is the gist of the relevance problem?
4. How is the efficiency of an IR system determined?
5. What is the purpose and structure of the schematic diagram of an IR system and how do its components interact?
6. Describe the operations of a library and the use of an abstract journal in terms of IR systems?
7. What is the advantage of the inverted organization of AS over the serial one?

Literature

1. Fairthorne,R.A. Towards Information Retrieval. London, Butterworth, 1961, XXIII, 211 p.
2. Kent, A. Textbook on Mechanized Information Retrieval. New York, Wiley, 1962.
3. Sharp, J.R. Some Fundamentals of Information Retrieval. London , Deutsch, 1965, 224 p.
4. Vickery, B.C. On Retrieval System Theory. 2nd ed. London, Butterworth, 1965, XII, 191 p.

6. CONVENTIONAL INFORMATION RETRIEVAL SYSTEMS

Information and bibliographic publications and catalogues of libraries and bibliographical indexes, which were the subject of our third chapter, may serve as particular examples of conventional IR systems or their separate parts.

We revert to them in this chapter in order to consider the principles of their design and operation in terms of information retrieval.

We know already that the basic types of conventional document retrieval systems are the author, alphabetical subject and classified systems. Since in conventional IR systems the matching criterion is not explicitly stated (it is applied by the searcher intuitively), their sole component is the information retrieval language. In the author systems the IR language is formed by the rules for bibliographical description and alphabetization; in the alphabetical subject systems, by lists of subject headings and methods of subject indexing; and in the classified systems, by hierarchical or faceted classifications and classification methods. These conventional-type IR languages will be the subject of this chapter.

Author systems

Author systems originated many centuries ago as bio-bibliographical lists and were designed mainly for the identification, first of different manuscripts, and subsequently of different editions of the same work. The turning-point in their development and use was the mid-19th century when they enjoyed widespread acceptance as author catalogues in large public libraries. It was then that the theory of cataloguing and author catalogue arrangement began to develop giving rise, by the 20th century, to the various national cataloguing codes, and in our day, to the internationally approved principles of descriptive cataloguing.

The main idea behind the rules for bibliographical description is to isolate in every type of work a bibliographic particular (e.g. name(s) of the author(s), title, sponsoring institution) which would best characterize it. These basic data, extracted from a publication according to certain rules and recorded in a symbolic form, constitute what is called an index entry; the totality of these data makes up the vocabulary of an IR language of an author system. The other data, also extracted and recorded according to special rules, constitute the text of the entry and form an integral part of the search pattern of any document in an IR system. The entries are arranged alphabetically by headings in conformity with strictly defined rules which serve as the grammar of the given IR language.

The need for elaborate rules for bibliographical description comes from the IR system requirements which we formulated in the previous chapter, namely, the uniqueness of the vocabulary and the formality of the grammar. But bibliographical data are, unfortunately, far from unique and their manner of presentation varies in different publications. For this reason, descriptive cataloguing methods rigidly regulate the sources, kinds and elements of an entry, as well as its language and spelling. In principle, the source of description is the document as a whole, but primarily its title- or front-page which is characterized in the language of the original but in modern spelling. Abbreviations of certain words, prescribed by a special list, are used in the text of the entry.

The bibliographical entry contains the following particulars: (1) Author(s), or the heading. (2) Title. (3) Sub-heading data - the data explaining and qualifying the theme and content of the document, its nature and purpose; data concerning the approval or endorsement of official documents; data on the persons participating in the compilation of the document or its preparation for publication; data on the edition or the official character of the publication. (4) Imprint - place of publication, publishers, and year of publication. (5) Collation - number of pages and the presence of illustrations. (6) Information before the title - the name of the institution sponsoring the publication; the title of the series in which the publication appears, and its serial number. (7) Bibliographical notes - data on the authors given in the publication (if the notes disagree with the data accepted for the entry, or if the publication has more than two authors); data on the variant readings of the title, on the availability of a bibliography in the document, on publication defects. Fig. 19 gives a break-down of bibliographical data on the title-page of a publication and in its entry.

Different approaches to descriptive cataloguing arise in connection with the authorship, number of volumes, frequency of their appearance and the polygraphic independence of a publication. Authorship is the most individualized attribute for some of the more widespread publication types. Works by one, two and three authors are entered under the surname and forename (initials) of the first author. An added entry is made for the second and third co-author (or they are listed in the alphabetical index of names), so as to reflect a given work in the common alphabetical sequence of their works. The names of these authors are presented in the headings in a unified form, in order to ensure the unique location of the works of the same author in the common alphabetical listing.

The authors are called compilers in such publications as dictionaries, guides, handbooks, study aids and guidance manuals, reading books, anthologies, technical and scientific information material, advanced methods exchange material and fundamental bibliographical indexes. All these types of publications are indexed under their compilers, who have done a piece of independant research.

Works by four or more authors, collections of papers of different authors, syllabuses, procedure manuals, statistical handbooks and similar publications are entered under the title, even where the compiler or editor are indicated. (In some countries, e.g. in English-speaking countries, such publications are indexed under compiler or editor). Added entries are made for the first author, if a work of four or more authors is indexed, and for the compiler or editor, if a collection is indexed, which is a great help to those users who will remember a book by such names rather than by its title.

Official publications are issued on behalf of institutions, organizations or agencies as their documents for which they bear full responsibility and which are directly related to their activities. These include statutes, regulations, manifestos, appeals, reports, plans, decrees, resolutions, instructions and other similar material. Their titles usually begin with these standard words and contain the name of an institution or organization, e.g., "The Charter of the Communist Party of the Soviet Union". The salient feature of an official publication, which determines information requests for it, is the name of the issuing body. For this reason official publications are commonly indexed under the name of the institution, which in this case appears as a corporate author (not to be confused with joint authors, viz., several persons who have collaborated in writing a work). The same indexing procedure is used for many non-official publications, such as "Transactions", "Proceedings" or "Papers" of scientific institutions, congresses, symposia etc., if their title begins with or is confined to these standard words, and contains as an integral part the name of the organization or institution; for example, "Transactions of the Krupskaya Institute of Culture in Leningrad".

According to the number of volumes (issues) and their frequency of appearance, publications are divided into multi-volume and series works, and periodicals. The typographic make-up of a multi-volume publication or a series is such that much bibliographical data recurs in each of its volumes or issues. For such publications a series entry is made, with the general part giving all the recurring details, and the specification listing only those details which distinguish one volume from another. This saves a lot of entry space and indexing time. It is important though that not only the similarities but also the differences between multi-volume books and series should be realized. A multi-volume publication is limited by its content to a definite number of volumes, while a series comprises quite independent publications united only by their common nature or by the similarity of their subject matter. Therefore, the series entry for a multi-volume book is its main — and normally its only — entry, giving full bibliographic details. For series publications, on the contrary, main entries will be the entries for individual issues, which ought always to be made. Here the series entry is an auxiliary one; it is very brief and made only in those cases when a series as a whole can be of scientific interest.

Periodical publications, which include serials, journals and newspapers, are also indexed by means of a series entry usually under the title. The only exception to the general rule is constituted by the serials which occupy an intermediate position between the book series and the journals. Occasionally, they are entered under a corporate author with, in addition, separate entries made for the individual volumes or issues which may have an author or be collections of works having a title of their own.

To index a part of a publication which is in itself an independent work (e.g. a paper in a collection or journal), an analytical entry is made under the author and title, as with other independent publications. In place of the imprint however details of the publication in which the paper is included are indicated. These details may vary according to whether the publication in question is a collection, serial, journal or newspaper. An entirely different matter is making an analytical entry for a part of a work, such as a chapter or section of a book, which cannot be viewed autonomously out of the context of the publication to which it belongs. Therefore, chapters or sections are entered under the author and title of the relevant publication. But instead of the collation such an analytical entry gives the number and title of the chapter or section indexed, as well as the number of its pages in the book.

The same procedure applies to reviews, or critical comments on books, which usually appear in journals, serials or collections of papers. What the reader wants to know is primarily whether there is a review for this or that book (often he will not know who wrote this review), what title it has and where it was published. For this reason reviews are entered under the author and title of the book reviewed. The entry lists all the particulars which are usual in book entries, but instead of collation data it gives details of the review and of the publication in which it appears. The most important reviews may also be indexed as separate papers. The above rules for description of the main types of publications are schematically displayed in Fig. 20, with illustrative examples shown in Fig. 21.

Rules for alphabetical filing of entries, which form the grammar of this conventional IR language, specify the various methods of such filing within one and within several natural languages. There are two commonly used methods of alphabetization of words, names or sentences, which are both based on the order of letters in the first words. By the first method ("word-by-word"), entries are filed in the sequence of words, by the second ("letter-by-letter"), strictly in the sequence of letters of the alphabet, without consideration for the division of sentences into words. These methods can be explained by the following:

"Word-by-word"	"Letter-by-letter"
New England	Newark
New wives for old	New England
Newark	Newman
Newman	New wives for old

Most alphabetical systems use the "word-by-word" method as better common-sense and less formal. It applies not only to prepositions and other function words (articles, particles), but also to complex and compound words, including forenames and surnames. In some systems this rule is made still more complicated: if the first words coincide, it is not the alphabetic sequence of the words that follow which is taken into consideration but the heading category -- entries are listed first under the names of the authors, then under the names of institutions, and finally, under the titles beginning with the same word. Since this facilitates locating entries in long alphabetical sequences, the practice of listing these three categories of headings in three different alphabetic sequences is sometimes adopted.

There is another range of problems in alphabetization of index entries connected with the fact that different languages use different alphabets and systems of writing. Most of the nationalities of the USSR, for instance, make use of Russian letters; many nations in Europe, America and Africa use Roman letters; some nations of Asia employ national systems of writing, e.g. Arabic, Ancient India (Sanskrit), Chinese hieroglyphics, etc. Moreover, different languages using the same letters may have different alphabets. Thus, the alphabet of the Azerbaijan language differs from that of the Russian language, and the English alphabet differs from the Polish one. Because of this, entries in languages which have different letters and alphabets have to be placed in different alphabetic sequences. If occasion demands a common alphabetic sequence for entries in different languages within the same system of letters, the basis is provided by some artificial alphabet that usually ignores auxiliary letters and diacritics.

In conclusion, it should be noted that until quite recently alphabetic systems have been the only means of identification of publications, of searching for specific titles in IR systems, and of searching for publications by specific personal and corporate authors. The excessive work and time-consuming efforts resulting from the existing cataloguing rules and their insufficient standardization at the international level should be mentioned as the principal faults of those systems. The recent proposal to use standard numbers for books and serials as their identifiers in information work, is an attempt to rectify those inadequacies. The main idea of the proposal is to assign to every new book and every serial title a unique number associated with the respective entry in a national bibliography. Implementation of this proposal will considerably facilitate the input of documents in automated IR systems and the exchange of bibliographical data both on a national and international scale.

Subject systems

We have already seen, as in the case of permuted-title and citation indexes, that a listing of some particulars of a bibliographic entry can, under certain conditions, provide a subject searching capability. This fact has long been utilized in subject (or alphabetical subject) systems. Subject catalogues and alphabetical subject indexes, which were discussed in the second chapter, are examples.

Subject cataloguing is a special kind of document classification which is concerned with the concise formulation of document contents by means of one or several standard words called subject headings. These are arranged alphabetically and followed by either bibliographical entries, as in subject catalogues, or by classification numbers, bibliographic entry numbers or book page numbers, as in alphabetical subject indexes.

The task of subject cataloguing is to identify the subject which is the main theme in a document, as well as to establish its main features and its relationships with other subjects. The different features of a subject are as follows: its history(origins and development), structure, composition, properties, state, purpose, interaction with other subjects, investigation, evaluation, etc. In subject cataloguing, documents are grouped by subjects not belonging exclusively to any one area of knowledge, which makes it possible to form different complexes ("subjects") such as a personality, human activity, geographical feature, natural phenomenon, social phenomenon, material, property, etc. Thus, the subject heading "X-rays" will bring together documents which are concerned with the nature of X-rays(physics), their use in medical diagnostics (medicine), and design of X-ray apparatus (X-ray engineering). As subject headings are listed in a common alphabetical sequence, it is not necessary - while searching for a document on a specific subject - to know the place of this subject in the general systematics of scientific knowledge. Subject headings of quite different content will be listed side by side in alphabetical subject systems, e.g.:

Apricots
Archangelsk
Aristotle
Automobiles.

On the other hand, alphabetical arrangement brings together words of the same root which often prove to be associated semantically; this will help to create in a subject system a broad thematic complex, e.g.:

Atom
Atomic bombs
Atomic electric power station
Atomic energy
Atomic ice-breakers
Atomic nucleus
Atomic spectra
Atomic thermal capacity.

A subject heading may be selected from a special previously established list or it may be formulated by the indexer; in either case it recognizes only the main aspects of a document's content, which in practice is usually covered by a very few headings. The major problems faced in building up the IR language of a subject system are the use of "generic" and "specific" headings, the inverted word order in the formulation of subject headings, and the use of sub-headings. Without going deeply into the methodology of subject classification, solutions for each of these problems can be illustrated by a few examples.

A subject heading must be most specific in defining the subject of a document. Thus, a document dealing with motor-car carburettors should be entered under "Carburettors" and not under "Internal combustion engines". In this case the subject heading is *specific*, or rather, adequate, i.e. corresponding to the main subject of the document. If, in anticipation of possible requests by users of the system, it is thought desirable to enter a given document under a heading expressing a more general idea (in our example, Internal combustion engines), such a heading is considered to be *generic* or, more correctly, generalizing.

In order to bring together related subject material in a subject system, the words in a heading can be regrouped (inversion). Thus, an adjective is often placed after a noun, e.g.:

Microbiology - agricultural
Microbiology - industrial
Microbiology - medical
Microbiology - veterinary.

In some cases, however , inversion is not necessary; for example, when an adjective defines an important attribute of a subject (Computational technology, Electronic valves), or in the names of historical events and other fixed phrases(Roman Empire). In order to divide the material within a subject heading, subheadings are used, e.g. topical(Water - Analysis, Evaporation, Purification, Chlorination), geographical (Animal Breeding - USSR, Great Britain, France, USA), form (Standards - Bibliography, Classification Schedules, Indexes), chronological (Periodicals - History - 17th Century, 18th Century, 19th Century). A special position is occupied by the standard or common subheadings, e.g. those used in the subject cataloguing of the literature on materials (Concrete — Analysis, Defects, Substitutes, Testing).

Subject systems are most effective in searching for narrowly defined subjects and are far less useful for broad thematic searches in a particular area of knowledge. This is their peculiarity, not to be viewed either as an advantage or disadvantage - it merely defines the conditions of their application. On the other hand, subject systems have a number of unquestionable advantages. They are simple in operation and require no preliminary training on the part of the user; introduction of new subject headings does not involve any changes in the existing headings and re-indexing of the documents already in the system. Growth in the number of subject headings and documents in a system does not create any special difficulties either for the indexers or for the users. Subject systems are easily implemented as either manual card catalogues and card files or as bibliographical lists.

But subject systems also suffer from a number of shortcomings. They retrieve only those documents whose main subjects match document requests, and they do not give details on other topics discussed in scientific documents. (This, however, is the common inadequacy of conventional IR systems oriented to manual handling). Another shortcoming which is equally common to all conventional IR systems is the difficulty of multi-aspect searches, since a subject system does not permit simultaneous retrieval of information on several documents entered under different headings, that is to say, it is impossible to conduct a search using any combination of characteristics, or their logical product. The retrieval languages of subject systems employ the vocabulary of a specific natural language and are therefore unsuitable for international use. Finally, in comparison with other conventional IR systems, the setting up and maintenance of a subject system is a very laborious task for highly skilled personnel.

Hierarchical classifications

Classification is the grouping of subjects or relations in classes according to a common characteristic which is inherent in all the subjects of a given kind and distinguishes them from the subjects of the other kinds, the grouping being done in such a way as to put each class in the system in a definite, fixed position with respect to the other classes. The characteristic according to which classification is done is called a principle of division.

The classification process obeys some formal rules of logic: subjects or relations may be divided according to only one principle of division at a time, and the resulting classes and subclasses should be mutually exclusive, the division into classes should be balanced and continuous, without leaps. Classifications, in which every subclass has only one class that immediately precedes it(relations of strong hierarchy) and all subclasses subordinated to only one more general class (collateral subordination relations) are called hierarchical. By way of visual demonstration of a hierarchical classification, Fig. 22 shows its graph, that is a diagram consisting of points (apices) and the lines (edges) that connect them, as well as an Euler-Venn diagram indicating the relationships between the classification divisions.

Classification of documentary information, which is one of the most important types of an IR language in conventional systems, consists in the grouping of scientific documents by areas of knowledge according to their content. Such document classifications are more or less associated with classifications of sciences, as their main classes will broadly correspond to specific fields of knowledge, and further division of these classes will correspond to the structure of these fields of knowledge.

However, these classifications, which have been called library classifications, are not the same as classifications of sciences. Library classifications must be built according to some formal rules of logic, because only under this condition can they serve as IR languages and provide unique identifications to the documents. But rules of logic are inapplicable to classification of sciences, since there are no clear-cut boundaries between different sciences. The other distinctions also spring from the strictly practical nature of library classifications and the peculiarities of the subjects classified. Apart from the division by document content, a library classification must include division by types of publications (books, periodicals, special types of technical publications), by their purpose (scientific, popular science, instructional), by language, etc.

Library classification schemes are generally published as schedules consisting of two parts: the main tables and the tables of auxiliaries or common subdivisions. Main tables list all fields of knowledge and their sections in a logical order with each level of division according to only one characteristic; each heading may have a number of subordinate headings forming a certain hierarchy. Because of these characteristic features, conventional library schemes are sometimes called linear-hierarchical. Tables of auxiliaries or listings of common subdivisions display the recurring characteristics of different subjects, e.g. the characteristic of place and time to which a document content refers, or that of type of publication.

Each division of a library classification is assigned a conventional symbol called a code number. Classification code numbers constitute the notation which may be numerical, alphabetical, or mixed. The merits of a numerical notation are based on the fact that a numerical sequence is more obvious and familiar than an alphabetic one, that any combinations of digits are easily pronounceable as numbers or figures, and that the Arabic numerals are understood by all nations whatever their spoken or written language.

At the same time numerical notations have a serious drawback, namely their limited base: because there are only ten digits (from 0 to 9), very lengthy numbers with many digits have to be formed in order to designate complex or very specific notions. A solution is frequently sought in the use of mixed code "numbers" which include letters as well as numerals.

From the standpoint of structure, notations may be non-structural and structural or hierarchical. Non-structural notations make use of the serial numbers of classification subdivisions in a common numerical sequence. Such notations have little mnemonic value, do not reflect the hierarchy of classification subclasses, and make further division of these subclasses difficult. For this reason, most of the modern library classifications employ structural notations which may be either numerical, alphabetic or mixed. Structural notations represent the conceptual structure of a classification, since each class is designated by one symbol, and all primary divisions of this class by two symbols, of which the first stands for the class, and the second for the corresponding subclass. Structural notation permits the detailing of a classification scheme to any desired length or depth.

Universal Decimal Classification

Library classifications have been with us from very early times, and this is hardly surprising, as the need for them arose simultaneously with the emergence of written records. Nowadays, hundreds of different classifications are in use.

The Universal Decimal Classification is certainly the most significant of the classification schemes developed around the turn of this century, and also the most extensively used in all countries of the world. It owes its origin to the work of two eminent Belgian bibliographers, Paul Otlet (1868-1944) and Henri LaFontaine (1854-1943). The international conference of bibliographers , which was convened in 1895 through their efforts, formed two organizations: Bureau International de Bibliographie and Institut International de Bibliographie which were later to unite under the second name (IIB). The Institut International de Bibliographie which had its headquarters in Brussels was entrusted with the task of compiling a Universal Bibliographical Repertory, i.e. a bibliography that was to cover the literature in all fields of knowledge published in all countries in all languages from the earliest times to our day. This obviously unfeasible task was never realized but required for its fulfilment the provision of a depth classification scheme encompassing all parts of knowledge, i.e. a comprehensive or universal classification, and usable on an international scale.

The basis for this new classification scheme, developed by a group of experts headed by P. Otlet, was furnished by the Decimal Classification of Melvil Dewey, which had been considerably revised and extended. The new version of the decimal classification has appeared in French in separate issues since 1897: in 1905 it was issued as a complete publication, and in 1907 reprinted under the title "Manuel du répertoire bibliographique universel".(1) Subsequent editions of this version and other versions in different countries and different languages, have been sponsored by the International Federation for Documentation, successor to the Institut International de Bibliographie,and are known as the Universal Decimal Classification or UDC.

The UDC comprises main tables, auxiliary tables and an alphabetical subject index. The main tables display the numbers by which the documents are systematized according to their contents; each concept presented in the main tables must have a definite UDC number. The sum total of human knowledge is divided into 10 main divisions, shown in Figs. 23 and 24. Each successive digit added to the designation of a main division does not change its general content but merely qualifies it.Such a method of building UDC numbers makes it possible to divide any general idea into narrower specialized topics. The greater the depth of division, the longer will be the notation. Thus, 'compressive strength of soils' in construction engineering is denoted by 624.131.439.4. To facilitate the reading of the notation, every three digits are followed by a full-point. The numbers are pronounced as a sequence of integers, e.g. six-two-four (point) one-three-one (point) etc. Fig. 25 gives an example of interpretation of a UDC number, 621.22 Water power. Hydraulic machines.

---

1. Manuel du répertoire bibliographique universel. Bruxelles, 1907.

UDC numbers are filed, within each division, according to the principle 'from the general to the specific'.Their order depends only on the sequence of figures in each division and not on the length of a number.

In the main tables are listed, apart from the main classification numbers, special (analytical) subdivisions which represent the characteristics typical of a limited range of concepts; they are limited to use in the section under which they are listed. If applicable in many subdivisions of a given section they are joined to the main numbers by a hyphen; if applicable only in the subdivision concerned, they are joined by means of .0 (point 0).

The auxiliary tables list the general or common auxiliaries, which represent recurring characteristics according to which documents can be divided in all areas of knowledge and which are applicable in all the sections of UDC, as well as special signs which serve to join several main UDC numbers. The common auxiliaries may be of language, of form, of place, of time, of race and nationality, of viewpoint. They are used in all sections of the scheme with the same meaning. When occasion demands, letters of the Roman or Cyrillic alphabets, separate words or names can be attached to UDC numbers;e.g. 92 Einstein, for a biography of Einstein.

For a more exhaustive and accurate representation of the subject-matter of documents, the UDC permits - in addition to the main numbers and auxiliaries - the connection of several numbers with the help of different symbols. The common auxiliaries and the connective symbols are shown in Fig. 24. The addition (plus) sign is used in those cases when a document deals with several separate concepts of equal value. The extension (stroke) sign is used to join the first and last of a series of consecutive numbers denoting adjacent related concepts. The relation (colon) sign permits the classifying of documents dealing with concepts that are conceptually related. The synthesis (apostrophe) sign is used, inter alia, for documents on chemical compounds and alloys.

The task of classifying by UDC is made much easier by the availability of an alphabetical subject index to the tables. Having established the primary subject of a document one must look up in the index the UDC division or subdivision that corresponds to the document content, then scan the array of the relevant section, select the most appropriate heading and write down its number. In establishing the heading it is not necessary to follow the last level of division in the tables used. One should be guided here by the availability of material on this subject and the prospects of its growth. The use of too general numbers is undesirable, however, because the depth of classificatory analysis of a document will then be sacrificed.

In classifying documents in the mathematical natural and applied sciences (UDC divisions 5 and 6), the most important action is to delimit their topics. It should be remembered that the natural sciences study the laws of nature, while the applied sciences use their findings for practical purposes. X-rays, for instance are studied in physics, and the different aspects of their application in technology and medicine: hence, documents describing direct observations or results of laboratory studies are entered under 537.531 X-ray and Gamma-ray physics (non-corpuscular rays in discharges); documents on the design and manufacture of X-ray apparatus under 621.386 X-ray tubes and accessories (electrotechnology) and material on the diagnostic application of these rays under a subdivision of medicine, 616-073.7. The use of other subdivisions is also possible here: for example, if a document deals with the subject of films for X-ray photography it should be indexed by 771.531.34 in Photography and cinematography.

The UDC has the following specific features: (1) coverage of all fields of knowledge (the recent UDC editions contain more than 100.000 subject headings); (2) the decimal principle of division, which permits unlimited division of subclasses without violating the basic structure of the scheme; (3) the use of exclusively numerical notation, which is relatively easy to memorize and equally comprehensible to specialists using different languages; (4) the availability of an elaborate system of auxiliaries; (5) the use of the principle of synthetic notation; (6) the possibility of classifying any number of documents to any level of division.

The primary distinctive features of the UDC are its elaborate system of auxiliaries and the synthetic notation. It was precisely because of these features that the emergence of the UDC in a way meant a breakthrough in library classification. The former 'enumerative' schemes with previously established headings and ready-made numbers were ousted by a new flexible classification in which the necessary headings are formed during the classification process either by combining numbers with auxiliaries or joining together two or more numbers. The existence of a well organized international system for keeping the UDC up-to-date is its unquestionable merit and places it in an advantageous position with respect to other present-day classifications. Naturally, the maintenance of such a system takes much time and its operation presents great difficulties and entails considerable expenditure, but it is only through such a system that the development and perfection of a classification is possible.

All the same, it should be noted that the division of the whole universe of human knowledge into a mere ten classes, and the order of these classes in the UDC, do not correspond to the present level of the development of science. Despite this and some other defects, the UDC has become widespread and found application in thousands of institutions in many countries of the world. Presently the size of the complete UDC schedules is approximately 10 volumes each of 300-500 pages.

In conclusion, it would not be out of place to list the principal merits of the UDC which are: universality, international usability, the decimal system of notation, the number-building principle and a smoothly functioning body for keeping it up-to-date. At the present stage of development of classification theory and practice, the UDC remains the only internationally accepted universal system capable of revealing the contents of reference collections in sufficient detail, ensuring speedy information retrieval, and making for closer international co-operation. These merits of the UDC point with certainty at the advantage of using it in scientific information agencies and in special libraries for classifying literature in the natural sciences and technology. The Council of Ministers of the USSR decreed on 11 November 1962 the compulsory use of the UDC by information centres and technical libraries throughout the country for classifying literature in the natural and technical sciences.

Decimal classification has been in use in the USSR since 1921. The public libraries still utilize the different versions of the library classification schedules which are based on a modification of UDC made by Soviet bibliographer L.N. Tropovsky. Work on a new Soviet scheme of classification for research libraries, based on the Marxist-Leninist classification of sciences, has been going on for many years; it has been appearing in separate issues since 1960 (1).

The new Soviet scheme makes wide use of the techniques which determined the success of the decimal classification: the decimal structure of notation for the subdivisions of each class, the synthetic principle, the minutely elaborated tables of common subdivisions. All the same, the mixed (numerical-letter) notation employing different punctuation marks seems rather complicated in comparison with the UDC. Below are the main divisions of the scheme:

А	Marxism-Leninism
Б	Natural sciences in general
В	Physico-mathematical sciences
Г	Chemical sciences
Д	Earth sciences (geodesical, geophysical, geological and geographical sciences)
Е	Biological sciences
Ж/О	Engineering and technological sciences
П	Agriculture and forestry. Agricultural sciences
Р	Health protection and medical sciences
С	Social sciences in general
Т	History. Historical sciences
У	Economics. Economical sciences
Ф	Communist and workers' parties. Socio-political organizations of labour
Х	Government and law Juridical sciences
Ч	Military science
Ц	Culture. Science. Education
Ш	Philological sciences. Literature
Щ	Art. Art criticism
Э	Religion. Atheism
Ю	Philosophical sciences. Psychology
Я	Literature of universal content.

Specific features of hierarchical classifications

Classification is one of the major tools of humanity in the process of cognition, one of the normal methods people use to define an object. V.I. Lenin wrote in this connection: "What is meant by giving a 'definition'? It means essentially to bring a given concept within a more comprehensive concept". (2) This is one of the strong sides of hierarchical classification as an information retrieval language.

---

1. Bibliotechno-bibliograficheskaya klassifikatsiya. Tablitsy dlya nauchnykh biblioteк. Vyp. I-25. M., 1960-1969. (Gos. b-ka im. V.I. Lenina). (Bibliothecal-Bibliographical Classification. Schedules for research libraries. Issues 1-25. Moscow, 1960-1969. (Lenin State Library of the USSR)).

2. V.I. Lenin. Collected works. Vol. 14. Moscow, Foreign Languages Publishing House, 1962, p. 146

However, it should be recognized that hierarchical classifications, like any other IR languages, have certain limitations. It will be recalled that the development of science shows two conflicting tendencies, namely differentiation and integration, or the division of scientific fields into ever new trends and disciplines, on the one hand, and the interpenetration of the related and even remote fields and disciplines, on the other. The process of science differentiation is more or less adequately recognized in the linear hierarchical classifications, but the integration and interpenetration of sciences cannot be adequately reflected in these classifications. For example, it is an extremely complex task to find in the UDC a place for the scientific trends and topics which emerge on the borderlines of chemistry, geology and biology, or of mathematics and linguistics.

This limitation of hierarchical classifications, which strictly delimits separate sciences in accordance with formal rules of logic, runs counter to the synthesizing trend in the progress of science. The same rules will not permit multi-aspect indexing of documents by means of an hierarchical classification, and information searching for any combination of characteristics. It would be wrong to assert categorically that this constitutes a fault of hierarchical classifications; this is their inherent quality, which renders them higly effective for broad thematic searching when conducted in the conditions of traditional realization in the form of card catalogues. Like other IR languages, hierarchical linear classifications have a restricted range of applications.

Moreover, they are undergoing certain modifications intended to overcome these limitations. Already the UDC offers some facilities for synthesizing notation, which to a certain degree permits recognition of different aspects of a subject classified. Thus, using the relation (colon) sign one can index the multidimensional subject "Bibliography of atomic physics" as 016:539.1 where 016 stands for Special Subject Bibliographies and 539.1 for Atomic Physics. The synthetic notation principle has a very significant role to play in the development of classification.

In this respect, one of the leading figures in classification theory, the British information scientist D.J. Foskett, has this to say: "There is nothing to be gained by discharging hierarchical classification altogether, provided that we recognize that classification, in the modern sense, can and should mean more than this; that it can cope with the most detailed forms of subject analysis.....It is evident that a scheme of terms, whether systematic or alphabetical, however well equipped with cross-references, cannot hope to predict all the contexts in which every term may appear at some time or another. It can, however, provide a set of roles - operating procedures by means of which such contexts may be freshly created out of the scheme as occasion demands. This means that a classification scheme should no longer set out to provide a 'place' for every document, in the sense that a term, or set of terms, will be found in the actual schedules of the scheme for every subject that may be found in a document. In a modern scheme, the art of the classifier in a library is to construct a symbol that is, in effect, a translation of a subject".(1)

---

1. Foskett, D.J. Some fundamental aspects of classification as a tool in informatics. In: On Theoretical Problems of Informatics (FID 435), Moscow, 1969, p. 65

Faceted classifications

The first step in this direction was made by the UDC: the introduction of the common subdivisions and special auxiliaries and the connective symbols, in particular the colon, has appreciably enlarged its capacity for building numbers for complex and multi-aspect subjects. But it is to the eminent Indian library scientist S.R. Ranganathan that we owe a fundamental and consistent approach to the solution of this problem. The Colon Classification which he developed in 1933 was a further elaboration of the synthetic principle in classification.

S.R. Ranganathan opposed the practice of building minute 'enumerative' classification tables, in which the compilers sought to provide a separate number for every subject and concept. Instead of a single order to divisions in every main class, he worked out tables each based on one characteristic or aspect, later called 'facets'. The classification number for any document is built from the symbols used in every table, connected by means of the ' colon '. Hence the name - Colon Classification. This classification scheme has not found wide application, but the idea of facet analysis, which forms its most important premise, gave a powerful spur to the development of the library classifications called faceted or analytico-synthetic.

Facet analysis is essentially as follows: first, a field of science or technology is thorougly analysed, a faceted classification scheme is built, and a collection of documents in the relevant field is studied. The analysis yields a list of main facets, while the documents provide the essential terms in the subject field, grouped by the appropriate facets. Each term in a facet is called a focus. A special significance is attached to the order of the foci within a facet, and the facets within a classification scheme. This order is called the facet formula. S.R. Ranganathan suggested a facet formula that includes five categories:

Personality (class, subclass, subject)
Matter (material)
Energy (operation, process, action)
Space (place, territory)
Time.

A number of Indian classification experts and a group of British scientists are presently engaged in the further improvement of this formula, increasing the number of the categories to be included. The indexing procedure in a faceted classification begins with the formulation of the main subject of a document. It is expressed by a chain of foci which are taken from the facets and arranged in a fixed order. Often instead of the foci their numbers will be used. Such a procedure enables classes to be formed for those documents whose subjects are expressed by a combination of characteristics viewed from different aspects. To illustrate this rather complex procedure, we shall present an example from the field of medicine (denoted in Ranganathan's system by the letter L):

Facet O

Organs of human body

1 Organism as a whole
2 Digestive system
23 Esophagus
24 Stomach
25 Intestine
3 Circulatory system
4 Respiratory system
45 Lung

Facet P

Problems of medicine

1 Preliminaries
2 Morphology
3 Physiology
4 Disease
42 Infection
421 Tuberculosis

Facet H

Care and Treatment

1 Nursing
2 Etiology
3 Symptom and diagnosis
4 Pathology

In classifying documents dealing with the "diagnosis of infectious diseases of the intestine", they will be indexed L 25:42:3, and the"pathology of lung tuberculosis" L 45:421:4.

Compared with hierarchical classifications of the enumerative type, faceted classifications offer a number of attractive features: they greatly facilitate the multi-aspect indexing of documents, bringing together in one place all aspects under which a subject or theme is discussed; they are more hospitable to new terms; they provide greater depth of indexing with shorter notation. However, even the faceted classifications will not ensure searching by any combination of characteristics, as this would necessitate the classified catalogue providing places for all possible groupings and regroupings of the foci taken from the different facets - with card catalogues it is as complicated as in hierarchical classification. Furthermore, the present level of development of faceted schemes permits their effective use only in very specialized document collections.

We have discussed the principal types of conventional IR systems: author, subject and classified. With knowledge of the rules for bibliographical description and the filing of entries in an alphabetical sequence, the techniques of subject cataloguing and classification open the way to mastering IR systems, which have evolved in the course of centuries of development and will for a long time to come serve as large-scale tools of information work. Naturally, the active utilization of these systems requires a deeper understanding of their languages and practical experience in their use, but if you have formed an idea of the merits of these conventional systems and the limitations in their use which are imposed by their structure, it will be easier for you to proceed to the study of the descriptor-type IR systems, which are of fairly recent origin. They will be discussed in the next chapter.

Questions for self-checking

1. What is the purpose of the basic types of conventional IR systems and what are their specific features?
2. In what context do the problems of bibliographical entry occur and how are they resolved?
3. What properties distinguish the IR languages of subject IR systems?
4. What is the structure of a hierarchical library classification?
5. What are the merits and demerits of the UDC?
6. What are the advantages of the faceted schemes over the hierarchical classifications?

Literature

1. Foskett, D.J. Some Fundamental Aspects of Classification as a Tool in Informatics. In: On Theoretical Problems of Informatics (FID 435). Moscow, 1969, p. 64-79.

2. Guide to the Universal Decimal Classification (UDC). B.S. 1000C:1963. (FID No. 345). London, British Standards Institution, 1963, 128 p.

3. Metcalfe, J. Subject Arrangement and Indexing of Information. Notes for Students. Sydney, Bennett, 1966, 181 p.

4. Needham, C.D. Organizing Knowledge in Libraries. An Introduction to Classification and Cataloguing. London, Deutsch, 1964, 259 p.

5. Sayers, W.C.B. A Manual of Classification for Libraries. 4th ed. London, Deutsch, 1967, 404 p.

6. Vickery, B.C. Classification and Indexing in Science. 2nd ed. (enl). London, Butterworths, 1959, XIX, 235 p.

7. Vickery, B.C. Faceted Classification. New Brunswick, Rutgers Univ. Press, 1966.

7. DESCRIPTOR INFORMATION RETRIEVAL SYSTEMS

We have made it clear in the preceding chapter that the conventional information retrieval systems, both subject systems and those based on hierarchic and faceted classifications, along with definite advantages suffer from a certain limitation: they do not provide for document search by any combination of characteristics that has not been established beforehand. The conventional IR systems, particularly alphabetical subject systems and hierarchical classifications, become too cumbersome when confronted with the necessity of providing for multi-aspect search, i.e. searching for documents by multiple characteristics belonging to the different aspects in which a subject or phenomenon is viewed. This limitation becomes even more pronounced when a search for information is carried out in a file of documents that have many such aspects, and the aspects do not fit into a natural hierarchy of generic relations.

An instance of the limitation of conventional IR systems

To remind you of this feature of the conventional IR systems, we shall use as example an ordinary and deliberately simplified case where these systems appear in a schematic form. Let us assume that we conduct an information search in a file of motion pictures according to four aspects containing two characteristics each. Motion pictures are normally divided, according to their content, into documentary and art(non-documentary) films; according to their format, into normal and wide-screen; according to their colour, into black-and-white and colour; according to their length, into short and full-length.

In a *subject system* the name of each of these eight characteristics will form a heading and all of them will be arranged alphabetically:

Art (non-documentary) films
Black-and-white films
Colour ....
Documentary ....
Full-length ....
Normal ....
Short ....
Wide-screen ....

The divisions under each of these headings will list the address codes (numbers) or titles of those motion pictures which possess the relevant characteristics. You remember that this system is easy to use, practically does not require any preliminary study by the user of the principles of its build-up, and will quite readily accept new characteristics. For example, if we were to index educational or popular science films, stereoscopic or cinerama films, we would only have to insert into the alphabetic listing the subject headings corresponding to these characteristics. Then in answer to a request for a film possessing any of these characteristics the system would promptly respond with the needed information.

The situation is radically changed, however, when choice of a film by a combination of several characteristics is desired. The system is unable to provide an answer to a multi-aspect query such as this:

"what colour wide-screen art films or films with alternating colour and black-and-white images are available?" To supply an answer, all the relevant subject headings would have to be looked up and the matching film titles identified. The subject system is not designed for this task. We know that a way of overcoming this limitation is to form compound subject headings which combine several characteristics to provide for the would-be requests. However, it is not always possible to foresee all possible requests, and apart from this, the system then becomes too complex and cumbersome.

Now let us look at a *hierarchic classification*. Taking the first letters of the listed characteristics as the codes of relevant subject headings, and observing the formal rules of construction of hierarchic classification, we shall obtain the classificatory tree presented in Fig.26. This example obviously demonstrates that the hierarchy of generic relations will not at all necessarily be a natural one. Here the choice of the aspects to serve as the characteristic of division at the upper levels of hierarchy depends on the nature of the potential information requests. But they are hard to predict, for they depend both on the purpose of the show, on the type of film-goers, on the time they are ready to spend, and on the projection equipment available.

With the classification chosen in Fig. 26, the system is unable to provide an answer to a multi-aspect request on the availability of short or full-length motion pictures (eight subject headings will have to be scanned) or the black-and-white and colour pictures (four headings will have to be scanned). It is also clear that the necessity of following the formal rules of logic in the construction of a hierarchic classification contributes to the formation of rarely used headings.Thus, according to the above given example, the headings ANBS, ANCS, AWBS and AWCS will not contain very many entries, since short art films are produced infrequently.

To enable the system to give answers to any request using any combination of characteristics, it would be necessary to employ successively each aspect of each characteristics at each level in the hierarchy. This is impracticable because of the very great number of subject headings that would be formed. Another important point to make is that in order to introduce new characteristics, e.g. for the indexing of educational and popular science films or stereoscopic and cinerama films the system would have to be considerably revised to accommodate new classes and subclasses with all the subordinate headings.

Motion pictures where the colour and black-and-white images are alternatingwould be entered under several subclasses at the same time. To enable the system to retrieve only these films, new subclasses uniting both characteristics would have to be formed within it, thereby further increasing its complexity. The system would be made ever more complicated if it were necessary to incorporate into a hierarchical classification of films new principles of division, for instance, according to their country of origin (domestic, foreign), or according to the age group of the film-goers for which they are meant (for children,for adults), and so on. It would result in new levels of hierarchy and many new subject headings.

Faceted classifications are able to eliminate some of these inadequacies. Since in compiling them no attempt is made to list as many combinations of characteristics as possible, but instead a sort of building-block set is offered in the shape of foci (characteristics) grouped in facets (aspects), the indexer can omit practically unessential characteristics without running the risk of disturbing the hierarchy. He can also form new subject headings of a specific narrow interest. In our example, for indexing short documentary newsreels a generalizing subject heading D:S could be set up, and for art films we could use the subject headings A:N:B; A:N:C; A:W:B; A:W:C. If necessary, to these can be added, leaving the system intact, subject headings for motion pictures with alternating colour and black-and-white images A:N:BC; A:W:BC, as well as subject headings for the new characteristics (foci)- educational films, cinerama films, etc.

It would not be too difficult to insert new aspects of pictures (e.g according to their place of origin, the film-goers age group, etc.)into the faceted classification but it will still not ensure a search for any combination of characteristics, because for this it would have to comprise no fewer subject headings than a hierarchic classification. In our example the system fails to answer a query on all short colour films or black-and-white wide-screen films.

Therefore, the conventional IR systems, which have for centuries been evolving to provide answers to broad thematic and single-aspect requests, have proved to be poorly equipped for specific and multi-aspect searches, and for searches for any combination of characteristics not previously established. The most extensively used conventional IR systems, those based on hierarchical classifications, have great difficulty in catering for the ever increasing number of multidisciplinary problems often without any clear-cut generic relations.

These circumstances led to the emergence, some twenty years ago, of a new method of information retrieval which was to be called *coordinate indexing*. This method forms the basis of the descriptor-type IR systems, discussed in this chapter.

Set-theory terminology

In presenting the essentials of coordinate indexing, it is convenient to use the elementary terms of set theory, some of which it may be useful to introduce here. A *set* is a collection of objects: letters of the alphabet, articles in periodicals, books on shelves, the numbers 1,2,3,4, 5 - each of these is an example of a set, which may contain as little as one element or be empty (contain no elements). Sets may also be infinite, e.g. the infinite number of points on a circle. A set is generally denoted by writing its elements within braces,e.g. $\{1,2,3,4,5\}$. The statement "A is a subset of B" (i.e. every element of A is an element of B) is written in a contracted form as $A \subset B$ or $B \supset A$ (the *inclusion* relation). An empty set is denoted by zero; a set which comprises all the objects in a given field is called universal.

The main operations on sets are the logical sum, logical product, logical difference, and logical complement. The *union*(sum) $A \cup B$ of two sets A and B is a set each element of which is an element of A or of B or of both. The union of the set $\{1,2,3\}$ and the set $\{2,3,4,5\}$ is the set $\{1,2,3,4,5\}$. The *intersection* (product) $A \cap B$ of two sets A and B is a set each element of which is simultaneously an element of A and of B. The intersection of the sets $\{1,2,3\}$ and $\{2,3,4,5\}$ is the set $\{2,3\}$. The *difference*, A - B, between two sets A and B is a set of all elements of A that are not elements of B. The difference of the sets $\{1,2,3\}$ and $\{3,4,5\}$ is the set $\{1,2\}$. The *complement*, A/B, of a set A in a set B is a set of all elements of B that are not elements of A. The complement of the set $\{1,$

2,3} in the set {2,3,4,5} is the set {4,5} . The completement, A or A', of a set $\bar{A}$ in the universal set is a set of all elements of the universal set that are not elements of A. The graphic representations of the relations between sets are given in Fig. 27.

Coordinate indexing

The coordinate indexing method is based on the assumption that the semantic contents (subject) of a document and information request can with sufficient accuracy and completeness be expressed by an appropriate list of so-called key-words which are explicitly or implicitly contained in the text being indexed. By *keywords* we mean the words which are most essential for expressing the main meaning of a word or phrase, which have a nominative function. Most nouns, adjectives, verbs, adverbs, numerals and pronouns can be used as keywords. Prepositions, conjunctions, connectives, particles and other functional words cannot be keywords. To put it differently, coordinate indexing is a method of expressing the primary subject of a document or information request by a given number of keywords. In pure coordinate indexing, the keywords in the search patterns are not related to each other and function independently. In contrast to search patterns of documents, search request formulations are presented as logical sums, products or complements of the classes designated by the corresponding keywords. In order to locate a document matching an information request it is necessary to do certain logical operations on the classes designated by the keywords in the search patterns of the documents.

In the simplest case, when a search request is formulated as the logical product of a certain set of keywords, the document is held to be a match if its search pattern contains all the keywords of the search request formulation. In a sense, this is equivalent to the operation of logical product of classes designated by those words of the search pattern which coincide with the keywords constituting the search request formulation.

To illustrate this point we shall again take the example given in the beginning of this chapter. Suppose that our collection of films is indexed by keywords denoting the characteristics that have been chosen. To take out of this collection the short films with alternating black-and-white and colour images (a difficult question for conventional IR systems)it will be sufficient to state the request as the logical product of the three keywords describing these characteristics: $S \cap B \cap C$. Another, more abstract example was given in Fig. 16, at the beginning of the fifth chapter, where we conducted a search by means of keywords designated by the letters A to H. We wanted all documents containing characteristics C and D or C and F. Using the terminology that we have introduced we can now say that the request formulation was composed as a logical sum of two products of keywords $(C \cap D) \cup (C \cap F)$.

Externally, the coordinate indexing method appears to resemble an IR language of a subject system. In either case we are dealing with classes designated by appropriate natural-language words and phrases, that is, by keywords and subject headings. Each keyword, e.g. 'Electric', covers a class of subjects and concepts, the designations of which include this word. The same is true of the formulation of subject headings. But the similarity ends there.

In coordinate indexing, in order to locate a document whose search pattern is composed of a given number of keywords, it is necessary to perform the logical operations of sum product or complement on the classes designated by these keywords. This provides an opportunity of conducting a multi-aspect search for any previously unspecified combination of keywords. In the subject systems, on the other hand, as well as in the hierarchical and faceted classifications, each subject heading or classification number appears independently, listing the address codes (numbers) of all relevant subject material. In conventional IR systems simultaneous searching for several subject headings or class numbers is a difficult matter indeed, and sometimes a downright impossibility, as they are not designed for such use. For example, a search for documents using any combination of five mutually exclusive characteristics will require, in a system based on coordinate indexing, a search request formulation as a logical product of five keywords. In any conventional IR system the same task will involve scanning the contents of as many subject headings as there are permutations of five headings of classification numbers, i.e. $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. This constitutes one of the most important advantages of coordinate indexing over the IR languages of the conventional systems.

Nevertheless the procedure of pure coordinate indexing described above is inadequate since it does not ensure the necessary levels of recall and precision. Some of the main reasons for this are as follows. Natural language words from which keywords for coordinate indexing are chosen have the property of *synonymity:* the same thing or idea may be denoted by different words. Such words which are mutually replaceable in similar contexts without giving a change in meaning are called synonyms. Examples of synonyms are 'test' and 'trial', 'linguistics' and 'science of language', 'common salt' and 'sodium chloride'. It can be easily understood that if the same subject is designated by different words in the search pattern of a document and in a search request formulation, the document, though relevant to the request, will not be retrieved.

Another source of the inadequacy of pure coordinate indexing is the multiple meaning of natural language words. It manifests itself in *polysemy,* which is a transfer of meaning from one subject to another having common qualities or characteristics, and in *homonymity,* which refers to the accidental coincidence of different words that have nothing in common either in respect of origin or meaning. Examples of polysemantic words are 'star' - a celestial body or a geometrical figure, 'bureau' - a kind of table and a collegiate body; examples of homonyms are 'mail' - the system of communication and an armour; 'ray' - a beam of light and a sea-fish. Evidently, the multiple meaning of keywords may bring about the retrieval of non-relevant documents which means lower precision values and increased 'noise'.

One more source of insufficient effectiveness of information retrieval by coordinate indexing may lie in *unspecified relationships* between keywords, which result in the non-retrieval of relevant documents - that is, in poor recall. Let us assume that a search request is for documents on the properties of liquids, and the search request formulation includes the keyword 'liquid': if it is in no way indicated that water is a liquid, the documents dealing with the properties of water will not be retrieved. The most important relations between keywords are the generic relations, but certain significance is attached to other relations, particularly to the associative relations, i.e. the relations between keywords by association: by contiguity ('chair' and 'table'; 'fall' and 'autumn'), by similarity ('clock', 'scales' and 'thermometer'; 'pond', 'lake' and 'sea'), by contrast ('black' and 'white'; 'light' and 'heavy'), etc. All the conceptual relations between keywords, based on the relations between the ideas, things or phenomena they designate, are called paradigmatic relations.

Descriptor information retrieval language

It follows from the examples discussed above that, in order to effect a considerable increase in the precision and recall of information retrieval based on coordinate indexing, it is necessary at least to establish control over the vocabulary of the IR language used. Such a control should consist in the elimination of synonymity and in compiling special dictionaries, tables or charts graphically demonstrating the most essential paradigmatic relations between the keywords. With full vocabulary control, the coordinate indexing of search requests and documents will make use of only those words contained in a given standard list, in which their synonymity and polysemy has been eliminated and the paradigmatic relations are shown. Such keywords have been termed 'descriptors'.

Descriptors are standardized keywords, designed for coordinate indexing of documents and information requests, selected according to certain rules from the main vocabulary of a particular natural language and artificially (by cross-references and notes) freed from synonymity, polysemy and homonymity. A *descriptor language* is a specialized IR language, whose vocabulary is composed of descriptors and whose grammar consists, in the simplest case, of a method of building search patterns and search request formulations by correlating relevant descriptors.

It is impossible to establish now who was the originator of the idea of coordinate indexing. The British information scientist B. Vickery has noted that coordinate indexing with inverted file structure was apparently known in Sumer three thousand years ago. Clay tablets have been found, each assigned to a disease, with the names of all diseases characterized by a particular symptom also listed on the relevant tablet. Obviously, if according to the symptoms observed in a patient the corresponding tablets are selected and the name of any disease repeated on each of them is noted, it will in all probability be the disease from which the patient in question is suffering. These Sumerian tablets are the prototype of modern diagnostic machines. (1)

The principle of coordinate indexing is the basis of the so-called superimposable cards, which will be considered in the next chapter. In 1915, they were used by the ornithologist H. Taylor, in the 1920's they found applications in mineral identification and personnel selection, and in 1939 W. Batten (of Imperial Chemical Industries, Great Britain) used them for patent searching. The credit for developing the information retrieval systems which formed the basis for the development of modern descriptor languages, however, must go to the American scientists C.N. Mooers and M. Taube.

---

1. International Study Conference on Classification for Information Retrieval. Dorking, 1957. Proceedings. London, Aslib, 1957, p. 106.

C.N. Mooers in 1947 designed a mechanized document searching device which he named "Zatocoding system", and it was he who coined many of the terms which are now in wide use among information scientists, including 'information retrieval', 'information retrieval system', 'retrieval language ', 'descriptor' and 'descriptor dictionary'. In one of his earlier works, he has described the indexing method in his system thus: "Each document or unit of information is characterized by a set of descriptors taken from the vocabulary of descriptors. Each descriptor of the set applies to, or is true in some way, of the information content of the unit of information. The descriptors operate independently in this type of characterization. The fact that there are several descriptors in the set may mean that they formed some interacting combination in the original document, or it could just as well mean that they relate to independent ideas scattered through the document. Using descriptors in this fashion drops almost all relationship between the ideas represented by the descriptors.(1)

Mr. Taube made a great contribution to the theoretical substantiation, development and popularization of the ideas of coordinate indexing. He defined coordinate indexing as a 'method of analysing items of information so that retrieval is performed by the logical operations of the product, sum and complement on the codes in the store'(2). In 1951 he developed the socalled Uniterm system which has found extensive application since then. The main difference between the Uniterm and Zatocoding systems is that the former deals with words expressing concepts while the latter deals with concepts expressed by words.

Uniterm system

Taube's Uniterm is a keyword (usually a simple one) which may have an appropriate cross-reference or scope-note helping to eliminate its synonymity or multiple meaning. In contrast to the descriptors of C. Mooers, the Uniterms have no references specifying the paradigmatic relations between them. Uniterms may be keywords expressing single ideas, as well as proper, geographical and trade names. All Uniterms have an equal hierarchical rank: none of them occupies a conceptually superior position with respect to any other Uniterm (as in hierarchical classifications) and none is used in a pre-established combination with any other Uniterm (as in subject headings).

This method of building an IR language vocabulary contributes to a great reduction in its size. The list of subject headings in the Library of Congress subject catalogue contained some 50,000 entries; conversion to Uniterms resulted in a list containing only 3,000 words. The important feature of the Uniterm system is that the vocabulary is built in the process of using it and not developed beforehand, as in the Zatocoding system. During the initial period of operating the system, the number of Uniterms will grow steadily. Gradually, synonyms and polysemantic words will be detected and marked on Uniterm-cards. With the growth of the document collection, the rate of growth of the vocabulary will tend to slow down and eventually come almost to a standstill.

---

1. Mooers, C.N. Zatocoding Applied to Mechanical Organization of Knowledge. "American Documentation", 1951, V. 2, No. 1, p.26.
2. Taube, M., Wooster, H. (eds.). Information Storage and Retrieval. Theory, Systems and Devices. New York, Columbia Univ. Press, 1958, p.8.

Using as an example this system, one of the first and simplest systems based on the coordinate indexing method, it will be easy to become familiar with the practical techniques of input and searching in a descriptor-type IR system. For the physical implementation of his system, Taube developed a special Uniterm card having either the usual catalogue card size (75mm x 125mm) or larger in size (203mm x 125mm), on which a grid formed by one horizontal and ten vertical lines is printed. The horizontal line is used for writing a Uniterm. Uniterm cards are envisaged for all Uniterms in the system and are filed alphabetically, with due account taken of the inversion of those Uniterms consisting of more than one word.

The vertical columns are for the codes (numbers) of documents whose search patterns include the Uniterm indicated in the horizontal line of the card. The peculiarity of the system is the "terminal digit" order of writing these numbers on the card. This means, for example, that number 127 is written in the 7 column, number 239 in the 9 column, and number 270 in the 0 column. Since the documents are numbered as they are entered into the system, the numbers will be posted in the columns in ascending order. Such a system of recording considerably facilitates the visual scanning and locating of document numbers which are common to all the cards being matched. The procedure in this case amounts to the operation of logical multiplication of the Uniterms contained in a search request formulation. The method of work with Uniterm cards is demonstrated in Fig. 28.

Let us assume that we are to enter into the system a paper dealing with methods of 'corrosion protection of gas turbine blades'. For the coordinate indexing of this paper we need the Uniterms TURBINES, GAS, BLADES, CORROSION and PROTECTION, which jointly form its search pattern. The paper is assigned its serial address code 2005. Then the cards assigned to these Uniterms are removed from the file. If in doing this, the Uniterm PROTECTION or its synonyms were found not to be in the system, a card would be made for this particular Uniterm. The number 2005 is put in the 5 column of each of the relevant Uniterm cards, which are then replaced in the file. Similar procedures are performed for each incoming document.

An information search in a file of Uniterms is conducted in the following way. Suppose that a request is put to the system for papers dealing with 'gas turbine blades'. The search request formulation is made to contain the Uniterms GAS, TURBINES, BLADES. The cards for these features are extracted from the Uniterm file and the numbers common to all these cards are located. It is advisable to have as the basis for comparison the card which contains the least number of postings (in this case, the Uniterm card for BLADES). Matching should likewise begin with the columns with the least number of postings: columns that are empty in any of the cards need not be matched at all; where the base card has columns with one posting it is not necessary to match the postings which have been made before this single posting (i.e. smaller numbers). The time it takes to locate all common postings is usually negligible. Thus, in our example, it can be seen at once that the three upper Uniterm cards have only three matching numbers: 526, 1027 and 2005, which are the address codes of the documents sought. If the request is broadened to cover documents on gas turbines (two cards at the top), the output will increase to include five documents: 195, 294, 526, 1027, 2005. If, on the other hand, the request is made more specific to read "corrosion/of gas turbine blades" (four upper Uniterm cards) only two relevant documents will be retrieved, namely 1027 and 2005.

Thesaurus and its construction

In the Uniterm system, the control over the retrieval language vocabulary, as we have seen, is limited to the elimination of synonymity and polysemy. Full lexical control, however, which alone can ensure maximum precision and recall, especially with thematic searches and searches for documents only partially matching the request, necessitates the recognition of paradigmatic relations between the indexing terms. For this purpose special normative reference dictionaries called thesauri are compiled.

An information retrieval *thesaurus* is a reference dictionary designed to help the information user to state his information needs in terms of the descriptor language and to provide for finely detailed indexing of documents and information requests by these terms. It must contain all the descriptors used by the language of a given system, clearly displaying their conceptual relationships, and also the keywords within the system which are considered to be synonymous with these descriptors. Thus, the thesaurus will help to eliminate synonymity and polysemy of keywords, and the lack of explicit relations between them, which give rise to the defects of 'pure' coordinate indexing already discussed.

Synonymity of keywords is eliminated in the following way. In constructing a thesaurus, a list of keywords used for the coordinate indexing of documents is first compiled. Then groups of words which can be considered synonyms are brought together. From each of these groups a word or phrase is taken to represent the whole group, which becomes a descriptor. The rest of the words in the group are considered synonyms of the descriptor and linked to it by 'see' or 'use' references. Each descriptor is linked with all its synonyms by reversed 'includes' or 'used for' references, e.g.

ABSTRACT <u>includes</u>	speculative theoretical
speculative <u>see</u>	ABSTRACT
theoretical <u>see</u>	ABSTRACT.

Polysemy of keywords is also provided for, in building up a dictionary of descriptors, by affixing, to each multiple-meaning keyword an alphabetic or numerical symbol and a word qualifying its meaning. Another method is to replace polysemantic keywords by descriptors composed of one-value phrases, e.g.:

atlas(geographical) see geographical atlas
atlas(vertebra) see cervical atlas
filter - 1 (chemical)
filter - 2 (electrical)
filter - 3 (gas)
filter - 4 (optical)

The conceptual (paradigmatic) relations between discriptors are also displayed in the thesaurus: according to our definition, these are relations based on the existence of objective connections between the ideas, objects or phenomena denoted by the descriptors. In this case synonimity is not taken into account, because in a descriptor language it is eliminated and does not exist within one system. The following paradigmatic relations can be cited as the most important ones:

species - genus (genus - species);
collateral subordination;
similarity (functional);
cause - effect (effect - cause);
part - whole (whole - part).

Of particular importance are the generic relations and the collateral subordination relations realizable through them. It is these relation types that form the groundwork of hierarchical classifications. Almost every concept can be both generic (i.e. it can reflect major features of a class of objects that includes other classes of objects which are species of this genus) and at the same time specific. The notion of 'rectangular', for example, is generic to the notion 'square' and specific to the notion 'parallelogram'. Exceptions are constituted only by the broadest concepts or categories (e.g. 'matter', 'space', 'time'), which have no generic concepts, and by the most narrow, unique notions (e.g. 'UDC') which have no specific concepts. Concepts which are equally subordinated to one generic concept are said to be collaterally subordinated, e.g. the concepts 'phonetics','lexicology' and 'grammar' are collaterally subordinate to the generic concept 'linguistics'. It should be noted that, in contrast to hierarchical classifications, descriptor languages take into consideration generic relations irrespective of the hierarchic level to which they belong.

A thesaurus indicates generic relations independently of other paradigmatic relations which may sometimes outwardly resemble them. This applies to the functional similarity relations (clock - scales - thermometer), causal relations (tiredness - sleep), and particularly to the relations of the 'part - whole' kind: of all these, it can be said that they differ from generic relations in that they represent relations between objects and not between concepts; 'genus' and 'species' are abstractions, while 'whole' and 'part' are concrete things.

Thus, an information retrieval thesaurus generally consists of three parts:

1. The vocabulary part is a normal alphabetic list of descriptors, together with keywords regarded in this system as synonyms of these descriptors. Descriptors are commonly made more prominent in the listing (e.g. by printing them in capitals) and are linked by cross-references to and from all their synonyms. Polysemy and synonymity of keywords are eliminated using the method described above.

2. The 'semantic map' of the retrieval language vocabulary is a network of conceptual classes in which all the descriptors of a given language are grouped. This part of the thesaurus provides a graphic demonstration of the essential paradigmatic relations between descriptors, at least of their generic relationships. These relations are expressed in one of two ways: either by combinations of the alphabetically listed thematic classes (fields) containing a multiplicity of thematic groups of descriptors also listed alphabetically, or by charts in which the basic paradigmatic relations are indicated by arrows.

3. The rules of conversion of keywords and key phrases of the natural language into a descriptor-type IR language determine the procedure of substitution of descriptors for these keywords and phrases. The rules define the conversion of the names of institutions, chemical compounds, biological species, and other similar categories of terms, and also include the rules for lexicographical editing of search patterns and search request formulations, e.g. rules for complementing them with descriptors connected with the main descriptors by generic and other paradigmatic relations.

To illustrate the thesaurus structure two examples will be given. The first published thesaurus to use the graphic method of displaying paradigmatic relations between descriptors was the "Euratom-Thesaurus" (first edition) for nuclear physics and engineering.(1) It contains 4,470 descriptors, including 1,836 names of inorganic chemical compounds and 1,404 names of isotopes. The 42 subject classes, numbered 00 to 94 (with lacunae) are in the form of charts in which the arrows are directed from the generic to specific descriptors. The collaterally subordinate descriptors are marked by two-way arrows. These classes cannot be considered as classifications; they broadly correspond to those subject fields that are of interest to Euratom. The vocabulary part provides a common alphabetical list of descriptors with references to the relevant subject classes, and a common alphabetical list of their synonyms which are associated with the corresponding descriptors and their subject classes. The generic relations are thus completely ignored in the vocabulary part. Fig. 29 shows the subject class 71 'Mathematics' from the "Euratom-Thesaurus" (first edition). The arrows which go beyond the limits of a subject class join its descriptors with the descriptors of the other subject classes and their numbers. To the right, there is an alphabetical listing of all descriptors which belong to this particular class, and at the bottom there are numbers of all subject classes used in the thesaurus.

Another example is the "Thesaurus of Engineering Terms" which was published in 1964 by the Engineers Joint Council (USA) (2). In contrast to the Euratom thesaurus, it indicates generic and other paradigmatic relations by means of a system of references and scope notes. The thesaurus contains 10,515 words of which 7,750 are descriptors. The listing of the descriptors and their synonyms in the vocabulary part is alphabetical. The synonyms are linked with the descriptors meant to replace them by 'Use' references. In the 'semantic map', which has the form of an alphabetical listing of the basic (heading) descriptors, each entry includes the heading descriptor, its synonyms, the specific descriptors, the generic descriptors and the descriptors connected with it by other paradigmatic relations. To designate them the following notes are employed:

UF (used for) - for synonyms;
BT (broader term) - for generic descriptors;
NT (narrower term)- for specific descriptors;
RT (related term) - for other descriptors.

---

1. Euratom-Thesaurus, Keywords used within Euratom's Nuclear Energy Documentation Project. EUR 500.e (1st ed.) Brussels, 1964, 80 p. (European Atomic Energy Community).

2. Thesaurus of Engineering Terms: a list of engineering terms and their relationship for use in vocabulary control, in indexing and retrieving engineering information. 1st ed. New York, Engineers Joint Council, 1964.

The ampersand (&) is used to mark those descriptors which can be used in place of other descriptors expressing more narrow concepts and marked by the symbol # .

The second major thesaurus published by the Engineers Joint Council, "Thesaurus of Engineering and Scientific Terms", 1st ed. (New York, 1967), uses some other symbols. The dagger (†) in front of a term signifies that two or more descriptors are to be used in coordination for that term. The dash (–) symbol in front of a descriptor indicates that the descriptor has narrower terms (not shown) and that the main entry should be consulted to determine these. Below is a simple dictionary entry from the "Thesaurus of Engineering and Scientific Terms":

Scientists 0509

UF Scientific personnel
BT Personnel
   Professional personnel
NT Chemists
   Physicists
RT Engineers

A few words about the term 'thesaurus' (from Greek 'thesauros', literally meaning a treasure, treasury or storehouse) might be appropriate The word seems to have been used first in its present meaning by Florentine Brunetto Latini (1220-1294) in his encyclopaedia "Li Livres dou Trésor". In the 16th century, the word appeared in the names of the Latin and Greek lexicons "Dictionarium, seu Lingue Latinae Thesaurus" (1532) and "Thesaurus Linguae Graecae" (1572), which were published by the Estiennes, well-known French philologists and publishers. In contemporary usage the term refers to the dictionaries of concepts, or ideological dictionaries, which are in effect inverted explanatory dictionaries. Of these, the greatest popularity is enjoyed by the "Thesaurus of English Words and Phrases", compiled by P.M. Roget in 1852 and since reprinted at least 90 times. Similar dictionaries, linking all words in thematic groups or subject classes, are to be found in French, German, Spanish and other languages.

One of the first to recognize the need for this type of dictionary in information retrieval work was the American bibliographer C.L. Bernier who wrote in 1957: "A limited thesaurus would seem to be another effective way of bringing the relevant terms to the attention of the searcher if the vocabulary proves too large to be read completely each time for selection". (1)

Grammatical resources of descriptor IR languages

In using descriptor languages, the search pattern of every document and search request formulation is stated in the form of an unordered set of descriptors. In addition to their paradigmatic relations, however, which in some way or other are recognized in the indexing and in searching assisted by a thesaurus, there exist different relationships between descriptors that derive from the document contexts. Such relations between the descriptors are said to be *syntagmatic*. If these re-

---

1. Bernier, C.L. Correlative Indexes. II: Correlative Trope Indexes. "American Documentation", 1957, V. 8, No. 1, p. 48.

lations are ignored, the descriptors belonging to a document search pattern may form false combinations or 'false drops' which result in the retrieval of non-relevant documents, increased 'noise' and hence lower precision.

The following examples will illustrate this point. Suppose that our system includes a document on the 'production of sulfuric acid and catalyst purification'; its search pattern will contain the following descriptors - PRODUCTION, SULFURIC ACID, CATALYST and PURIFICATION. These descriptors may, during a search, form false combinations which will result in the retrieval of this document in answer to a request on the 'purification of sulfuric acid' and on the 'production of catalyst', although the document does not deal with either of these. (1) Another example: a document dealing with the 'coating of copper tubes with lead' is indexed by the following descriptors - LEAD, COATING, COPPER, TUBES. If a request for documents on lead tubes is entered into the system, the request formulation will be indexed by the descriptors LEAD and TUBES, and the document will be retrieved, although it is not relevant to the query. (2)

Roles (role indicators) and links are the principal grammatical resources used in descriptor languages to reduce noise. *Roles* are special symbols which are attached to a descriptor and reduce the scope of the concept it stands for. This is achieved by indicating the logical role which a given descriptor plays in a particular context. In the first of the examples given, in order to prevent false combinations it will be enough to affix to the descriptor SULFURIC ACID in the search pattern the role indicator "B" showing that it is a product of chemical or industrial reaction, and to the descriptor CATALYST the link "C" indicating an undesirable component (waste, impurity, admixture, spoilage). The document search pattern will then be indexed - PRODUCTION, SULFURIC ACID-B, CATALYST-C, PURIFICATION,whilst the request formulation will read - CATALYST-B, PRODUCTION, SULFURIC ACID-C, PURIFICATION. The document will then not be retrieved in response to the queries asking for documents on catalyst production and sulfuric acid purification.

Links are also special symbols attached to the descriptors in the search patterns of documents (or their address codes) and designed for the conceptual grouping of these descriptors. In the second of our examples , the search pattern of the document on the 'coating of copper tubes with lead' would assume this form by the addition of links - LEAD R_1, COATING R_1, COPPER R_2, TUBES R_2. The link R_1 would join the descriptors LEAD and COATING, and the link R_2 the descriptors COPPER and TUBES. If the condition is stipulated, that a document is to be produced on request only when the matching descriptors in its search pattern and in the search request formulation have the same links ascribed to them, this document will not be retrieved in answer to a query on 'lead tubes'. But then it would also not be retrieved in response to a request for documents dealing with 'tube coatings' generally. This would mean an information loss. The search pattern can be made more complex by the addition of several links: LEAD R_1, COATING R_1R_3, COPPER R_2R_3, TUBES R_2R_3.

---

1. Holm, B.E. Information Retrieval - a Solution. "Chemical Engineering Progress", 1961, V. 57, No. 8, p. 74-75.
2. Taube, M. Notes on the Use of Roles and Links in Coordinate Indexing. "American Documentation", 1961, V. 12, No. 2, p. 98-100.

If now the condition stipulates at least one common link in the matching descriptors, the document will be retrieved in response to questions on 'tube coating', 'lead coating' and 'coating of copper', but will not be retrieved by a false correlation in response to a request for 'lead tubes'

Experience indicates that the best results are produced by the joint utilization of links and roles. They reduce noise by 10-15%, but considerably increase the cost and time of indexing.In relatively small document collections (up to 30,000 items) the increased precision of retrieval achieved by such grammatical means will not balance the increased time and cost of indexing and searching, which means that under certain circumstances the use of descriptor languages without grammar is more advantageous.

Further development of grammatical means of the descriptor languages has resulted in specialized IR languages in which the semantic relations between descriptors are developed to a greater extent. These relations and the descriptors themselves are designated by complex codes and the search pattern of a document has the form of an encoded 'telegraphic abstract'. In one such language the notion 'thermometer' is treated as a machine or device(MACH.) which effects (U) measurement (MUSR) and is affected (W) by heat (RWHT.) and is designated by the descriptor MACH. MUSR.RWHT. Among the best-known of such languages are the WRU Semantic Code, SYNTOL, the language of RX-codes and a few others. They are still in the experimental stage and designed primarily for automating information retrieval with the aid of computers. Familiarization with these languages is beyond the scope of our introductory course.

We have now considered some of the fundamental problems relating to descriptor-type IR systems: the reasons for their appearance and their advantages over conventional systems; the principles of coordinate indexing and its inherent weaknesses; the basic concepts of descriptor languages; the operation of some of the simpler systems based on these languages; the construction of the thesaurus and its range of applications and, finally, the grammatical resources of descriptor IR languages. To conclude the discussion of information retrieval problems we have only to consider the technical means of implementing IR systems, which will be the subject of the next chapter.

Questions for self-checking

1. How do the inherent limitations of conventional information retrieval systems manifest themselves?
2. What is coordinate indexing and what are its advantages over the IR languages of conventional systems?
3. What are the definitions of a descriptor language and of its principal concepts?
4. How is a Uniterm system designed and how does it work?
5. What is a thesaurus, its structure and purpose?
6. What is the function of the grammatical resources of a descriptor language?

Literature

1. Costello, J.C. Training Manual and Workbook for use in Abstracting and Coordinate Indexing. Training course. Columbus, Battelle Memorial Institute, 1964, IX. 117, 11 p.
2. Howerton, P.W. (ed.). Information Handling: First principles. Washington, Spartan Books, 1963.
3. Lancaster, F.W. Information Retrieval Systems, Characteristics, Testing and Evaluation. New York (a.o.), Wiley, 1968, XIV, 222 p.
4. Taube, M. and Wooster, H. (eds.). Information Storage and Retrieval. Theory, Systems and Devices. New York, Columbia Univ. Press, 1958.
5. Vickery, B.C. On Retrieval System Theory. 2nd ed. London, Butterworths, 1965, XII , 191 p.

8. TECHNICAL FACILITIES FOR INFORMATION RETRIEVAL

While discussing information retrieval systems in general in the last three chapters we hardly mentioned the technical facilities for their implementation. The only exceptions were library systems and Uniterms, based on such elementary facilities as bookshelves and catalogue cards. As you may remember, we did this on purpose, being aware that diverse technical means can be used to implement one and the same system and that it is therefore convenient to consider all these together.

Before dealing with them, however, we must first discuss a number of general aspects of mechanization and automation of information retrieval.

Mechanization and automation of information retrieval

Technology is rapidly penetrating all spheres of human activity but the system of scientific communication undoubtedly still remains in this respect at the level of past centuries. It was no accident, therefore, that the first paper to state that scientific information work was one of the major scientific problems urged the development of special information machines. The paper was by an American scientist, engineer and public figure V. Bush, who wrote: "Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose."(1)

Certain successes have been achieved in that respect during the past quarter-century, but we are still far from a solution to the problem of mechanized and automated information retrieval. Futile attempts to design a 'big information machine' finally brought the realization that better results would be achieved with standard punched card and electronic digital computers in combination with microfilming and other reprographic facilities. But some people still overestimated the role to be played by machines in solving information retrieval problems, and this attitude still survives today despite the fact that information retrieval problems can obviously be solved only through a profound study of man's thought processes.

One is reminded of the French mathematician L. Brillouin who wrote: "Too many laymen picture science as a sheer accumulation of facts, and take a scientist for a living encyclopaedia. They think it possible to solve every problem by means of giant machines which might register all human knowledge in their enormous, magnetic memories. Do you need some information? Push a button and the machine will answer! This is ridiculous, and stems from the poorest science fiction. More important, it proves that the role of the scientist is too often completely misunderstood". (2)

---

1. Bush, V. As We May Think. "Atlantic Monthly", 1945, V. 175, No. 1, p. 101; reprinted in: Readings in Information Retrieval. New York-London, 1964, p. 20.
2. Brillouin L. Scientific Uncertainty, and Information. New York-London, Academic Press, 1964, p. 39.

This, of course, does not imply that mechanized or automated retrieval is unnecessary. But complex facilities should only be introduced into information processes if they bring real economic gains by reducing the operational costs of retrieval systems, appreciably speed up retrieval, or permit some retrieval functions that in principle are not feasible in manual systems.

It should be kept in mind that machines are not the priority problem of automation and that it is not with them that automation must start. Machines may be entrusted only with those operations that have been previously formalized or presented in the form of algorithms. An algorithm is defined as a precise and unambiguous statement of all the operations comprising a process, and of their sequence. Hence, the first thing to do is to learn to make formal statements of the complex thought processes that occur during document analysis, abstracting, indexing and retrieval. This is an important general scientific problem which goes far beyond the scope of scientific information activities and informatics.

We have dealt at some length with the relative merits and shortcomings of conventional and descriptor systems. Some believe that the former are difficult to mechanize while the latter are easy to mechanize: this is not exactly true. Conventional systems, particularly those based on UDC, can also be mechanized; but this is not usually necessary since they are efficient enough when implemented as card catalogues. The case is different with descriptor systems for which mechanized or automated retrieval is to some extent imperative, since manual searches for descriptor combinations are feasible only in relatively small files.

Material information carriers

Technical facilities in wide variety - ranging from the simplest information retrieval devices up to the latest generation of electronic computers - are currently used to mechanize and automate information retrieval. Pursuing the objectives stated earlier, we shall not go into detail but concentrate on some simpler devices. Nevertheless it seems desirable to obtain a general picture of the whole range of equipment used. Usually information retrieval devices are classified by the information carriers they use. These are conventionally divided into discrete and continuous carriers, and a possible classification of some of the principal information carriers is as follows:

Discrete information carriers
- Catalogue cards
- Uniterm cards
- Punched cards
  - hand-sorted
    - edge-punched
    - body-punched
    - superimposable
  - mechanically sorted
- Microfiches

Continuous information carriers
- Microfilm
- Punched tapes
- Magnetic tapes (disks, drums)

Catalogue cards and Uniterm cards have been described in the preceding chapters, so we shall come straight to punched cards.

Punched cards

A punched card is a stiff paper rectangle with either holes punched along its edges or positions for punching such calibrated holes marked over the whole card surface. Punched cards were first applied in the 1780s by J.M. Jacquard who used them to control his looms. Early in the 1890s the American H. Hollerith introduced a pantograph punch and an electric accounting tabulator for body-punched cards which he offered for processing the U.S. census data: punched cards of this particular type are therefore sometimes called 'Hollerith cards'. He founded a firm which later grew into a major computer company - International Business Machines (IBM).

The division of punched cards into hand- and machine-sorted is purely conventional since machines can be used to handle all kinds of cards, though edge- and body-punched cards, as well as superimposable cards, are usually processed manually. It should be emphasized that punched cards and all the other information carriers were not created expressly for information retrieval; they are used in record-keeping, statistics and for some other purposes.

Edge-punched cards are used in document retrieval systems with serial organization of AS. One or more rows of holes are punched along the card edges. Each hole is a binary code position which permits coding the presence or absence of a characteristic. The presence of a characteristic is indicated by clipping out the partition separating the hole from the card edge. The only thing needed to select the cards having a notch in a specified position (that is the cards on which the respective characteristic is recorded) is to insert a sorting needle in a card pack and shake it slightly, when the cards with the notch will drop out.

Edge-punched cards are produced in different sizes and with varying numbers of positions. For their visual alignment in the file, the right-hand upper corner of every card is cut off. Most popular are cards of 207 x 147 mm with two rows of holes (a total of 215 positions of which 200 are used). Various types of auxiliary equipment are used to manipulate such cards. These include tongs for edge-notching individual cards, keyboard devices - both hand and electrically operated - for notching card packs, sorting-needles, and diverse sorting boxes and selectors.

The maximum number of recorded different characteristics is the most important parameter of an edge-punched card. Since the characteristics are recorded on the card by their codes, the number of holes along the edge is called its code field. The information capacity of a code field, or the number of different characteristics it can accommodate, is determined by the number of holes and the type of code applied. According to the method of representing the code words on an edge-punched card, a code may be direct or combined; according to its designation - selector or sequence; according to the type of encoded symbols - numerical or alphabetical; and according to positioning of the records in the code field - local or superimposed.

The *direct code* puts every characteristic into one-to-one correspondence with a hole in the card's code field. The direct code can be used to record as many different characteristics as there are positions in the code field. Advantages of the direct code are: no "retrieval noise" and minimal time and labour costs of encoding and card sorting. The drawbacks are the limited number of characteristics by which documents can be encoded and searched and the fact that only mutually exclusive characteristics can be encoded in every pair of holes in the case of a double-row perforation. The first shortcoming necessitates the application of *combined codes* - selector or sequence; the second shortcoming is common to all local codes, and to get rid of it a superimposed code is applied.

The *selector code* is a combined code that provides for retrieving the required cards from a file by means of a minimal number of sortings. The selector code puts every characteristic into one-to-one correspondence with a configuration of the code symbols recorded in a certain part of the code field (a subfield) reserved for a given group of characteristics. Every word encoded by the selector code must comprise the same number of code symbols. Only one characteristic out of a given group may be encoded in a subfield whatever combined code is applied. This is a shortcoming of all local codes which makes them suitable for mutually exclusive characteristics only. The selector code is applicable both with single- and double-row perforation.

The *triangular (pyramid) code* is a variety of selector code. It records every symbol with two notches in the positions situated at the upper ends of two diagonal columns of the coding network at the intersection of which the symbol to be coded is located. The advantage of the triangular codes as compared with linear ones is the smaller code field required. When the triangular code is applied to cards with double-row perforation the symbols are recorded with deep and shallow notches. The deep notches are made on the side on which the symbol in question lies at the intersection of the diagonal columns. In case the symbols are located one above the other the upper symbol corresponds to the right-hand deep notch.

The *sequence code* is a combined code providing for arranging the cards in a required order by a minimal number of sortings. With this code, every characteristic is put into one-to-one correspondence with a configuration of code symbols inside a code subfield. The configuration may consist of any number of symbols, provided their number does not exceed the number of the positions in the subfield. The sequence code has the common shortcoming of the local codes and therefore it is fit for mutually exclusive characteristics only. But its advantage is that it takes smaller code-field space than the direct and selector codes. The price of this, however, is greater retrieval noise due to more false drops of cards with coinciding notches.

All the above mentioned types of code may be either *numerical* or *alphabetical*. An alphabetical code may be regarded as a version of a numerical code since each letter of the alphabet is, explicitly or implicitly, uniquely designated by its ordinal number in the alphabet. When a sequence code is used to represent letters, their rapid alphabetic ordering is possible.

The *superimposed code* is a code meant to record several characteristics in one and the same code field. The superimposed code uniquely assigns to each characteristic a specific configuration of code symbols; the number of symbols per code word must be constant. The advantage of the superimposed code is a substantially increased utilization of the code field which is equivalent to increasing the latter's capacity. The shortcoming of superimposed coding is the retrieval noise caused by false configurations likely to arise when two real codes are superimposed. For example, if codes 17 and 23 are recorded in the same code field, false configurations 12, 13, 72 and 73 will appear. Some of these may coincide with true code configurations, which will result in retrieval noise. The noise will be the greater the more codes are recorded in the same field and the less symbols every code word consists of. Typical codes used with edge-punched cards are shown in Fig. 30. Codes used with body-punched and machine-sorted cards are essentially similar.

It should be pointed out that edge-punched cards are recommended only for mechanized information retrieval in files of up to 10,000 documents in the case of manual sorting, and up to 30,000 documents with vibration or shock-type selectors. They are mainly applied to individual files in narrow fields of science and technology containing one or two thousand documents at most.

Body-punched cards are a modification of edge-punched cards and they are likewise intended for information retrieval systems with serial organization of AS. The code field of a body-punched card is shaped as a rectangular matrix of calibrated holes with equal spaces between the rows and the columns. A code position in a body-punched card is the partition between two adjacent holes either vertically or horizontally, depending on the card type. The coding of retrieval characteristics consists in cutting out a partition which results in a slot appearing in its place.

Searching is carried out in an unordered file. The cards are placed in the sorting box of a special selector. Sorting needles are inserted in the holes made in the box and in the cards, namely the holes bordering from above (or on the side) the required positions in the code field. Then the sorting box is tipped by 180° or 90°. All the cards having slots in the corresponding positions are moved out of the file to the extent of a slot. The box remaining in the tilted position, an auxiliary needle is inserted in a hole located at the uppermost (or sidemost) edge of the cards and the box. This needle connects the bulk of the file to the box except for the cards that have been moved out (their edges being below the hole through which the needle was inserted).

The selector box is then restored to its original position, but the auxiliary needle prevents the cards that have been moved out from dropping back into the file. The sorting needles can now be removed to take out of the file the cards whose code fields contain the sets of characteristics that constitute the search patterns of the required documents. The chief advantage of body-punched cards lies in their being better fitted for multi-aspect searching than edge-punched cards, that is for searching for several characteristics at a time. However, a special selector is needed with them. The working principle of body-punched cards is schematically shown in Fig. 31 (top).

Superimposed cards are used with inverted AS organization. This means that every descriptor has a corresponding card upon which identification numbers of all the documents having that descriptor in their search patterns are recorded. The recording is effected by punching a hole in the location whose coordinates indicate the document's call number. During retrieval, the cards concerning all the descriptors of the search request are extracted from the file, laid one upon another and inspected against a light source to spot the coinciding holes. The numbers indicated by the coinciding holes are the identification numbers of the documents sought. Thus, a superimposable card is in a way a modification of the Uniterm card permitting mechanized identification of the matching numbers in the cards selected for searching. The working principle of superimposable cards is schematically shown in Fig. 31 (bottom).

The main drawback of these cards is their limited information capacity which depends on the density of holes and the reduction of the space occupied by one hole. The card sizes vary between 187 x 82.5 mm and 445 x 445 mm, with capacities ranging from 400 to 40,000 positions. They are usually most efficient with files of up to 100,000 documents. To facilitate punching, a grid of coordinates is usually printed on cards, otherwise transparent stencils are applied. Punching or drilling devices are used to make the holes, and light screens are used to spot the matching holes. Superimposable cards are either filed according to strictly alphabetized descriptors or are sometimes provided with edge notches enabling them to be filed at random. It should be noted that superimposable cards, like Uniterm cards, can be used to build grammarless descriptor languages.

Superimposable cards were invented, as already mentioned, in 1915 by the American ornithologist H. Taylor who applied them to define different bird species by combinations of their characteristic features. It was, however, not until the 1940s that they were first applied to information retrieval. Superimposable cards are sometimes called optical coincidence cards, feature cards or peek-a-boo cards.

Mechanically sorted cards, which are a variety of body-punched cards, are of standard size - 187.4 x 82.5 mm. The face side of a machine card is covered with columns of figures printed over the whole field leaving free only a narrow horizontal band at the top. The figures constitute a matrix. The 45- , 80- and 90- column cards are the most common types in use. In contrast with hand-sorted cards, they have the left-hand upper corner cut off. Comparatively inexpensive computing and punching machines currently in production are used to handle machine cards. Systems based on this equipment can handle files of over 200,000 documents, and even several million documents, provided a preliminary ordering (subsorting) is done. The equipment is efficient, simple and reliable. Both serial and inverted organizations of AS may be applied to such mechanized IR systems. The former type of systems are mostly built around sorting machines, while the latter type are based on card-selecting equipment.

In the case of a serial organization of AS, every card carries an encoded search pattern of a document and its identification number. Some information systems have the bibliographic details of the document and even its annotation printed on every card. During search, the sorting machine scans the card file - or some part of it - matching the codes of document search patterns recorded on the cards against the search request code entered into it. When the adopted matching criterion is fulfilled, the sorter selects a relevant card and sends it to the receiver, since this card indicates the identification number of the document (and in some systems, other data as well, which permits regulation of the search by the intermediate retrieval results).

In the case of an inverted organization of the AS, every card carries the code of a descriptor and the identification number of every document, whose search pattern includes that descriptor. The rest of the code field may be used for a bibliographic entry and annotation which serve the same purpose as with serial storage organization. In the course of information retrieval, subfiles of cards corresponding to the descriptors of the search request are selected. Note that whenever standard card selecting equipment is used, a separate card is reserved for every document. If an ordered file is used, this selection is done manually. A sorting machine is used to select the required descriptor cards from an unordered file.

Then the selected card subfiles are arranged by their identification numbers and entered into a card-selecting machine, which selects out of the first pair of subfiles being processed only such cards which contain matching numbers. The resulting subfile is then matched against the third subfile, etc., until the cards with matching identification numbers have been selected from all the subfiles. These numbers indicate the documents sought. It should be mentioned that despite the fact that a card sorter has an operating speed of 250 cards per minute, the advantage of inverted over serial AS organization is practically reduced to zero by the need for repeated increase of the card file.

Modified card sorters are now under development that will be able to handle cards carrying dozens of identification numbers. This will cut down the file size and hence the retrieval time.

Thus far we have made only some brief remarks on the contents of the card space lying outside the code field. In some cards - edge-punched, body-punched, and mechanically sorted - this space may be sufficient to hold the identification number and bibliographic details of a document and even an informative abstract. Moreover, there exists a special type of card having a calibrated window (aperture) for inserting a microfilmed copy of a document. These cards are called *aperture cards*. They are also sometimes called Filmsort cards by the name of the company that first produced them on a commercial basis. With various types of cards the aperture size may be between 105 x 77 mm and 30 x 41 mm. Most widespread are mechanically sorted 80-column aperture punch cards with one frame of 35-mm non-perforated microfilm. Different types of aperture cards are shown in Fig. 32.

The chief advantage of aperture cards is an 80 per cent reduction of storage space used for real-size documents, as well as the facility for rapid search and retrieval of documents and for their reproduction. Usually these cards are used in single-circuit IR systems in which AS and PS are combined in a single file of unordered aperture cards. Such systems are particularly efficient for storing technical drawings - a case where problems of space and retrieval speed are especially acute: suffice it to say that in designing a jet aircraft tens of thousands of drawings are used weighing sometimes as much as the aircraft itself. An aperture card is a combination of a punched card and another information carrier, a microfiche.

Microfiches

Transparent microcards, or microfiches, are pieces of negative or positive photographic film carrying the images of documents or their parts, and should be distinguished from microcards made of an opaque material, usually photographic paper. They are of different sizes, but the currently most used size is 75 x 125 mm, that is the size of a standard library card. The upper horizontal band of a microfiche is reserved for the bibliographic reference readable with the naked eye. The document text is micro-reproduced (depending on the reduction ratio one microfiche can accomodate from 48 to 200 images) and can be read by means of special equipment. Specimens of microfiches are shown in Fig. 33.

The fault with single-circuit mechanically sorted, aperture-card IR systems is that the film is soon ruined through scratching during sorting For that reason systems with separate AS and PS are much more practical. It is possible to have AS implemented on normal mechanically sorted cards, with PS consisting of hand-sorted microfiches. The major merits of microfiches - simple filing, with easy storage and production of enlarged copies make them a most promising information carrier for PS in systems numbering millions of items.

Quite sophisticated IR systems with specially designed retrieval devices are developed on the basis of microfiches. An early system of this kind was *Filmorex* built by the French physician J. Samain in 1950; it has been considerably modified and improved since then. The information carrier used is a 35 x 60 mm microfiche, shown in Fig. 33. It is divided into two zones, one of which is for the micro-image of the document and the other one for its search pattern. The maximum code field capacity is 20 descriptors, one horizontal column per descriptor. The Filmorex system comprises the following hardware: a recording camera, a guillotine for cutting film into frames, a selector, a reader-printer, and card files - subject, auxiliary and author.

Special code cards are used to record document search patterns on microfiches. For every descriptor there is one such card carrying its five-digit number and the binary code for that number. The operator of the recording camera pulls out of the file the code cards of the descriptors of the document being processed and lays them one upon another in such a fashion that each subsequent card covers the underlying card but leaves open the edge with the binary code of the descriptor. Thus the code of the document search pattern is formed and it is photographed together with the document. The last (topmost) card remains entirely open, so that in addition to the binary code of that descriptor, its numeric, alphanumeric or alphabetic code is also photographed. This code is reproduced on a microfiche before the code field, and is easily readable. The basic advantages of this code-forming technique are its simplicity and the fact that it is almost completely error-free.

During input, each document is photographed as many times as there are descriptors in its search pattern; the last, uppermost code card is every time taken off and placed under the rest of the cards so that it again will become the first one. As the result, the code of the next descriptor is laid open and recorded on the next microfiche. Thus, every following card is marked with the code of the subsequent descriptor. The cards are then manually sorted according to these codes between the corresponding sections of the operational files. It takes an average of 2 minutes to input a document with a search pattern of 5 descriptors.

Search is performed with the help of a special selector. The search request is recorded in a complementary code, that is in such a way that the positions which are transparent in the basic descriptor code are black in the complementary code and vice versa. When superimposing a complementary code upon the corresponding basic code, the line of the code field becomes completely hidden. Cards can be searched by not more than three descriptors at a time. For that, a 'negative' mask of the codes of the descriptors is prepared; this mask is a piece of black film into which the complementary codes of the descriptors have been punched on a keyboard punch. The mask is entered into the selector and during the sorting of the cards is superimposed, as it were, on their code fields. If the mask completely blacks out the code field of a microcard this means that the code field contains the required descriptor codes. The coincidence of a code with its negative counterpart is established by three light cells that close the electrical circuit of the .sorter magnet which transports the microfiche to the selector holder. The output device of the system is a reader-printer which produces enlarged document copies ready for use. The selector can handle up to 600 cards per minute, and the reader-printer up to 40 cards per minute. The operation and the principal units of the Filmorex system are shown in Fig. 34. Its cost is in the order of $ 25,000. Several dozen such systems are currently operational in a number of countries.

Other, more complex systems based on transparent microcards have also been developed, for example Minicard, Magnacard, Walnut, with costs from hundreds of thousands to millions of dollars. However, only a few of these expensive systems exist and they have not become widespread. As a matter of fact, the growing popularity of microfiches is explained just by their being efficiently applicable without any complex and costly equipment.

Microfilm

The same is nearly as true of microfilm, which, however, is a kind of continuous information carrier. The division of material information carriers into discrete and continuous is as conventional as any other classification. Indeed, if we cut microfilm into small strips of several frames (called strip microfilm) they can be regarded as discrete carriers. For that matter, even roll films are continuous only within a roll. The continuous nature of a carrier creates certain difficulties for ordering the file of documents it stores. On the other hand, equipment used with continuous carriers ensures very high speeds of operation. The obvious advantage of microfilm is that it requires no special film-handling equipment since all the existing devices of the film industry are available.

Information retrieval of microfilmed documents, however, does require special *microfilm selectors*. The concept of such selectors was first suggested by the German E. Goldberg back in 1927. The first operational prototype microfilm selector 'Memex' was built in the late 1930s by V. Bush of the Massachusetts Institute of Technology whose words were quoted at the beginning of this chapter . The final version of the device, called 'Rapid Selector', was constructed by R. Shaw in 1949, since when it has been repeatedly modified.

The Rapid Selector handles 35-mm unperforated positive film. Recording of documents is carried out at reduction ratios from 2.5:1 to 20:1. The frame size is 29 x 44 mm; every frame on the film is preceded by a code field for a document search pattern. The capacity of the code field is 315 positions. The search pattern is entered into the code field from an 80-column card punched on a keyboard punch. The punched card is fed into the microfilming camera, and the punched holes are automatically converted into a binary light code which is projected on to the microfilm.

Search requests are likewise entered into the selector from 80-column punched cards. A switchboard panel is used to specify the logical relations between the search pattern descriptors. As the film moves, a special photoelectronic unit matches the search request against the search patterns of the documents whose microrecords are on the film scanned. The documents produced by the selector are copied onto the reproduction microfilm in a dynamic mode, that is without stopping the main film. At the film drawing speed of 2.7m/sec and a main reel capacity of 1,800 m, the average retrieval time in a reel, including development of the reproduction film, does not exceed 12 minutes.

A 'Poisk-OK' microfilm selector was developed in the Soviet Union in 1962 by engineer Yu. Ya. Klyachkin (VINITI) for searching micro-records of documents by their identification numbers. The information carrier used is again 35-mm unperforated positive film with documents filmed at a reduction of 15:1. Identification numbers are entered from a numeric keyboard located on the console of the filming device. The number is projected against a coding scale and recorded together with the document. On the film, the document (two pages) and its number are placed in a frame of 31 x 43.5 mm. The density of filmed records is 19 frames/m. The code field with a document's call number precedes every frame containing a copy of that document. 250-270 m reels are used, each containing up to 5,000 frames.

Requested numbers are keyed into the selector and the identified images enlarged to the original size are projected onto the display screen. The average retrieval time in one reel (holding up to 10,000 pages) at a film-moving speed of 0.6 m/min. does not exceed 4 minutes. There is a facility for stopping the film to make natural-size copies of the images from the display on usual paper: hence the name of this model - 'Poisk-OK' which means 'Searching with stop for copying'. A new model of this selector has now been developed, named 'Poisk-DV', which stands for 'Searching with dynamic copying'. While retaining the main features of the previous model, the new version permits the requested images to be copied onto 35-mm roll film without stopping the film drive. Automatic control of the selector is built in.

There exist more complex selectors with better performance characteristics. These include, among others, US-manufactured equipment used in IR systems - MIRACODE, RADIR, CRIS, FLIP. It should be noted, however, that microfilm selectors are mostly used as PS in IR systems with a relatively limited file. This explains the trend towards simplification of these selectors. Some recent popular models can even be regarded as ordinary reader-printers supplemented with elementary facilities for searching filmed images and a mechanical film drive.

Roll microfilm plus simple and inexpensive readers and copiers can well be used by individuals for building up personal scientific and technical libraries.

Electronic computers and their role in automated information retrieval

Of all the continuous information carriers listed above we have still not spoken of punched paper tape and magnetic tape, disks and drums. Punched tapes are much used in devices employed to compile various information publications, permuted-title indexes and union catalogues, and specific types of these devices will be discussed in the following chapter. Magnetic tape is used in some special-purpose information retrieval devices, for example in the Videofile system (USA, 1958-1964). These devices will in all probability have good prospects once television techniques have taken root in scientific information activities. At present, however, they are not much used because of their complex design and high costs, and for that reason we shall not dwell on them.

Continuous carriers of the above types are mainly used for information input and also as external memories of electronic computers, which are more and more regularly applied to information retrieval tasks. The design and operation of different types of computer would make the subject of a different and rather ample course. Nevertheless, discussing the mechanization and automation of information retrieval would be incomplete without assessing the part electronic computers play in this area. Successful applications in various fields of human activity have inspired a great number of people with what amounts almost to a mystical faith in their infallibility and omnipotence, though the real capabilities of computers are far from boundless.

The essential advantages of a computer over the human brain are uniformity of operations, reliability, minimal errors, high speed, and eventually lower cost of certain operations. All these advantages can be realized only provided there is an algorithm and a program for the performance of the task concerned. In contrast to humans, the computer will always give identical indexing for the same document, and it can be far faster and more exact in comparing an information request against the abstracts of documents or even their complete texts.

However, unless we understand the inner mechanism of such thinking processes as document content analysis, abstracting,, annotating and indexing, we shall not be able to algorithmize these processes. This, in turn, requires deep-going studies in the psychology of thought, linguistics, mathematical logic, semiotics and other disciplines. All attempts at using computers for information retrieval have been based on mere statistical properties of document texts and have not taken into account their semantic properties. But we are not even certain that the identity of meaning of two texts can be established by purely statistical methods without any semantic analysis of their contents involving recourse to extra-linguistic information, i.e. information not immediately contained in the texts being processed.

The greatest promise therefore appears to lie in finding the optimal modes of man-machine interaction so that their respective advantages may be utilized and support one another in the best possible way. As Norbert Wiener, the father of cybernetics, phrased it in his last work: "Render unto man the things which are man's and unto the computer the things which are the computer's. This would seem the intelligent policy to adopt when we employ men and computers together in common undertakings. It is a policy as far removed from that of the gadget worshipper as it is from the man who sees only blasphemy and the degradation of man in the use of any mechanical adjuvants whatever to thoughts". (1)

To recapitulate, we have outlined in general terms the technical means of information retrieval. In keeping with our chief objective - to introduce a student to informatics - we have concentrated on the aspects essential for mechanization and automation of information retrieval. As to particular information-retrieval media, the emphasis has been on those using punched cards, which can be considered the most generally available and reasonably efficient under certain conditions. We have spoken more briefly of the devices based on microrecords of document texts, despite the fact that these have a major part to play in mechanizing scientific information activities. To some extent this gap will be filled by the following chapter where the copying and reproduction of scientific documents will be dealt with as the starting point and essential prerequisite for the utilization of scientific information.

---

1. Wiener, N. God and Golem, Inc. A Comment on Certain Points where Cybernetics Impinges on Religion. Cambridge, The M.I.T. Press, 1964, p. 72-73.

Questions for self-checking

1. Under what conditions is it advisable to mechanize and automate scientific information activities?
2. What are the existing information carriers and how are they applied?
3. What are the applications of the different types of punched cards?
4. What are the advantages and shortcomings of the principal punched-card codes?
5. What are the principles of handling edge-notched, body-punched, and superimposable cards?
6. What are microfilm selectors used for?
7. How can computers contribute to the solution of information retrieval problems?

Literature

1. Artandi, S. An Introduction to Computers in Information Science. Metuchen, Scarecrow, 1968, VIII, 145 p.
2. Becker, J. and Hayes, R.M. Information Storage and Retrieval: Tools, Elements, Theories. New York, Wiley, 1963.
3. Introduction to Data Processing. "Library Resources & Technical Services", 1965, vol. 9, N. 1, p. 5-103.
4. Punched Cards. Their Applications to Science and Industry. Ed. by R.S. Casey, J.W. Perry, M.M. Berry, A. Kent. 2nd ed. New York, Reinhold, 1959.
5. Scheele, M. Punch-Card Methods in Research and Documentation (With special reference to biology). New York, Interscience Publishers, 1957.
6. Taube, M. Computers and Common Sense. The Myth of Thinking Machines. New York-London, Columbia University Press, 1961.

9. REPRODUCTION OF SCIENTIFIC DOCUMENTS

Throughout the entire course we have repeatedly stressed the fact that scientific documents are a major attribute of scientific information work. They are a mode of existence of science, for they are the final product of any research, serve as one of the primary means of dissemination of scientific information, and their utilization poses highly important organizational problems of scientific research.

For this reason, document reproduction - i.e. copying and duplicating - cannot be considered as a purely technical matter. The level of reproduction techniques and facilities has a direct bearing on the effective use of scientific information and, eventually, on the unhampered advancement of science. For centuries since Gutenberg's invention, the printing process remained the sole method of large circulation printing of textual documents. Photography, invented in the 1830s, has found extensive application as a document copying method only in recent decades.

Document reproduction processes

The need for rapid and low-cost copying techniques and facilities, which arose several decades ago in the sphere of clerical and office work, has led to the development of a variety of copying and duplicating processes. In addition to the early silver halide and diazo processes (i.e. photography and dyeline copying), the thermographic and electrographic processes, which first appeared in the early 1950s, have quickly made their way into the document reproduction field. It is now beyond any doubt that the swift spread of new reproduction techniques will play as great a role in increasing the efficiency of scientific research as the invention of book printing did five hundred years ago. The theoretical and practical aspects of these processes are dealt with by a new technical discipline which has been called *reprography*. Knowledge of the essentials of reprography is important both for information officers and for subject specialists in the different areas of research and practical work. The development and use of the various information retrieval devices discussed in the preceding chapter only becomes possible through the application of the latest copying and duplication techniques.

Document reproduction processes can be broadly classified as follows:

Document copying
- Silver halide processes
  - Photographic
  - Reflex
- Diazo processes (dyeline copying)
- Thermographic processes
- Electrographic processes

Document duplication
- Typographic printing
  - Letterpress
  - Gravure
  - Flat-bed (lithographic)
- Spirit duplicating
- Stencil duplicating
- Offset duplicating
- Use of addressing machines

Division of all document reproduction processes into copying and duplicating is for the sake of convenience only; the criteria of division are the number of copies produced and the use of a specially prepared master. Copying is generally held to include those processes which yield a maximum of 12 or 15 copies from any original which has not been specially prepared for this purpose. But, then, many copying processes permit the production of more than 15 copies, and also make certain demands on the originals being reproduced.

Photography

Copying of documents by photographic means dates back to the invention of photography. As far back as 1839, when J.M. Daguerre for the first time demonstrated his photographic process to the members of the French Académie des Sciences, Albrecht Breyer, a medical student at the University of Liège, began experimenting with silver halide-coated paper to produce copies from printed book pages. Nowadays, standard photographic processes are used to obtain high-quality copies able to withstand prolonged archival storage. These processes, however, are very costly and time-consuming and require specially equipped darkrooms. Standard photographic processes are mainly used in scientific information work for producing half-tone copies of illustrations with further multiple duplication (by the photostatic method), and for producing reduced copies (microcopying) or enlarged copies (reproduction copying).

Production of greatly reduced copies of documents is called microphotography or microfilming. *Microfilming* (with reduction ratios of from 8 to 200) permits the reduction in size of the original documents by as much as 95%, standardization of the document size, and prompt execution of copying and small-run duplicating jobs. Microcopy negatives serve as starting material for the 'real-size' reproduction of originals. Microcopy positives are used directly for reading with the aid of microfilm readers.

Microfilm readers may have reflexive or translucent screens, and many stationary, desk-type and portable readers offering a magnification range from 8x to 200x are produced. In most of them microphotographs are read against a light source. To obtain enlarged copies of microfilm, both standard enlargers and special electrographic devices may be used. Increasing use is made of the reader-printer devices which make it possible, after reading a microfilm in roll or microcard form, to obtain immediately an enlarged copy on special paper.

Reflex processes

To overcome the difficulties attending standard photographic processes, to facilitate silver halide copying processes and to eliminate the need for darkrooms, reflex processes have been developed. They are based on the method of contact copying, where the exposure of the sensitized paper is effected by the light reflected from the surface of the original. The most widespread of these are the stabilization, diffusion-transfer-reversal and gelatin-dye-transfer processes.

In the *stabilization process*, silver halide-coated paper is laid with the coating on the original to be copied and is exposed to light from the side of the copy paper. The light passing through the sensitized material is absorbed by the dark areas of the original and reflected from the light areas back onto the coating. This results in the reflex exposure of the copy paper. The paper is dipped in an activating solution which converts the exposed silver halide to metallic silver in the areas corresponding to the non-image portions of the original. After development, the copy paper is put into a stabilizing solution (thiourea) whose constituents react with the unexposed silver halide to form complex compounds insensitive to light. The next step is the squeezing and drying of the paper. The first reflex copy obtained is a negative; using it as an original, further reruns produce any desired number of positive copies. It should be noted that the need to dry the original makes the stabilization process too slow (though less slow than standard photography which requires time-consuming washing after fixing).

The *diffusion-transfer-reversal (DTR) process* is much faster, but requires two types of special paper, negative and positive. The negative paper is coated with a gelatin emulsion containing silver halide and other light sensitive substances; the positive paper has a coating of colloidal silver and is insensitive to light. The negative paper is exposed by the reflex method, as in the previous process, and following exposure, removed from the original and immediately, without preliminary development, placed face to face with the positive paper. The two are immersed in a solution containing a developer (e.g. methanol and hydroquinone) and a solvent for silver halide (e.g. sodium thiosulphate). In the areas of the latent image of the negative, soluble silver salts are formed and transferred by diffusion to the positive paper where under the effect of colloidal silver they form metallic silver. The positive image is thus produced. DTR is the quickest of the reflex processes, but has some drawbacks: notably, one negative can produce only one copy on which brownish spots may appear and which may become yellow with time.

The *gelatin-dye-transfer process* (known under the trade name Verifax) depends on the use of a printing matrix, which is a sheet of light sensitive paper coated with a gelatin emulsion containing silver halide, a non-diffusing developing agent and a dye-forming agent. The matrix is exposed by the standard reflex method and then placed in a solution of alkali known as the activator solution which has the effect of hardening the gelatin in the lighted areas, while the dye-forming agent converts to insoluble blue dye. This dye can be transferred onto a sheet of any paper together with the unhardened gelatin in which it is contained. The transferred gelatin contains not only dye but also silver halide which under the action of thiourea (with which the paper has been previously impregnated) gradually converts to metallic silver. After the dye has faded out, this metallic silver produces the copy image. One matrix will yield up to seven copies. Since the sensitivity to light of the gelatin coating on the matrix is fairly low, copying can be carried out in a room with subdued fluorescent lighting. The matrix is sensitive, however, to yellow; as in other reflex processes, yellow, orange and green originals are very difficult to copy. Moreover, the gelatin-dye-transfer process does not produce quality copies from half-tone originals and requires specially trained operators. The general diagram of the reflex and other copying processes is shown in Fig. 35.

Diazo

The diazo process, also called dyeline and diazotype, is based on the fact that many diazo compounds react with aromatic amines, phenols and some other compounds to form azo dyes. Diazo compounds are decomposed, however, when exposed to ultraviolet light and become incapable of forming azo dyes. For copying, an original is placed on diazo paper or film and exposed to an ultraviolet light source from the side of the original. This results in the decomposition of the diazo compound in all areas of the coating which are not shielded by the elements of the image being copied. The ability to form azo dyes is retained by those areas of the diazo coating which correspond to the elements of the image being copied. If, therefore, such a paper or film is subjected to appropriate chemical treatment, a positive copy of the original will appear on it.

According to the method of development, diazo processes may be dry or wet. In the dry method, the diazo coating contains azo components which react with the undecomposed diazo compound to produce the corresponding dye. In addition, such a diazo coating contains an acid preventing premature coupling of the diazo compound with the azo component. The exposed diazo coating is brought into contact with ammonia vapour which neutralizes the acid, resulting in the necessary reaction which yields the dye. In the wet method, the light-sensitive coating contains only a diazo compound: because of this, the diazo paper or film is treated with an alkaline solution containing a specific azo component.

The dyeline process is used primarily for the duplication of engineering drawings. It has come into wide use since the middle of the last cent after John Herschel, a British chemist and astronomer, invented a light-sensitive paper which was originally cyanotype. Diazocopies retain shar image for some 25 years, and their cost is four or five times lower than that of copies produced by any other method. The copied original, however must be on one side of a transparent or translucent material.

Thermography

Thermo-Fax, the most popular of the thermographic processes, is direct, i.e. it involves no image transfer. It was introduced in 1950 by the Minnesota Mining and Manufacturing Co. (3 M) as a result of investigations by a physical chemist, Dr. Carl S. Muller. This process makes use of heat-sensitive paper containing compounds that form a coloured substance when heated. The reaction usually involves a ferric compound, such as ferric stearite, and a phenolic compound, such as tannic acid; these materials are held in a binder, such as polyvinyl butyral. To prevent premature reaction between these compounds, a waxy ferric compound is used. The reaction which forms a coloured product takes place only after the ferric compound has melted.

For copying, heat-sensitive paper is placed face up on the surface of the original to be copied and is exposed to infrared radiation from the face side of the copy paper. Infrared rays pass through the heat-sensitive paper and, being absorbed by the printed areas of the original, generate heat that is transmitted to the areas of the copy paper with which they have contact. This absorption of infrared rays occurs only when the image being copied is printed in inks containing carbon black or metallic compound;

organic inks (of vegetable origin) do not absorb infrared rays. As a result of heating, the ferric compound is converted to a melted form and, reacting with the phenolic compound, produces a positive copy of the original on the heat-sensitive paper.

Thermography is applied to the copying of printed or typed originals, and of pencil inscriptions and drawings (pen and ball-point inscriptions cannot be copied), where it is necessary to produce copies for temporary use. This is an extremely easy and cheap process, but the copies it produces will become blackened and brittle especially if exposed to heat. There are also some indirect thermographic processes, e.g. the Dual Spectrum process, or the distillation process, which overcome the disadvantages of thermography, but at the same time slow down copying and render it much costlier and more complicated.

Electrography

Electrography is a name which is commonly applied to the various photoelectric and electromagnetic processes for forming images and printing without impression. As early as 1921, a Russian inventor, E.E. Gorin, proposed to use the properties of photoconductors for image reproduction and filed an application for the invention of what he called an 'Electrographic apparatus': he was apparently the first to use the term 'electrography'. The American physicist C.F. Carlson combined the principle of image forming which had been proposed by Gorin with the principle of latent electrostatic image development by the sprinkling of a dye (proposed by P. Selenyi, a Hungarian) and in 1938 obtained the first electrographic copy. This process was subsequently perfected at the Battelle Memorial Institute (USA) and re-named "Xerography" (from the Greek words meaning 'dry writing'). Commercial exploitation of electrography began in 1950. Electrography consists of processes involving image transfer (Xerography) and those without image transfer (Electrofax). Both are presented schematically in Fig. 35.

Xerography works in the following way. A selenium coating on an aluminium plate is used as photoconductor. To make the plate light-sensitive it is given a positive electrostatic charge by a corona discharge in the dark. During exposure of a charged plate in a reproducing or projecting device, the lighted areas of the selenium coating are depolarized. This results in the formation of a latent electrostatic image on the surface of the photoconductor coating, for the development of which an oppositely charged toner is cascaded across the plate. The toner particles are attracted by the charged areas of the coating and held by them permanently. Fine synthetic or natural resin powder with the addition of a pigment serves as the toner, which is electricized by so-called carriers (small glass or polystyrene beads).

A sheet of common paper is then laid over the powder image thus formed and it is given an electric charge by means of a corona discharge of the same sign as the charge of the latent image and opposite to the charge of the powder image. As a result, the toner particles are attracted to the paper surface and settle on it. The powder image is fixed on the paper surface either by heating the toner to its melting point or by exposing it to the vapours of a strong solvent, such as acetone or sulphuric ether.

An electrographic copy does not take more than 3-4 minutes to produce, even when all the operations are manual. At present, fully automatic rotary-type devices are used, in which the selenium coating is applied to a rotary drum and not to a plate. Such devices have high operating speeds and enable a copy to be produced in a few seconds. Xerography uses a dry method which requires no darkroom, and the materials are readily available and fairly inexpensive. An image can be copied not only on paper, but also on fabric, films and even on metal masters which are used in offset duplicating. The main drawback of Xerography is insufficient quality in reproducing half-tone images and originals with large dark areas. Moreover, the equipment - particularly for the rotary automatic machines - is very expensive. A selenium-coated plate can produce not more than 2,000 copies, and a drum must be replaced after about 50,000 copies.

The *Electrofax method*, or direct electrostatic process, was introduced by the Radio Corporation of America (USA) in 1954. It makes use of photoconductor paper coated with zinc oxide on which electrographic copies of the original are directly formed. Electrographic paper is produced by applying to a paper sheet a coating of a dielectrical binder, such as silicone resin or polystyrene, and organic dyes that sensitize zinc oxide to light. These properties of zinc oxide were discovered by the soviet researchers Academician A.N. Terenin and E.K. Putseyko. Electrostatic paper is charged and exposed in the same manner as in the xerographic process. The only distinction is that here the paper is given a positive charge and the pigment a negative one.

A visible image is produced on electrographic paper by either a dry or wet method. Dry development makes use of the 'magnetic brush' technique. The developing powder contains the carrier in the shape of fine iron particles which, together with the charged pigmented particles that cling to them, are attracted by an electric magnet forming a sort of 'brush'. As this brush moves across the electrographic paper with its latent image, the pigmented powder adheres to the oppositely charged areas of the photoconductor coating and develops this image. The iron particles remain on the magnet. This method makes a fine job of line drawings and half-tone images, as well as of dark images with large solid areas. The powder image is fixed on the electrographic paper by heating the pigment to its melting point. Liquid development is either by dipping the paper in the developer or by spraying the developer on the paper. The developer is usually a high-specific-resistance liquid (e.g. aircraft gasoline) in which an appropriate pigment is dispersed. It requires no further fixing of the image on the copy.

It should be noted that the Electrofax method permits the designing of high-speed automatic copying equipment and dispenses with restrictions on the number of originals that can be copied, but it requires special paper from which the images can be easily scratched off by metal objects. In addition to the electrographic processes discussed, a few others are used, including the electrolytic and thermoplastic processes. We have considered only the basic types, whose relative merits and demerits are tabulated in Fig. 36.

It was not our intention to discuss specific copying devices; such equipment includes thousands of models varying in complexity from the simplest, designed for manual operation, all the way to the most sophisticated machines designed for semi-automatic or fully automatic operation. Directly governed by their design and application are their speed, labour requirements and cost. Apart from that, in this new and fast growing field new equipment is constantly entering the market, and it is certain that by the time you complete your course, the existing equipment will have been replaced by new equipment having better performance characteristics. This explains why our introductory course is confined to giving a general survey of the new copying processes. Once you understand these, you have the key to any devices based on them. Concluding this survey, we should like only to point out that electrography is becoming the most widespread of all the processes discussed, and seems to be the most promising of all existing document reproduction methods.

Typographic printing

It has already been said that duplication is commonly understood to include all processes that produce more than 15 copies from a special master; the principal traditional method is typographic printing, which has been called classical printing, or graphic arts, as distinct from the more modern duplication techniques.

In the fourth chapter, speaking about the channels of dissemination of scientific information, we touched upon the work of a publishing office and a printing shop. We discussed only the process of preparing a forme for relief printing. In fact, three kinds of printing can be distinguished: letter-press (relief), gravure (intaglio), and flat-bed (planographic). In *letterpress printing*, the printing elements are raised above the rest of the forme surface, the forme in this case consisting of a set, which reproduces the text, and plates which reproduce the illustrations. This printing method is used in the production of the great majority of books, periodicals, newspapers and other predominantly textual material. In *gravure printing*, the printing elements are deepened with respect to the white space elements. The varying depths of the printing elements make it possible to reproduce gradual transition from the lighter to the darker areas on the original, which strongly recommends gravure printing for reproducing art illustrations. In *flat-bed printing*, all printing and non-printing elements are on the same plane. This kind of printing (the commonest is called lithography) is generally used for multi-colour printing of maps, posters, post-cards, etc. Different kinds of printing are schematically shown in Fig. 37.

A forme is used in the printing of a run, i.e. in the production of a desired number of copies. There are two techniques of doing this: with the first technique, the transfer of the ink from the forme occurs at the very moment when the paper is pressed against it; with the second technique, which is called *offset printing*, the ink from the forme is transferred first to an intermediary roller, from which it is transferred on to the paper.

Modern printing shop equipment includes platen, cylinder and rotary machines. In *platen machines* the forme and the surface which presses the paper (platen) are plane surfaces. In *cylinder presses*, only the forme, to which the paper is pressed by the impression cylinder, has a plane surface. In *rotary machines*, both the forme and the surface, which presses against the paper, are cylindrical. Substitution of cylinders for plane surfaces provides for a substantial speeding up of the printing machines. The working principles of various types of printing machinery are demonstrated in Fig. 38.

Office reproduction techniques

The demand for facilities permitting big-run duplication of documents - but faster, cheaper and easier than the conventional graphic arts methods - has caused the emergence of a variety of processes broadly termed 'office reproduction techniques'. The processes include the means of non-composition master production, flat-bed offset, stencil and spirit duplicating, as well as advanced methods of editorial document processing. The emergence of office reproduction is a breakthrough in the information communication techniques comparable in impact with the invention of book printing. The merits of office reproduction - low cost, high speed, small size and general availability - permit almost the same quantitative and qualitative results to be achieved as in graphic arts, with less labour and materials and less time spent.

The advantages of office reproduction techniques over graphic arts are not confined to reduction in the production cycle, and hence publication time, on account of the proof-reading step. Authors' and editorial corrections of the forme are the essential steps in any typographic process, as they considerably improve the content of the published documents. However, office reproduction facilitates and speeds up the correction processes by using new means of master preparation, which constitute the essential features of the new facilities and techniques of document duplicating.

Composing typewriters and phototypesetters

The majority of modern machines for the preparation of masters for office reproduction are traceable to ordinary *typewriters*, commercial manufacture of which was started a century ago in 1868 by the American firm Remington. The typewriter design proposed by engineers K. Sholles and K. Glidden allowed printing in single-size letters. In 1878, typewriter design was improved to accommodate both upper- and lower-case letters. Further perfections, including the introduction of the tabulator and the interline space-regulator, were introduced by another American firm, Underwood Typewriter Co., founded in 1895.

Owing to the introduction of carbon paper in the late 1860s and early 1870s, typewriters began to be used for duplicating textual material. In 1873, T.A. Edison invented a method of stencilling using a wax stencil; this marked the beginning of the use of typewriters as a means of master preparation. From the time the first typewriters appeared on the market up till the present day their designs have been undergoing continuous changes. One of the most promising trends in this respect is the development of electric and automatic portable composing machines for master preparation. The first such machine was apparently the Skoropechatnik (literally meaning "Fast-Typer"), designed in 1873 by a Russian inventor, M.I. Alisov, and exhibited three years later in Philadelphia, where it was awarded a medal.

Today's *composing typewriters* are distinguished by the great variety of type sizes, sets of type and styles of type-face, which closely resemble print type. These machines permit differential spacing according to the width of individual characters, whereas in ordinary typewriters the spacing is the same. Many composing machines provide line justification, which means that the right-hand edge of the printing master is an even line. The Varityper, which is one of the widely used machines of this kind, permits the composition of complex scientific and technical texts using mathematical, chemical and astronomical symbols; it belongs to the range of segment machines which have distinct advantages over ordinary machines. The characters are fixed on a changeable segment which is positioned on the type drum. Pressing a key rotates the type drum until it faces the paper with the corresponding character, after which the impression takes place.

The carbon-paper ribbon, which can be used only once, produces a sharp and clear image suitable for the immediate reproduction or preparation of offset masters. The Varityper can be operated at a speed of about 40,000 characters of mixed setting per working day, but with line justification nearly twice slower, because each text has to be typed twice: during the first setting, all the word spaces in a line are added up, and during the second setting the characters are spaced out and the line is thereby justified.

The Flexowriter composing typewriter produces simultaneously both a typescript copy which can be used for the preparation of a printing master, and a punched paper-tape which can be used for the second - automatic - retyping of the composed text. This tape can also serve as input to a computer. Because of this, Flexowriters are sometimes used as input devices with punched card equipment and electronic computers. A Flexowriter tape can be used for the preparation of permuted-title indexes, union catalogues and other information publications. The Justowriter composing machine does semi-automatic typesetting with justification without re-typing. The Justowriter consists of two units called the Recorder and the Reproducer. The Recorder is a composing machine which produces both a typed copy in galley form and a punched tape complete with justification codes. The tape is inserted into the Reproducer which interpretes it and automatically produces a justified copy on ordinary paper or on an offset master.

Like composing typewriters, *phototypesetting machines* have been increasingly used for the preparation of printing masters. These machines use a variety of type-face styles and signs of different alphabets and have great potentialities for changing type size. During composition, characters are projected onto a photographic paper or film, producing a negative or positive image of the text which, after proof reading, correction and make-up, is copied onto a printing master. Wellknown phototypesetter models are: Photon, Linofilm, Digiset and some others. The trend in the development of equipment for the preparation of printing masters and of office reproduction techniques generally is towards high-speed electronic printing devices permitting a direct output of textual and graphical information from computers.

The equipment described above enables us to obtain only the starting original for the subsequent preparation of a master. Methods of master preparation vary with the type of printing used, viz. spirit, stencil or offset.

Spirit duplicating

In the preparation of masters for spirit duplicating, paper or other moisture-resistant material is brought into contact with a special carbon-paper or ribbon covered with an alcohol-soluble dye. As the image is applied to the original, a laterally reversed image appears on the opposite side under the effect of the dye.

Spirit duplicating machines are called *hectographs*, the working principle is shown in Fig. 39 (top). The main parts of a hectograph are the impression drum (cylinder) on which the printing master is mounted, the pressure and leader rollers, the paper feeding and moistening devices. The paper feeding device feeds paper sheets one at a time into the machine. There the paper surface is uniformly damped by ethyl alcohol which dissolves the dye on the reverse side of the original serving as master, and transfers a small amount of the dye onto the imprint. The best imprints are obtained on writing paper which has optimal absorbency; use of other paper types will result in blurring and spreading of the image. By adjusting the damping rate and the pressure exerted upon the paper at the moment of contact with the master, it is possible to vary the brightness of the image and the number of imprints which can be obtained from one printing master. This number will vary within the range of 100 to 300 imprints of acceptable quality.

Hectographs are available with replaceable stamping mechanism which makes it possible to obtain imprints from masters in which modifications, additions, deletions and substitutions have been made. They are used for the production of multiple copies of forms to be filled out in different columns, e.g. order, accounting and invoice forms. These hectographs alternate printing from the fixed image master with that from the stamping mechanism which contains data varying for each imprint.

Spirit duplicating is the only office reproduction process which can produce copies in up to six colours in one operation. This is due to the use of carbon paper with different colour dyes. The most bright and permanent image is provided by the crystal violet dyestuff. The imprints produced by spirit duplicating will fade with continued exposure to light and are also rather soluble in water. But these disadvantages are offset by the cheapness of the process and the fact that it does not require complex equipment and specially trained staff.

Stencil duplicating

The method of stencil duplicating consists in forcing a dye through a printing master, or stencil, onto a paper surface. Preparation of a stencil involves the formation of printing elements in the shape of minute holes in the master through which the dye penetrates, and the formation of the white space elements which are impervious to the dye.

Stencil duplicating is done by means of an office-type rotary stencil duplicator. It consists of a drum (cylinder) made of metal mesh and covered with fabric on the outside; on its outer surface is mounted the master, inside the drum is the inking mechanism. As the drum rotates the ink-roller absorbs ink from a bath and transfers it to the surface of the ink-distributing roller from which it is squeezed through the drum perforations onto the inking fabric and further onto copy paper. The imprint, which appears as the paper passes between the rotating drum and the impression-roller, is caused by the forcing of ink through the holes in the master. The working principle of a rotary press is shown in Fig. 39 (centre).

A stencil master, commonly called a 'wax' stencil, is a sheet of long fibre tissue impregnated with paraffin, wax or other ink-impervious substance. This is inserted in a typewriter, and typing is performed without ribbon in order to displace or remove the ink-impervious coating. Because of the simplicity of the working cycle, and the low capital and operating costs of the equipment, stencil duplicating has found wide application in the production of small runs. However, this kind of printing has some shortcomings; notably, it cannot cope with illustrative material, wax stencils will quickly deteriorate when exposed to heat, and the quality of text in the imprints is often of rather inferior quality.

Improvement in the quality of stencil duplicating is achieved by the use of the electric spark method of preparing the master. An electronic copying device permits stencils to be produced from any printed, typed or inscribed original. The device has two drums rotating at the same speed; on one the original is placed, on the other the stencil. The light cell of the scanning device recognizes the textual images which are converted to high-frequency current and transmitted onto a special conductor rotofilm. The film is eroded by electric sparks according to the tones of the areas being scanned. The stencil thus produced allows runs of up to several thousand copies to be made. Electrically-driven rotary presses with automatic feeding devices produce up to 4,000 imprints per hour. It is a feature of this process that large quantities of ink are needed to produce an image, as well as an absorbent paper which is usually the notepaper or at least offset-paper type. The copies are better in appearance and last longer than those produced by spirit duplicating, but their cost is somewhat higher.

Offset lithography

Offset duplicating, as already mentioned, is a long-established process in the graphic arts. It is based on the transfer of ink from the printing master to the paper through an intermediary elastic surface (rubber roller). Use of this method results in reduced wear of the master and increased printing speeds. Offset duplicating in conjunction with flat-bed printing is favoured in office [illegible] it offers the possibility of quickly producing relatively [illegible] thousand copies) with comparatively easy preparation [illegible] printing master.

Office reproduction makes use of simplified, small-size printing presses called *rotaprints*. A rotaprint has three cylinders of equal diameter: the upper is the master cylinder to which the master is fastened; below are the blanket (offset) cylinder covered with a rubberized fabric, and the impression cylinder. Immediately adjacent to the master cylinder are wetting and inking devices with regulable supplies of water and ink. Paper from the feeding tray is automatically fed into the machine between the blanket and impression cylinders. The impression cylinder presses the paper against the rubberized surface of the blanket cylinder and the ink is transferred from the rubber onto the paper. The imprint thus produced is automatically stacked on the receiving tray. The printing rate can be adjusted within the range of 1,000 to 6,000 copies per hour. The working principle of the rotaprint is shown schematically in Fig. 39 (bottom).

Offset masters are made on aluminium foil which can be used several times. After the first run, however, the foil must be subjected to graining and complex chemical treatment. For printing small runs, water-absorbent paper masters are used, which are much cheaper and easier to prepare. Electrographic methods of offset-master production are particularly popular.

Offset duplicating can be included with the most expensive and complex processes. The price of a rotaprint greatly exceeds that of other duplicating equipment, and its servicing requires skilled operators. Offset paper must be moisture- and dust-resistant, durable and white, while offset inks must provide increased brightness. At the same time, however, it is the most perfect of all document duplication methods used in office work and finds application where there is a steady demand for high quality multiple copies.

Duplicating by addressing machines

Addressing machines (addressographs) are used primarily for printing the names and addresses of subscribers to periodical publications. They allow repeated reproduction in a specified sequence of a certain set of data. They can, however, also be used for the duplication of certain information materials.

Addressing machines may be based either on the spirit or stencil duplicating techniques, but the predominant method used is printing by embossing (a variety of non-composable letterpress). The masters are prepared by indenting characters in a plastic, fibre-board or metal plate. The laterally reversed raised text thus produced is imprinted on paper through an inked ribbon.

The advantages of duplication using addressing machines are the long life of the imprints, possibility of using any kind of paper, and practically limitless runs. But the primary application of these machines determines their deficiencies which are the small size of the master, containing only a few lines of text, inability to reproduce illustrative materials, low printing speed, and labour-consuming preparation of the master. With all these properties, addressing machines commend themselves for use in libraries for duplicating catalogue cards.

We have now considered the major techniques and facilities for copying and duplicating documents. A comparative table of their advantages and shortcomings is given in F 36. As regards selecting particular means of duplication, one can be guided by the table given in Fig. 40. We made it clear at the beginning of this chapter, that we intended to deal only with the essentials of reprography, the technical science which deals with the theoretical and methodological aspects,of document reproduction processes. If, as a result of this discussion, you have formed some idea of the new potentialities in this field, we shall have fulfilled our task. A knack in handling reprographic equipment comes through prolonged special training. But even this broad survey may help you to make more effective use of scientific information, and that, after all, is the ultimate purpose of this Guide, and why the next (and last) chapter will be devoted to the use of scientific information.

Questions for self-checking

1. What are the basic document copying processes?
2. What are the relative merits and demerits of the different copying methods?
3. What is the distinction between graphic arts and office reproduction techniques?
4. What methods of duplicating are used in office work?
5. What are the criteria for choosing the different methods?
6. What is the main impact of the development of new document reproduction techniques?

Literature

1. Bartlett, H.A. A Guide to Reproductive Processes. "Office Methods and Machines", 1965, vol. 12, 142 pp.
2. The Revolution of Office Copying. Special Report. "Chemical and Engineering News", 1964, vol. 42, No. 28-29.
3. Schaffert, R.M. Electrophotography. London-New York, The Focal Press, 1965.
4. Verry, H.R. Document Copying and Reproduction Processes. London, Fountain Press, 1960.
5. Verry, H.R. Microcopying Methods. London-New York, Focal Press, 1963.

10. USE OF SCIENTIFIC INFORMATION

This closing chapter will be devoted to the most difficult - and the least explored - problem of informatics, namely the use of scientific information, which is the ultimate goal of all scientific information activities. This problem in a way brings to a focus all the foregoing discussions since, after all, the diverse primary and secondary documents, the most complex information retrieval techniques, and the media and facilities used for it, all serve the common purpose of making the use of scientific information as efficient as possible. Accordingly, we shall have to come back to some of the points raised at the beginning of the present guide so as to identify the who, when, how and which of scientific information, which information is used, what kind of activity information work should be, and what future prospects there are for all these aspects of information work.

The role of scientific information activities

Scientific information work has an ever-growing role to play in science. It serves essentially to raise the efficiency of science, which is ever more obviously acting as a direct productive force in modern society, accounting for 3 per cent of the GNP of advanced countries and almost 3 million employees all over the world. The number of scientists and other specialists engaged in scientific information activities and the money invested in these have also been growing rapidly. In the Soviet Union alone, more than 100,000 persons, directly or indirectly, participate in these activities. Over $ 5,000,000,000 is spent annually on information work in the world. Hence the importance of ensuring the effectiveness of that work in all its aspects - practical, organizational, and theoretical.

We define *scientific information work* as an institutionally organized variety of scientific labour which is performed with the object of increasing the efficiency of research and development, which consists in collecting, analytical synthetic processing, storing and retrieval of recorded scientific information, and which supplies this information when needed to research and practical workers in a form convenient to them. So far we have been dealing with the processes enumerated in the opening part of this definition. But the utilization process mentioned in its second part deserves just as much attention.

It is no exaggeration to say that the biggest obstacle to increasing the efficiency of scientific information work is the ignorance on the part of the scientist and practitioner of the opportunities it offers him and his inability to use these opportunities. Science itself and the system for dissemination of scientific knowledge have been taking on a mass-scale, one could say, industrial nature, but the mentality of scientists in using scientific information still reflects earlier stages of the development of science when the seeking and initial processing of information was not a social activity but the personal business of every individual scientist.

A convincing analysis of this contradiction has been made by a French information scientist, M. Menou, who writes: "The most important characteristic of our times, in so far as knowledge transfer is concerned does not seem to us to be the increase in the number and the power of human knowledge but rather the ratio and the hiatus between our mental structures, so deeply marked by millenia of 'atechnical' behaviour, and the extent of the objects that they must enable us to handle. Even if they do not always have distinct special characteristics, one may find manifestations and consequences of these prehistoric structures both in scientific and technical activity and in daily behaviour, since being a scientist does not mean that one ceases to be a man". (1)

Main groups of users of scientific information

In order to learn to overcome this psychological barrier, one must get a clear - if only general - notion of the main groups of users of scientific information and of the diversity of their information needs. Three user groups are distinguishable according to the kind of activity they are engaged in:

a) scientists (research workers) mainly engaged in basic and experimental research in the fundamental sciences;
b) engineers (practical specialists) engaged in experimental design, projecting, and operational activities in the various fields of technology and industry; and
c) managers engaged in the spheres of science, technology and the national economy.

Although there is no clearcut borderline between science and technology, there does exist an essential difference between these two social activities that determines the specific information needs of scientists and of engineers. Science is an activity aimed at the production of knowledge with the cognition of nature for its object. The product of science is knowledge, or scientific information recorded in scientific documents. Technology is an activity aimed at the production of useful objects, of material wealth, and its task is the direct transformation of nature on the basis of knowledge gained by science.

As D. Price wrote: "If, when a man labours, the main outcome of his research is knowledge, something that has to be published open for a claim to be made, then he has done science. If, on the other hand the product of his labor is primarily a thing, a chemical, a process, something to be bought and sold, then he has done technology". (2) A researcher strives to learn the unknown, to discover a logical interrelation, an inner order in a seeming chaos of facts. If he knew exactly what knowledge is needed to solve the problem facing him, he would have had much of its solution already in his hand. A scientist can therefore never rely entirely on an information service for getting all the information he needs; the greatest task he can entrust to the information service is to select for him the scientific documents that are likely to be of relevance to the subject of his concern.

---

1. Menou, M. Knowledge Transfer from Autoconsumption to Mass Production. In: On Theoretical Problems of Informatics (FID 435). Moscow, 1969, p. 44.
2. Price, D. The Difference between Science and Technology. Edison birthday lecture. Dallas, 10th February 1968, p. 11.

As a rule, a scientist prefers to scan or peruse these documents himself, for in the process he gains a better understanding of his own information needs, comes across valuable associations, guesses and formulates hypotheses. For this reason, of all information services discussed above, most preferable for scientists appear to be: current-awareness information, abstract services, selective dissemination, and retrospective searching.

Entirely different are the information needs of the engineer, who usually knows more or less exactly what kind of knowledge he needs to solve his problem, is able to state clearly his information needs and consequently expects a clear and concrete answer from the information service. Reluctant to spend time on browsing, he is eager to entrust this work to the information service. Information service to engineers thus requires - first and foremost - reference books, analytical reviews, and special services extracting from scientific and technical literature and unpublished documents information they need. It was this circumstance that lead to the establishment, over and above conventional information units operating in technology and other applied fields, of specialized services called 'data banks' and 'information analysis centres'. Over a hundred such services are at present operational in the United States alone.

Information service to managers is again an altogether different matter, and the problem is complicated by the fact that the term 'management' itself is too broad. We are concerned not with the processes of management and control in general, in the cybernetic sense, but only with the particular processes of managing social systems, and primarily science, technology, industry, and economy. In that context, management can be viewed as the process of making decisions on the basis of certain available information, or as motivated choice of a course of action out of a range of possibilities. True, any purposeful human activity may be viewed as a decision-making sequence, but organizational decisions have their important peculiar features. A manager usually takes his decisions under the conditions of an acute shortage of necessary information and lack of time to process it, which results in the decisions being to some extent arbitrary. The higher the level of management, the greater the information and time deficits are and the more creative management is. If a manager had absolutely exhaustive information and enough time to consider it, the decision-making would not be a creative act, but merely a control function that could in principle be entrusted to a machine. Of course, we are speaking not of an absolute but of relative information shortage, since in fact a manager is usually flooded by very diverse data. But what he needs is not 'crude' information about the object of management but the specific quintessence of that information. And he has no time himself to extract the information he needs for management.

"The management of society and its sections, the organization of human labour is primarily an economic, social and ideological problem, and not a cybernetic, natural-scientific, technical or any other problem." (1) This implies that apart from information about scientific and technological accomplishments, a manager needs ideological, socio-political, economic and organizational information. A special place in management belongs to fore-

---

1. Afanas'ev, V.G. Nauchnoe upravlenie obshchestvom (Opyt sistemnogo issledovaniya). (Scientific management of society. Experience in system research). Moscow, "Politizdat" Publishers, 1968, p. 379.

cast information, since timely and adequate decisions are impossible without an analysis of the developmental trends of the phenomena concerned.

All these considerations suggest that management should have its own information service, of which the system of scientific and technical information would be one component. Information service to management ought to cover many more sources of information, subjecting them to a stricter selection, and deeper analytico-synthetic processing, and provide this information faster and reduced to optimal volume. Redundant information hampers decision-making, as the number of alternative strategies tends to grow with increased volume of information. Even the most perfect information service, however, will not relieve the manager from some information processing that he will still have to perform personally.

These considerations were mentioned back in 1963 in the report "Science, government and information" submitted to President John F. Kennedy by his Science Advisory Committee: "Because information about what is going on is necessary for making management decisions, improvement in scientific information systems is sometimes represented as a panacea for bad management of research and development. Though it is true that poor management can and does occur with the best of communication systems, poor communication almost always leads to bad management." (1)

Study of information needs

The three user groups identified above are very broadly defined. Moreover, they are by no means exhaustive. They do not include, for example, such large and particular user groups as teachers and students. A great number of scientists are engaged not in fundamental but in applied sciences. The information needs of engineers vary widely according to whether they are concerned with design or operational work. Many scientists and engineers are at the same time managers. Information services in many fields of science and practical activity, in particular in the social sciences, medicine and agriculture, have certain specific features. All these and many other aspects of the use of scientific information require a systematic study.

The information needs of scientists and other specialists, as well as their use of scientific information, are studied by sociological research methods, with ever deeper exploration of the psychology of scientific creativity. This problem, which surpasses the scope of informatics, has not yet been adequately investigated, although hundreds of studies are devoted to it every year in different countries. Only a few of these are of a general nature aimed at establishing common rules for the use of scientific information; most pursue narrower practical aims, e.g. to improve information service to this or that particular user group, or to establish the optimum pattern of functioning of particular information centres or networks. This research is still being conducted through scattered efforts and lacks any common methodology, with the result that the findings are difficult to compare. Even so, some common objects and methods of study are already identifiable.

---

1. Science, Government, and Information. The responsibilities of the Technical Community and the Government in the Transfer of Information. A report of the President's Science Advisory Committee. Washington, US Government Printing Office, 1963, p. 10.

First of all, information users themselves are investigated, as also are their typology, dependence of information needs on their occupations, the purposes for which they need information, qualifications, age, professional position, and a range of other characteristics. Some findings suggest that the occupation of the user and the kind of establishment he is employed in are the factors determining his information needs. The different working conditions influence the time spent by the user on information retrieval, use of information sources, and frequency of using libraries and information centres.

Whether the organizational structure of user groups is worth studying depends on the field of activity. Big, institutionalized user groups are mainly studied in technology and applied sciences, and in such natural sciences as chemistry and physics. In biology and other life sciences attention is focussed on use of information in small research teams; and in social sciences and the humanities, on individual user needs.

In this connection it is important to study the activities of different types of information centres, types of information service, and channels of information dissemination. No less attention should be paid to the study of sources of information, the dependence of the use of various primary and secondary documents on user activity, and other user characteristics.

All these and other aspects of the use of information are studied by sociological methods. Questionnaires are widely used. They are distributed among a representative sample of the user group studied. The replies are statistically processed and interpreted, so that the findings either corroborate or refute a previously formulated hypothesis about this or that relationship of factors influencing use of scientific information. This method is usually supplemented by interviews with those users whose replies are unclear or hard to interpret. Interviews can also be used whenever it would be difficult to formulate a questionnaire, provided the group studied is not too large. Other methods worth mentioning are: observation, and the keeping of diaries by users themselves. Complex studies using systems engineering and operations research techniques have been carried out lately. Promising results have been achieved in some aspects of information utilization studies by the analysis of citations in journal articles, feedback from SDI system users, and by the analysis of requests coming to libraries and scientific information centres.

All these methods, however, have their limits since they are based on user judgments that are naturally biased. The constant contact of a user with his information system results in his no longer expecting of the system anything beyond its capacities. This habit becomes a stable psychological factor likely to distort subtantially the results of any study based on some of the above mentioned methods or their combination; the purpose of a user study, after all, is to learn their actual needs and not what they think are their needs. This suggests that for studying different types of information needs at various stages of research and design work it is necessary for the explorer personally to participate in that work.

Types of information needs

The most general types of information needs were defined by M. Voigt, an American information scientist, in the early 1960s. (1) By questioning a hundred Scandinavian scientists in chemistry, physics and biology, he ascertained that they refer to information sources mainly in three cases:

1) while getting current awareness of the results both in their particular narrow field and the related disciplines;
2) in their day-to-day work, when they need some particular factual information on figures, methods and designs; and
3) when embarking on a new problem or a project, as well as when completing it and writing about it – they need then a retrospective search to identify as many published and unpublished sources on the subject concerned as possible.

The first case may be described as a natural need of any scientist or other specialist for keeping abreast of developments in science, technology or culture; the second case is the everyday use of information for reference purposes; and the third case is an exhaustive approach to scientific documents relating to a particular topic or subject.

These types of user needs have been since then repeatedly re-interpreted and specified in keeping with the different aspects of user studies. For example, as regards technology, it has been suggested that three stages at which information needs arise should be considered: the choice of the trend of an R & D or design project, its planning, and the execution of the project. In choosing the trend, the user needs general information on the existing approaches to the problem of his concern, on their relative advantages and drawbacks both of essential and of organizational and technical nature. When planning a project in conformity with the trend chosen, he needs more detailed data on the resources and their distribution (time, money, personnel, equipment, facilities). Lastly, at the execution stage, what he needs is very detailed information on all the elements of the object under development.

A more general approach to the differences in the information supplied at various stages of scientific and technical work is developed in the collective monograph by a group of Ukrainian scientists entitled "Information in scientific research" and published under the editorship of Ukrainian Academician A.N. Shcherban'. (2) They distinguish seven such stages:

1. Overall familiarization with the problem, and problem statement. Drawing up a plan and the provisional terms for the solution of the problems of primary and secondary importance. This stage comprises general acquaintance with information available on the subject.
2. Gathering scientific knowledge about the subject of study. Retrospective searching of the broadest possible scope of literature without any pronounced critical approach.

---

1. Voigt, M. Scientists' approaches to information. New York, ALA, 1961.
2. Informatsiya v nauchnykh issledovaniyakh (Information in scientific research). Otv. red. akad. AN USSR A.N. Shcherban'. Kiev, "Naukova Dumka" publishers, p. 119-120.
(Nauka, uchenyi, informatisiya (Science, scientist, information) by S.G. [illegible], A.V. Kozenko, V.V. Kosolapov, A.N. Polovinchik).

3. Coordination and interpretation of scientific data. A critical evaluation of the ideas and hypotheses of different authors. The relevance criteria for the information needed are specified and the amount of documents used is reduced.
4. Statement of hypotheses and choice of a working hypothesis, which is the most important stage of research in technical sciences. As to the use of information, it is characterized by deep processing rather than broad coverage. Sometimes, however, a working hypothesis may be chosen under the influence of what might appear 'irrelevant' information.
5. Proving the working hypothesis; the most important stage in basic research. The information used depends on the specifics of research; for instance, the proving of an assumption may require gathering factual data scattered in the literature.
6. Statement of conclusions and recommendations. Predictions, as well as generalizations, are frequently made at this stage. Information is often used to shed light on precedence and priority aspects.
7. Description of the research results. The information gathered and processed is, as a rule, minutely documented.

The above schemes are, of course, very far from perfect, and they may be greatly modified depending on the peculiar features of this or that study. Even so, they are evidence of the close association between the use of scientific information and the logical structure of scientific research.

There have also been attempts to assess the use of different sources as dependent on the kind of activity pursued by the user and on the stages of the scientific research process. It has been noticed, for example, that abstract journals, bibliographies and other secondary documents are more intensively used at the outset and at the end of elaborating a problem, while the need for reference aids and primary documents (particularly, special types of technical literature) is more acutely felt at the experimental testing stage of theoretical assumptions, equipment, process or product specimens.

A great many studies concerned with the channels through which scientists receive information indicate inadequate use of information services and publications. While the numerical results of these studies, which use different methods and different user groups, vary greatly, it is almost a rule that among the most important sources yielding more than half of all information needed for research such 'incidental' (not purposeful) and for the most part oral sources are provided by talks with colleagues and conference reports. Much less information (10-20 per cent) is extracted from abstract journals and bibliographic indexes, and even less from library catalogues. Thus, for all the inadequacies of methodology and unreliability of the findings of user studies, it may be asserted indisputably that the efficiency of scientific information activities still remains very low. The problem of evaluating this efficiency arises in this connection.

Efficiency of scientific information activities

This problem has become important since scientists and other specialists feel themselves forced to spend an ever greater share of their time on searching for information instead of doing proper research. A recent study of the time balance of American chemists showed that they spend an average of 45% of their working time looking for scientific and business information, while all their experimental work accounted for only 36%, and the processing of the results obtained less than 6%.

The efficiency of scientific information activities is a comparatively recent concept. It still lacks a unique definition and is diversely interpreted, sometimes as the total revenue from this activity, sometimes as the surplus of revenue over the expenditure, sometimes as the ratio of revenue to expenditure, This 'economic' approach to scientific information activities stems from the fact that ever greater funds are invested in this work and the requirements placed on it are growing. "The last factor gives rise to two questions: what benefits can be derived from information activities by institutions and enterprises which organize the information service or by the national economy (in case of socialist countries where national information systems are established), and what is the correlation between profits obtained and expenses incurred?" (1)

It is relatively easy to count up the expenditure on scientific information activities, which is, incidentally, quite sizable as we have seen. As to the economic benefits, they are much harder to determine, since they lack any direct economic expression. The difficulty is similar to that of numerical evaluation of the effect of education, Therefore, the existing system of the evaluation of economic activities proves inapplicable in this case. The Polish information scientist just quoted suggests that information efficiency should be measured by the ratio of the amount of relevant information supplied to users to the total amount of information gathered and processed at the information centre concerned. (2) He has formulated the corresponding efficiency criteria and calculation techniques and relationships.

However, other difficulties arise. We have already seen how relevance judgments may be subjective in measuring the efficiency of information systems.Here also, the assessment criteria are mostly based on a correspondence between the contents of the collected documents, on the one hand, and on potential user needs and actual user requests, on the other. But we are still unable to measure the objective value of scientific information in the same manner as mathematical information theory measures information contained in a message by the uncertainty this message serves to eliminate. And, it is not on the knowledge of an individual about the object or phenomenon described in the message that the measure of this uncertainty depends, but only on the objective properties of that object or phenomenon.

Despite the fact that scientific cognition is social by nature, we are able to measure the value of scientific information only with reference to separate individuals or their groups. Human society does possess a social mechanism which objectively evaluates scientific information but, unless and until this mechanism is explored and understood, measuring the efficiency of scientific information activity is bound to remain empirical.

The contents of scientific information activities

We have ascertained that any creative work may be viewed as the process of selecting and processing information for different purposes: in scientific research - to create new scientific information; in engineering

---

1. Piróg, W. Efficiency of Information Activities: Evaluation Criteria, Methods and Indices. In: On Theoretical Problems of Informatics (FID 435). Moscow, 1969, p. 93.
2. Ibid, p. 94.

and industrial activities - to build new machines, and develop new substances and engineering procedures; and in the management sphere for taking organizational decisions. But scientific information activity itself consists to a great extent in selecting and processing scientific information. What then are its specific features and what is the borderline between scientific information work and other creative activities?

Two diametrically opposed replies have been given to this question: the first viewpoint reduces scientific information work to just a variety of auxiliary service to research and development, and to science management activities; the other viewpoint fully identifies it with the rest of research work in general. Neither point of view appears tenable. If the scope of scientific information work is limited to a variety of auxiliary or technical assistance rendered to scientists and other specialists in furnishing them with documents they want, why has it become established as an activity in its own right in face of library and bibliographic activities that presumably have the same function? This view creates a dangerous illusion that the 'information crisis' could be cured simply by better special libraries without any essentially new methods and media of scientific communication; it virtually negates scientific information work as an independent activity, though this is contradicted by actual social practice.

On the other hand, to broaden the scope of scientific information activities so that it encompasses authentic research as well, is just as wrong; eventually this, too, boils down to a denial of its independent status. The erroneous identification of basic research with scientific information work stems from the fact that both the theoretical scientist and the information scientist are engaged in analysis and synthesis of scientific information at the stages of its selection and - especially - processing, and that they both deal almost exclusively with scientific documents. Outwardly, the work of U. Le Versier (who computed the orbit of Neptune, a planet then unknown) or of D.I. Mendeleyev (who discovered the Periodic Law) resembles the work of an information officer who prepares for a researcher the information he needs, extracting it from scientific documents and presenting it in a form convenient to the researcher. For all the outward likeness, however, the two kinds of work have obviously differing goals; while in the former the goal is to discover scientific laws and create new scientific information, in the latter it is only to provide prerequisites to authentic research.

Naturally, the border between the two kinds of activities is conventional and changeable; it depends on a great number of factors, in particular on the stage of development of the science in question, the traditions existing in it, the level of formalization it has reached, and on certain others. Still, this borderline exists quite clearly at a given moment under any particular conditions. Scientific information work constitutes, as it were, the opening and preparatory phase of any research. It creates conditions for more efficient investigation by selecting the documents pertinent to the subject of research and providing the scientist or expert with the information he needs, extracted from those documents and tailored to his needs. Part of this work is usually performed by the user himself. It is in this sense that the researcher himself is sometimes said to carry out scientific information work, which is not quite true since the term should be applied only to the socially and institutionally organized forms of activity.

Scientific information activities, as well as the system of scientific publications in which these activities first emerged, have existed and still exist as a natural condition of the development of science. Since their very emergence, they were unfolding within science as an important means of scientific communication. For ages, scientists - supported only by libraries - themselves searched, selected and processed scientific information. This was the reason that some important forms and methods of scientific information work first took shape in the library - until recently the only social institution to store, analyse and generalize the age-long experience in the field.

The advancement of science in the last decade, and its conversion to a major and mass-scale human activity, as well as the gradual penetration of science by industrial methods and organization of management, have all led to the establishment of scientific information work as an organic but autonomous part of scientific work. In the first chapter we spoke of the similarity of this separation to the division of research into theoretical and experimental. We can now bring more precision to this.

The process of social division of labour in modern science is unfolding along at least two different lines. Research divides into basic and experimental, primarily according to the methods employed. Simultaneously, a similar process is developing along a different, notably the functional line, where three independent varieties of scientific work are clearly seen - research proper, scientific information activities, and science management.

Prospects for scientific information work

While discussing the contents and place of scientific information work in science, it would be wrong to pass over its future prospects. This is the more essential as the scope of scientific information activities is continually expanding. C.N. Mooers, a pioneer of automated information retrieval, said in 1959, in a report at a conference: "Information retrieval is a term that I had the pleasure of coining only ten years ago, at another computer conference. The term has come a long way since. In thinking about a definition of information retrieval, and in considering the future of this field, we must take an evolving view. At the present time, information retrieval is concerned with more than the mere finding and providing of documents; we are already concerned with the discovery and provision of information quite apart from its documentary form. To encompass future developments, as we shall see, even this broad view of information retrieval will have to be modified and extended". (1)

A half of the term of this prediction now being over, one is entitled to say that it has proved to be valid. The proposition is especially important in forecasting the development of information activities which has its own rules and cannot be constructed in an arbitrary manner.

The rules of the growth of science and the specific features of the present stage of scientific and technological revolution are undoubtedly the key factors affecting the growth of scientific information activities.

---

1. Mooers, C. The Next Twenty Years in Information Retrieval. Some Goals and Predictions. "American Documentation", 1960, Vol. 11, N. 3, p. 229.

There are many facts in the practice of information service to indicate a trend towards a further shift of tasks from the scope of research proper to the sphere of scientific information work. The general eagerness of users to get as much information as possible whilst spending the minimum of effort is bound ever more insistently to press on information systems the tasks of synthesizing, creatively processing and tailoring information to user needs. Continuing interaction between information systems and different user categories will ultimately enrol users' participation in systems optimization and will increase the contribution by scientific information work to research endeavour.

To summarize: in this last chapter, we have profiled the role of scientific information activities under a new angle, discussed the varying patterns of the use of scientific information by the main user groups, highlighted the principal objects and techniques of studies of user needs and of the main types of user needs arising at different stages of research and development work, touched upon efficiency evaluation problems, specified the contents of scientific information work and outlined its development prospects.

As we have seen, the better organized the research conducted in the different fields of science and technology and the more formalized the special languages and methods used in those fields, the more sophisticated will be the information selection and processing tasks entrusted to institutionally independent information services. Knowledge of the basic rules, methods and media of scientific information work is indispensable for coping with these tasks. Providing a guide to that essential knowledge is the purpose of this book.

Questions for self-checking

1. What principal groups of scientific information users are distinguishable? What are the differences between their information needs?
2. What are the objects and techniques of user needs studies?
3. How is scientific information utilized at the different stages of research and development work?
4. How can the efficiency of scientific information work be evaluated?
5. Where does the borderline between research proper and scientific information work lie?
6. What are the major development prospects of scientific information activities?

Literature

1. Dean, H.H., Bryson, K.D. Effective Communication. A Guide to Reading, Writing, Speaking, and Listening. 2nd ed. Englewood Cliffs, Prentice-Hall, 1965, XVI, 560 p.
2. Fairthorne, R.A. The Scope and Aims of Information Science and Technologies. In: On Theoretical Problems of Informatics (FID 435). Moscow, 1969, p. 25-31.
3. Information Systems Workshop. The Designer's Responsibility and his Methodology. Washington, Spartan Books, 1962, V, 153 p.
4. Kent, A., Taulbee, O.E. (eds.). Electronic Information Handling. Washington, Spartan Books, 1965, VII, 355 p.
5. Scientific and Technical Communication. A Pressing National Problem and Recommendations for its Solution. A synopsis of the report of the Committee on Scientific and Technical Communication of the National Academy of Sciences - National Academy of Engineering, Washington, 1969, 30 p.
6. Wysocki, A. Study of User's Information Needs: Subjects and Methods. In: On Theoretical Problems of Informatics (FID 435). Moscow, 1969, p. 80-92.

SYLLABUS AND PROGRAMME

for the

INTRODUCTORY COURSE ON INFORMATICS/DOCUMENTATION

1. The subject and methods of informatics (documentation). *Laws of scientific development* that brought about the establishment of scientific information work as an independent discipline: differentiation and specialization, traditions and international scale of this activity, its accelerated development. *Objectives and stages* of scientific information activity: collection, analytical and synthetic processing, storage, retrieval, dissemination, and use of scientific information. *Fundamental concepts*: information, scientific information, scientific information activity, informatics (documentation). *Interrelations* between informatics and information theory, cybernetics, semiotics, linguistics, psychology, library science, bibliography, history of books, science of science, and technology.

2. Scientific literature as a medium of dissemination of knowledge (primary scientific documents and publications). *Phases of evolution of written documents*: cuneiform writing tablets, papyrus scrolls, parchment codices, manuscript paper books, printing, modern methods of publishing. *Scientific documents and their types*: published and unpublished, primary and secondary documents. *Description* of the principal types of primary documents: books (monographs, collective volumes, conference proceedings, textbooks and manuals, official publications, book series; book statistics); periodical publications (journals, serials, newspapers); special technical publications (standards, patent specifications, technical catalogues); unpublished primary publications (scientific and technical reports, theses, preprints, deposited manuscripts). *The laws* of distribution and obsolescence of scientific publications (Bradford's Law of Scattering, bibliographic coupling of documents, obsolescence of books and of journal articles, the 'half-life' of scientific literature).

3. Secondary scientific documents and publications. *Conventional* secondary documents: reference literature (encyclopaedias, reference handbooks, dictionaries), surveys (e.g. 'Progress in Science', 'Scientific and Technical State-of-Art Communications'), abstract journals (specifying the goals, scope, volume, and pattern of the principal abstract journals in natural and social sciences: 'Chemical Abstracts', 'Biological Abstracts', 'Science Abstracts', VINITI's 'Referativnyi Zhurnal', 'Bulletin signalétique', etc.), current awareness bulletins, library catalogues (classified, subject, dictionary, alphabetic), bibliographic indexes (main types as defined by contents, purpose, date and place of publication: 'Cumulative Book Index', 'British National Bibliography', 'Bibliographie de la France', 'Deutsche Nationalbibliographie', 'Deutsche Bibliographie', 'Knizhnaya letopis SSSR', etc.); survey of special retrospective bibliographies and bibliographies of bibliographies (Besterman, Malclès, Totok und Weitzel, etc.); *Unconventional indexes*: permuted (KWIC), KWOC, tabledex, citation indexes; current awareness publications and services (current contents, selective dissemination of information).

4. Organization of scientific information activities. *Scientific information channels*: personal contacts and written communication of scientists, exchange of preprints and reprints, participation in scientific meetings; 'invisible colleges'. Editing and printing processes and their place in scientific information activities. Relationships between libraries and information centres. *Information centres and facilities*: systems of information serving specialists in the major countries, description of the biggest information agencies, information practices and procedures. *Special (research) libraries*: national and larger independent universal libraries, libraries of R & D centres, university libraries, and libraries in industry; principal library practices and procedures. *International organizations* engaged in scientific information and library activities: Unesco, FID, IFLA, ISO/TC 46, ICSU, and others. International cooperation in these fields.

5. Fundamentals of information retrieval. *Information retrieval*, information retrieval systems (document and fact retrieval), information retrieval languages, *retrieval criteria* (relevance and pertinence of documents), *performance* of retrieval systems and methods for its assessment.

6. Conventional information retrieval systems. Library regarded as an IS & R system, library catalogues. *Alphabetical retrieval files*, the methods of bibliographic description. *Classified files*: principles of classification, hierarchical library and bibliographic classifications, UDC and its application. *Facet classifications*. *Subject classifications* and subject indexing procedures.

7. Non-conventional information retrieval systems. *Coordinate indexing* and its distinctions from conventional indexing systems. Descriptor languages and their creation: synonymity and multiple-meaning control. *Thesauri*: general principles, main types of thesaurus, representation of paradigmatic and syntagmatic relations, methods of thesaurus construction. *Specialized information retrieval languages*: WRU semantic code, SYNTOL.

8. Hardware of information retrieval. Retrieval devices operating with *discrete* information media: catalogue file cards, Uniterm cards, punched cards (manual and machine-sorted), microfiches ('Filmorex', 'Minicard'). Retrieval devices operating on *continuous* media: punch tapes, roll microfilm (microfilm selectors: 'Rapid Selector' system, 'Miracode'), magnetic tape ('Videofile'). Computer-based implementation of information retrieval systems.

9. Reproduction of documents. *Document copying*: silver processes (photography, reflex processes; stabilization, diffusion, matrix Verifax); diazo processes (diazocopying, Calfax); thermographic processes (Thermofax, Masterfax, Imagic); eletrographic processes (Xerography, Electrofax, transfer, and thermoplastic processes). *Document duplication*: printing, processes (relief and plane printing, photogravure); typesetting machines; offset, stencil, spirit duplication, addressing machines.

10. Use of scientific information. Study of information needs and requests. Efficiency of the use of information. Effective methods of document handling. Influence of scientific information activities on development of science, technology, economy and culture.

Training programme of the course "Introduction to Informatics/Documentation"

Chapters, subjects	Hours		
	Lectures	Seminars	Total
1. The subject and methods of informatics/documentation	2	-	2
Sources of scientific information			
2. Scientific literature as a medium of knowledge dissemination	2	2	4
3. Secondary scientific documents and publications	4	4	8
4. Organization of scientific information activities	4	2	6
Information retrieval			
5. Fundamentals of information retrieval	2	-	2
6. Conventional information retrieval systems	4	4	8
7. Non-conventional information retrieval systems	4	2	6
8. Hardware of information retrieval	2	2	4
Use of scientific information			
9. Reproduction of documents	2	4	6
10. Use of scientific information	2	-	2
Total	28	20	48

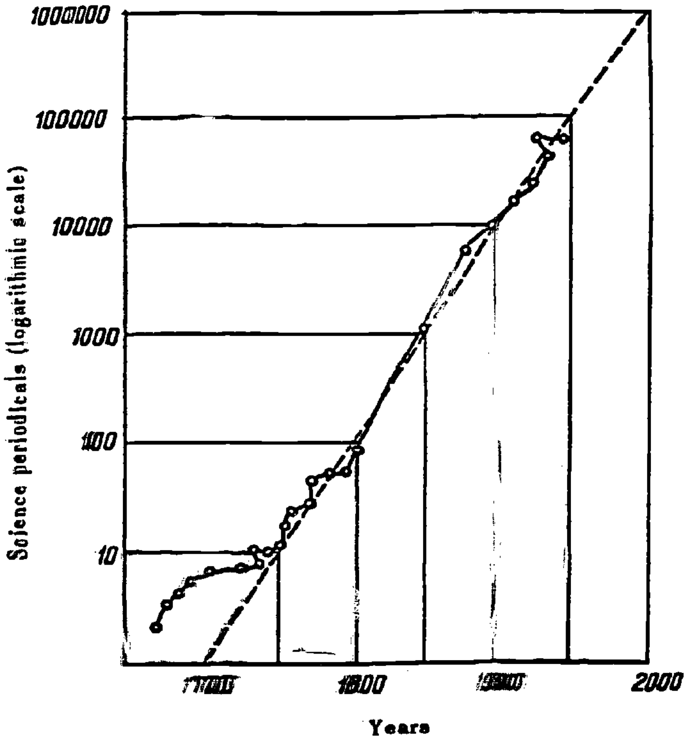


Fig. 1. Growth of world science periodicals.

The diagram shows the total number of periodicals, including those no longer published. The notion of a "science periodical" has been taken in a broad sense. Adopted from: Price D.J. de S. Science since Babylon. New Haven, 1961, p. 97.

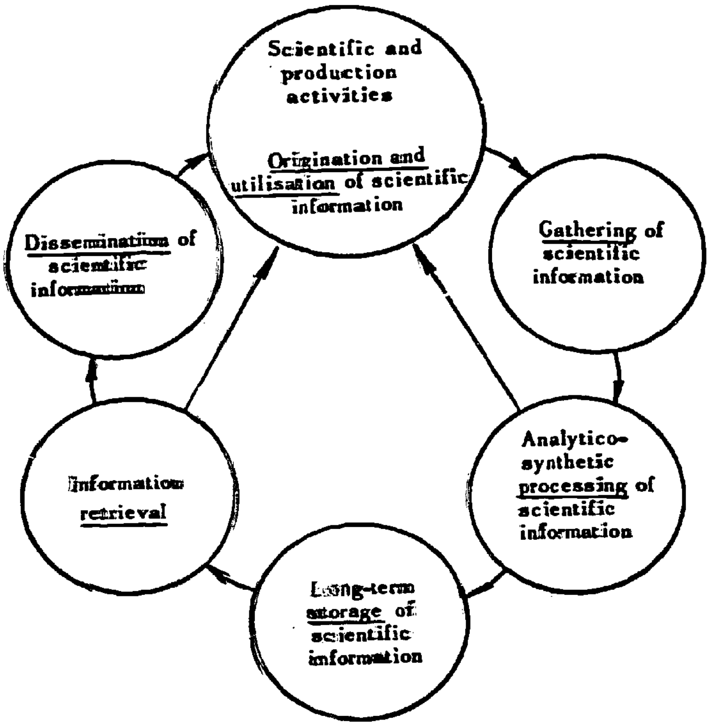


Fig. 2. Tasks and stages of scientific-information activities.

Origination of scientific information is not included in scientific-information activities by some authors. However, origination as well as utilisation of scientific information make up the organizational sphere of scientific labour and are obviously major aspects of scientific-information activities as a whole.

Scientific documents and publications

	Primary	Secondary
Published	Books and pamphlets	
	Monographs Collections of papers Conference proceedings Textbooks and manuals Official publications	Reference literature
	Periodical publications	
	Serials Journals and magazines Newspapers	Surveys Abstract journals Express information bulletins
	Special types of technical publications	
	Standards Invention specifications Technical catalogues Information leaflets	Standards indexes Invention bulletins Patent classifiers
Unpublished	Scientific and technical reports Dissertations Information cards Preprints Manuscripts and galleys Data files	Library catalogues Bibliographic files Translations

Fig. 3. Types of scientific documents and publications.

For clarity each type is mentioned only once. Actually many types could be indicated in more than one place, for example, bibliographic indexes are not always periodical, library catalogues may be published, etc.

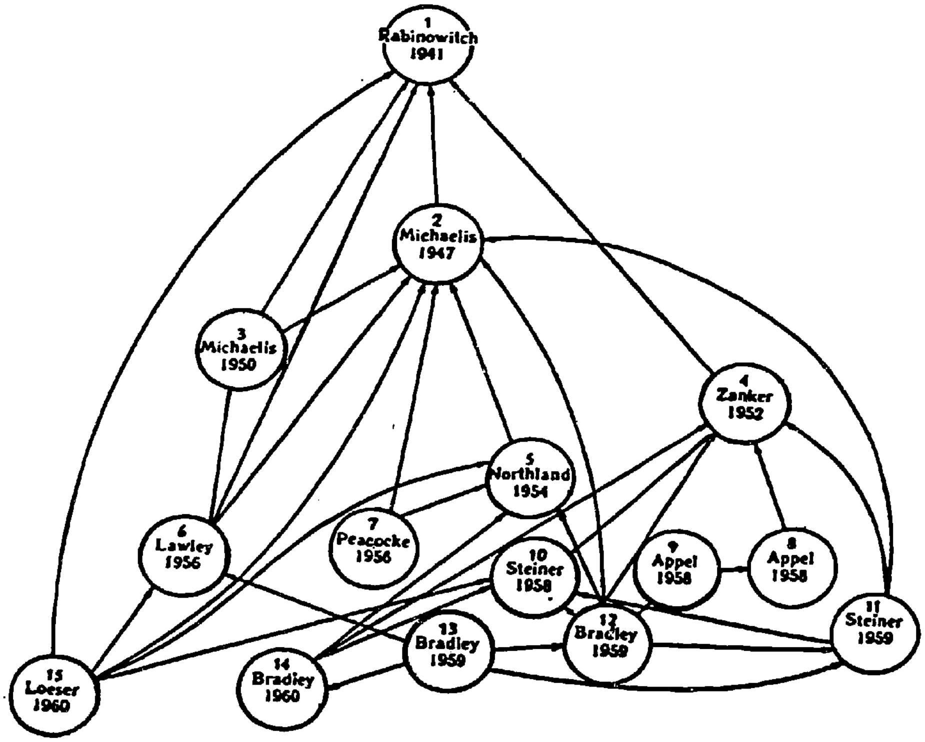


Fig. 4. Citation of science documents in later publications.

Circles indicate works on nucleic acids published between 1941 and 1960, e.g.:

1. Rabinowitch E. and Epstein L.F. Polymerization of Dyestuffs in solution. Thionine and methylene blue. - "Journal of American Chemical Society", 1941, v. 63, No. 69.
10. Steiner R.F. and Beers R.F. Spectral changes accompanying binding of Ao. by Poly A. - "Science", 1958, v. 127.

Arrows show references. The diagram is by G. Allen (Carnegie Institution). Adopted from the prospectus to "Science Citation Index" for 1963.

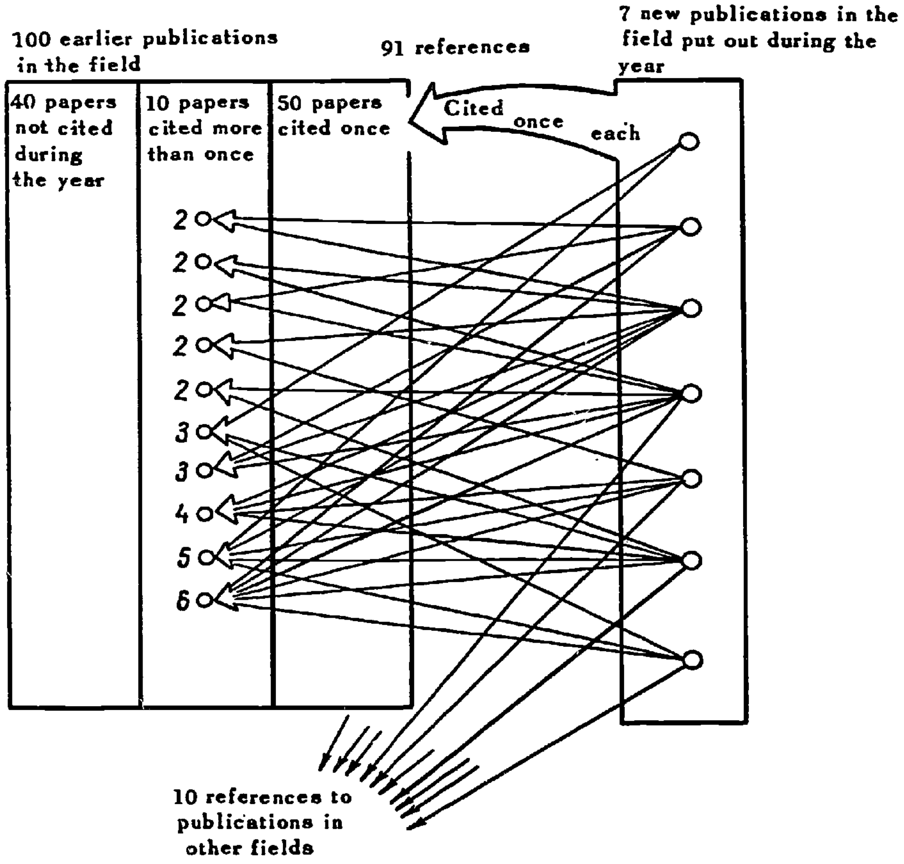


Fig. 5. Distribution of new references to earlier publications in a hypothetical field.

The number of publications was arbitrarily set equal to 100, the rest of the data refer to the average values for the different fields of knowledge derived from an analysis of the 1961 Science Citation Index. To make the diagram more illustrative an almost "closed" field is shown so that only 10 references are to publications in other fields. Due to the small number of papers in the right-hand column (7), the distribution of references they contain could not be shown adequately in conformity with the regularities indicated above.

Adopted from: Price, D.J. de S. Networks of scientific papers. - "Science", 1965, v. 149, No. 3683, p. 512.

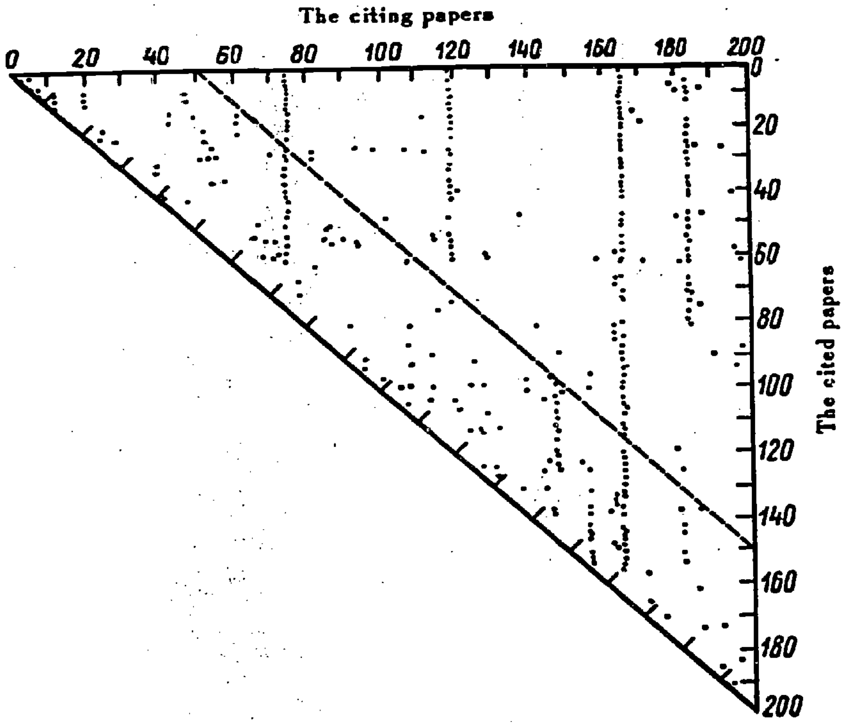


Fig. 6. Scatter of citations in an isolated special literature.

The horizontal and the vertical axes indicate numbers of the same papers ordered chronologically. All the papers refer to a narrow physical subject - the study of imaginary N-rays falsely discovered in 1904. Each column of dots designates the references contained in the paper indicated by the horizontal number opposite that column. Every dot designates a reference to the paper indicated by the number on the vertical axis opposite to it.

Adopted from: Price, D.J. de S. Networks of scientific papers. - "Science", 1965, v. 149, No. 3683, p. 514.

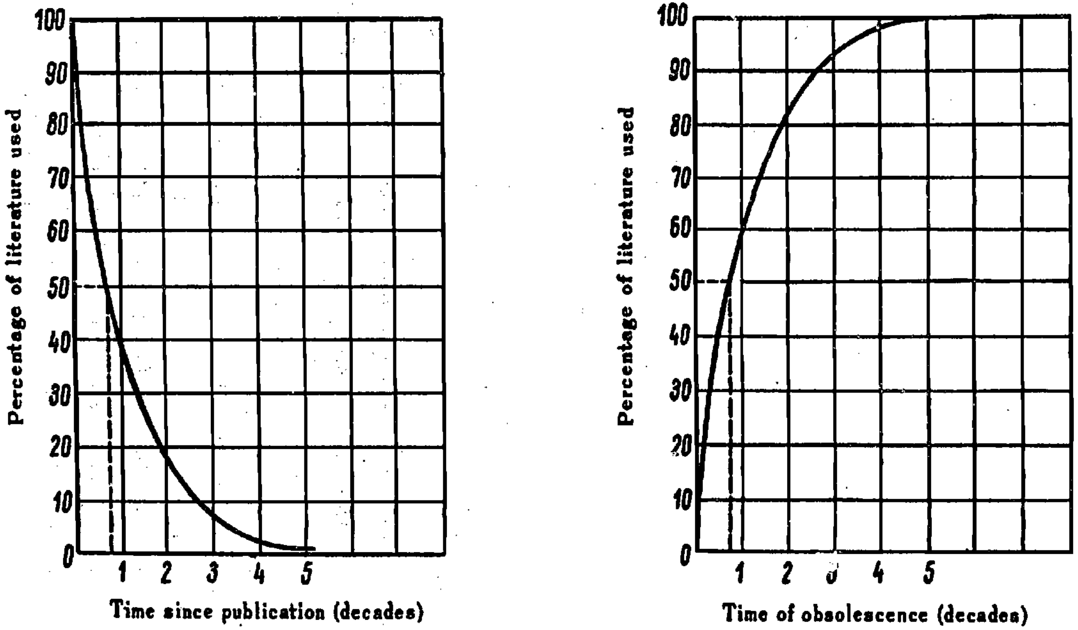


Fig. 7. Half-life of scientific literature.

Adopted from: Burton R.E. and Kebler R.E. The "half-life" of some scientific and technical literatures. - "American Documentation", 1960, vol. 11, No. 1, p. 18-20.

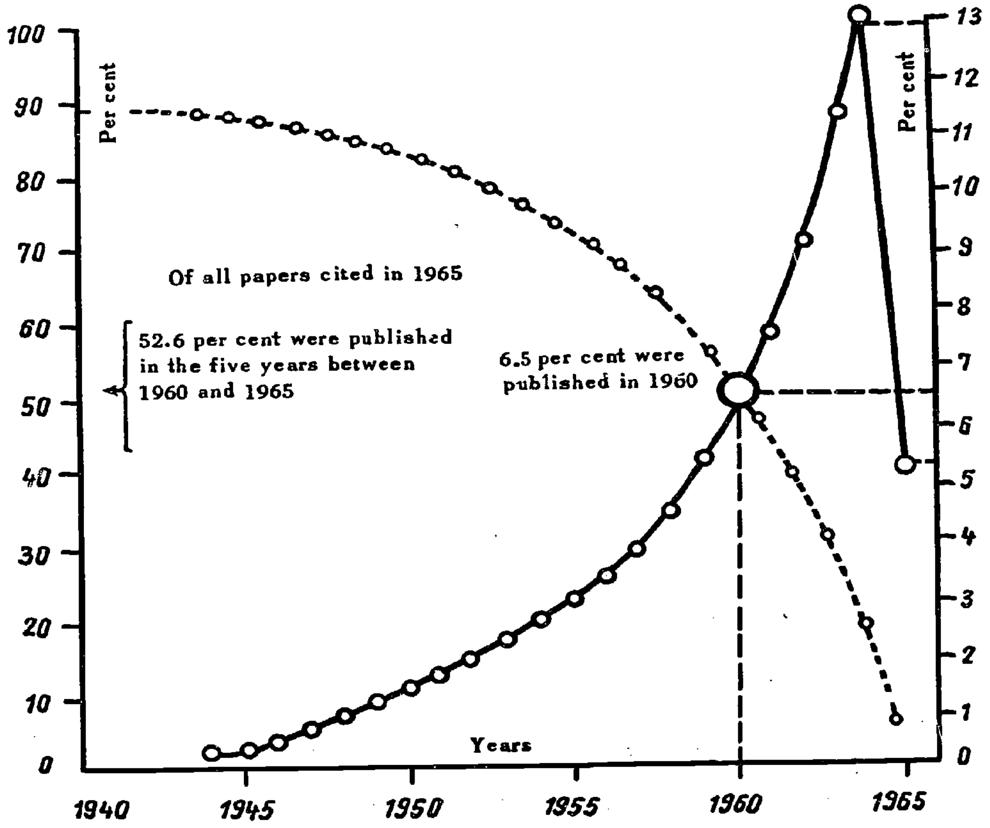


Fig. 8. Age distribution of papers cited in publications of 1965.

The horizontal axis shows the publication years of papers cited in 1965, in chronological order, the vertical axis shows the proportions of the paper of the specified publication years to the total number of references made in 1965. The dotted lines indicate the percentage of the total publications issued between 1941 and 1965 to the total number of publications cited in the publications of 1965. The diagrams are based on data taken from: Science Citation Index Guide. Philadelphia, 1966, p. 32.

2 Н115 П. Получение метансульфонилхлорида. Bennett Clifton F., Goheen David W. Making methanesulfonyl chloride. [Crown-Zellerbach Corp.]. Пат. США. кл. 260—543, № 3226433, заявл. 2.01.62, опубл. 28.12.65

CH_3SO_2Cl получают пропусканием Cl_2-газа в р-ционную смесь (от —20 до +25°, лучше от —10 до +10°), содержащую стехиометрич. кол-ва $ClCH_2S$-CH_3, воды (лучше 3—10 молей) и, по крайней мере, один орг. р-ритель (AcOH, CCl_4, хлф.) для $ClCH_2SCH_3$ и к-ты (HCl, разб. H_2SO_4, $\cdot H_3PO_4$). $ClCH_2SMe$ (81,2 *г*) смешивают (20°) в течение 5 мин. со 100 *г* воды, смесь охлаждают до 0° и через нее пропускают 2,6 моля Cl_2-газа (0—10°, 120 мин.). Орг. слой отделяют, сушат и разгоняют с получением 66,1 *г* $MeSO_2Cl$ (выход 68,7%). Р-цию можно ускорить и выход $MeSO_2Cl$ можно увеличить, если $ClCH_2SMe$ предварительно кипятить с водой в течение периода от 5 мин. до 2—3 час.; после этого хлорирование идет эффективней. Приведены примеры получения $MeSO_2Cl$ с предварительным кипячением, в присутствии р-рителя и подкисляющего агента. $MeSO_2Cl$ применяется в кач-ве р-рителя, полупродукта для получения различных орг. продуктов, включая полимеры, пластификаторы, детергенты, красители, фумиганты, инсектициды, огнестойкие составы и терапевтич. лекарства.
И. Дорман

Preparation of methanesulfonyl chloride from monochlorodimethyl sulfide. Clifton F. Bennett and David W. Goheen. (to Crown Zellerbach Corp.). **U.S. 3,226,433** (Cl. 260-543), Dec. 28, 1965, Appl. Jan. 2, 1962; 2 pp. Methanesulfonyl chloride may be readily prepd. from mono-, di-, or trichlorodimethyl sulfide by chlorinating the organic sulfide in a mixt. of H_2O–HCl–organic solvent at −10 to +10°. The preferred starting material is the monochlorinated dimethyl sulfide, and for optimum yields of product three moles each of Cl and H_2O must be used. Preliminary refluxing of the sulfide with H_2O for 5 min. to 2 hrs. allows the chlorination to proceed smoothly. The reaction proceeds by the following equation: $ClCH_2SCH_3 + 3Cl_2 + 3H_2O \rightarrow CH_3SO_2Cl + 6HCl + HCHO$. For example, 107 g. monochlorodimethyl sulfide was admixed with 75 g. H_2O and 29.8 g. 37% HCl, the mixt. refluxed 10 min., cooled to 0°, and maintained at this temp. while bubbling into it 248 g. Cl over a period of 60 min. The org. layer was sepd., dried, and distilled; the yield of methanesulfonyl chloride was 75.6%.
R. L. Littler

Fig. 9. Two different abstracts of the same paper as abstracted in the VINITI Abstract Journal "Chemistry" and the "Chemical Abstracts"

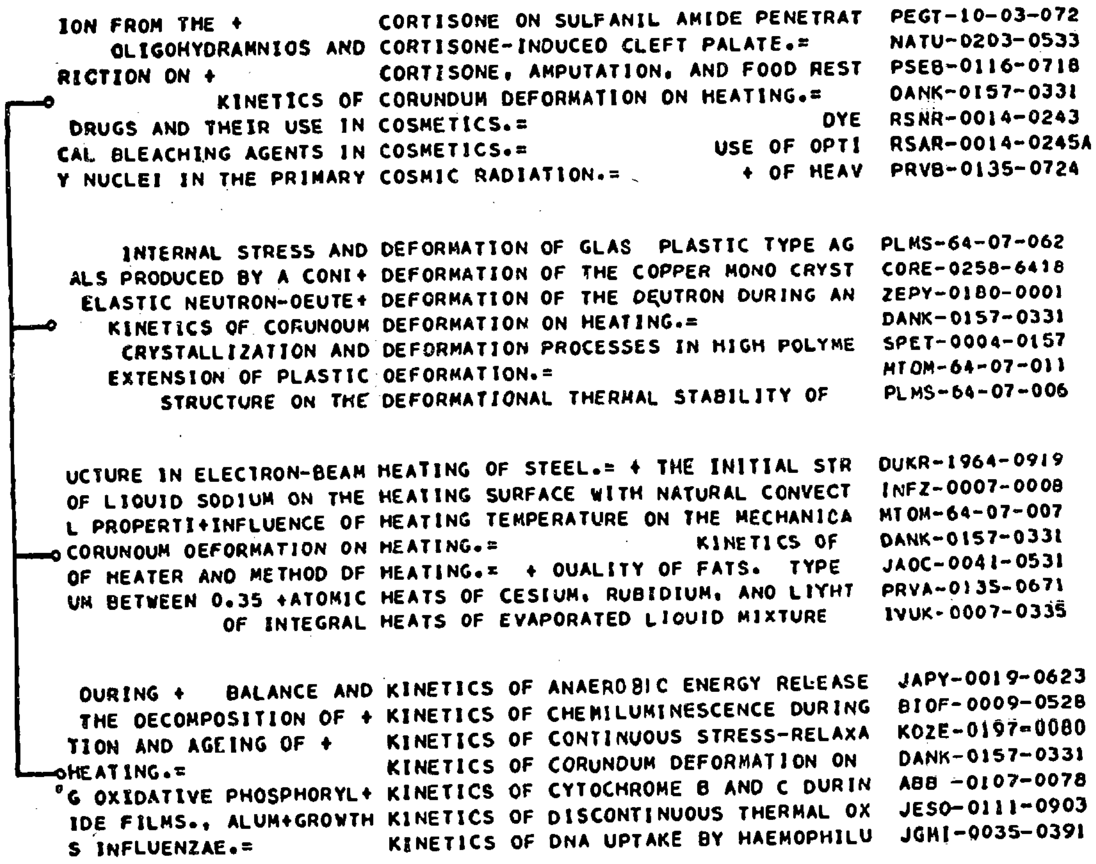
This example shows the use of the formalized chemical language for contracting the text of an original paper to its abstract. The abstracts have largely overlapping contents.

Classified order (UDC)	Alphabetical subject order	Author order
1. Philosophy McLuhan M. Understanding media. 5. Mathematics and natural sciences Wittgenstein L. Remarks on the foundations of mathematics. 6. Applied sciences. Engineering Keenan J.H. Thermodynamics. 7. The arts Rasmussen S.E. Towns and buildings. 8. Linguistics. Philology Ullman B.L. Ancient writing and its influence.	Architecture Rasmussen S.E. Towns and buildings. Mathematics Wittgenstein L. Remarks on the foundations of mathematics. Paleography Ullman B.L. Ancient writing and its influence. Social communication McLuhan M. Understanding media. Thermodynamics Keenan J.H. Thermodynamics.	Keenan J.H. Thermodynamics. McLuhan M. Understanding media. Rasmussen S.E. Towns and buildings. Ullman B.L. Ancient writing and its influence.

Fig. 10. Arrangement of entries for the same books in a classified, alphabetical subject and author order.

Characteristic of division	Types of bibliography	Types of indexed materials
Content and type of indexed materials	General	All literature irrespective of content and type
	Special	Literature of specific content or type
	Branch-oriented	Literature in a specific subject field
Purpose	Enumerative	The widest possible coverage by content or type
	Information	Literature for specialists in a specific subject field, theme or problem
	Recommendatory	Selected literature for a specific reader category
Time of publication	Current	Currently published materials
	Retrospective	Literature published during a certain time period
	Prospective	Literature prepared for publication
Place of publication	International	Literature appearing in different languages in different countries
	National	Literature appearing in one language or country
	State	Literature published in one country
	Regional	Literature published in one locality
Level	Primary	Published works
	Secondary	Primary bibliographies
	Tertiary	Secondary bibliographies (bibliographies of bibliographies)
Method of grouping of entries	Classified	According to a scientific system of knowledge
	Subject	Alphabetically by subjects
	Author	Alphabetically by authors and titles

Fig. 11. Basic types of bibliography.



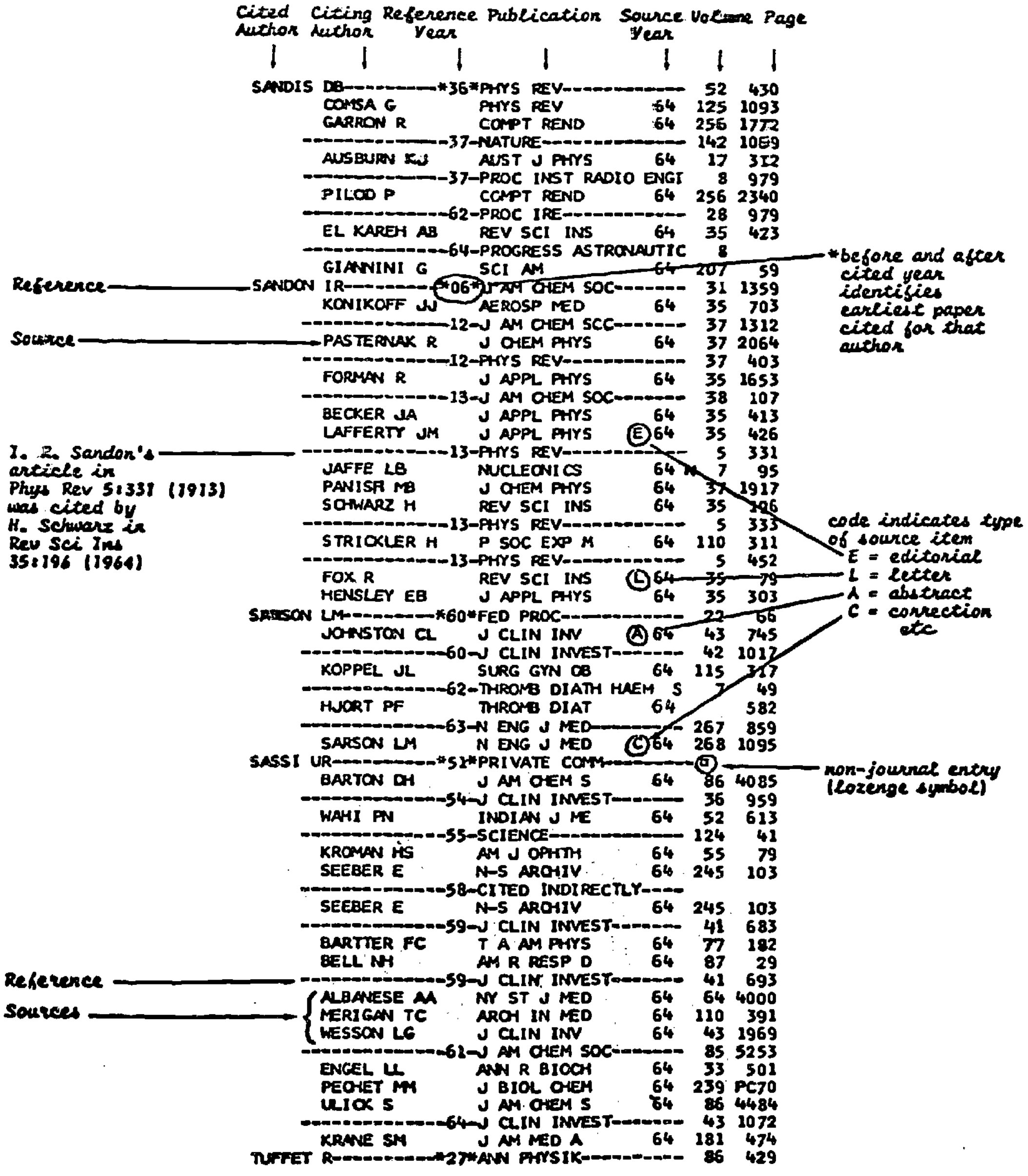
```
ION FROM THE +          CORTISONE ON SULFANIL AMIDE PENETRAT   PEGT-10-03-072
    OLIGOHYDRAMNIOS AND CORTISONE-INDUCED CLEFT PALATE.=       NATU-0203-0533
RICTION ON +            CORTISONE, AMPUTATION, AND FOOD REST   PSEB-0116-0718
         KINETICS OF    CORUNDUM DEFORMATION ON HEATING.=      DANK-0157-0331
 DRUGS AND THEIR USE IN COSMETICS.=                      DYE   RSNR-0014-0243
CAL BLEACHING AGENTS IN COSMETICS.=              USE OF OPTI   RSAR-0014-0245A
Y NUCLEI IN THE PRIMARY COSMIC RADIATION.=         + OF HEAV   PRVB-0135-0724

    INTERNAL STRESS AND DEFORMATION OF GLAS  PLASTIC TYPE AG   PLMS-64-07-062
 ALS PRODUCED BY A CONI+ DEFORMATION OF THE COPPER MONO CRYST  CORE-0258-6418
  ELASTIC NEUTRON-DEUTE+ DEFORMATION OF THE DEUTRON DURING AN  ZEPY-0180-0001
   KINETICS OF CORUNDUM DEFORMATION ON HEATING.=               DANK-0157-0331
    CRYSTALLIZATION AND DEFORMATION PROCESSES IN HIGH POLYME   SPET-0004-0157
   EXTENSION OF PLASTIC DEFORMATION.=                          MTOM-64-07-011
      STRUCTURE ON THE DEFORMATIONAL THERMAL STABILITY OF      PLMS-64-07-006

UCTURE IN ELECTRON-BEAM HEATING OF STEEL.= + THE INITIAL STR   DUKR-1964-0919
OF LIQUID SODIUM ON THE HEATING SURFACE WITH NATURAL CONVECT   INFZ-0007-0008
L PROPERTI+INFLUENCE OF HEATING TEMPERATURE ON THE MECHANICA   MTOM-64-07-007
CORUNDUM DEFORMATION ON HEATING.=                KINETICS OF   DANK-0157-0331
OF HEATER AND METHOD OF HEATING.=  + QUALITY OF FATS.  TYPE    JAOC-0041-0531
UM BETWEEN 0.35 +ATOMIC HEATS OF CESIUM, RUBIDIUM, AND LIYHT   PRVA-0135-0671
           OF INTEGRAL HEATS OF EVAPORATED LIQUID MIXTURE      IVUK-0007-0335

 DURING +   BALANCE AND KINETICS OF ANAEROBIC ENERGY RELEASE   JAPY-0019-0623
 THE DECOMPOSITION OF + KINETICS OF CHEMILUMINESCENCE DURING   BIOF-0009-0528
TION AND AGEING OF +    KINETICS OF CONTINUOUS STRESS-RELAXA   KOZE-0197-0080
HEATING.=               KINETICS OF CORUNDUM DEFORMATION ON    DANK-0157-0331
G OXIDATIVE PHOSPHORYL+ KINETICS OF CYTOCHROME B AND C DURIN   ABB -0107-0078
IDE FILMS., ALUM+GROWTH KINETICS OF DISCONTINUOUS THERMAL OX   JESO-0111-0903
S INFLUENZAE.=          KINETICS OF DNA UPTAKE BY HAEMOPHILU   JGHI-0035-0391
```

Fig. 12. Layout of entry in a permuted-title index (KWIC Index).



```
Cited   Citing  Reference Publication      Source Volume Page
Author  Author     Year                     Year
  |       |         |        |               |     |     |

SANDIS DB----------*36*PHYS REV------------     52  430
       COMSA G         PHYS REV            64  125 1093
       GARRON R        COMPT REND          64  256 1772
       -------------37-NATURE--------------    142 1069
       AUSBURN KJ      AUST J PHYS         64   17  312
       -------------37-PROC INST RADIO ENGI      8  979
       PILOD P         COMPT REND          64  256 2340
       -------------62-PROC IRE------------     28  979
       EL KAREH AB     REV SCI  INS        64   35  423
       -------------64-PROGRESS ASTRONAUTIC      8
       GIANNINI G      SCI AM              64  207   59
SANDON IR---------*06*J AM CHEM SOC-------      31 1359
       KONIKOFF JJ     AEROSP MED          64   35  703
       -------------12-J AM CHEM SCC-------     37 1312
       PASTERNAK R     J CHEM PHYS         64   37 2064
       -------------12-PHYS REV------------     37  403
       FORMAN R        J APPL PHYS         64   35 1653
       -------------13-J AM CHEM SOC-------     38  107
       BECKER JA       J APPL PHYS         64   35  413
       LAFFERTY JM     J APPL PHYS       Ⓔ64   35  426
       -------------13-PHYS REV------------      5  331
       JAFFE LB        NUCLEONICS          64    7   95
       PANISH MB       J CHEM PHYS         64   37 1917
       SCHWARZ H       REV SCI  INS        64   35  196
       -------------13-PHYS REV------------      5  333
       STRICKLER H     P SOC EXP M         64  110  311
       -------------13-PHYS REV------------      5  452
       FOX R           REV SCI  INS      Ⓛ64   35   79
       HENSLEY EB      J APPL PHYS         64   35  303
SARSON LM----------*60*FED PROC------------     22   66
       JOHNSTON CL     J CLIN INV        Ⓐ64   43  745
       -------------60-J CLIN INVEST-------     42 1017
       KOPPEL JL       SURG GYN OB         64  115  317
       -------------62-THROMB DIATH HAEM  S      7   49
       HJORT PF        THROMB DIAT         64       582
       -------------63-N ENG J MED---------    267  859
       SARSON LM       N ENG J MED       Ⓒ64  268 1095
SASSI UR-----------*51*PRIVATE COMM---------◊
       BARTON DH       J AM CHEM S         64   86 4085
       -------------54-J CLIN INVEST-------     36  959
       WAHI PN         INDIAN J ME         64   52  613
       -------------55-SCIENCE-------------    124   41
       KROMAN HS       AM J OPHTH          64   55   79
       SEEBER E        N-S ARCHIV          64  245  103
       -------------58-CITED INDIRECTLY----
       SEEBER E        N-S ARCHIV          64  245  103
       -------------59-J CLIN INVEST-------     41  683
       BARTTER FC      T A AM PHYS         64   77  182
       BELL NH         AM R RESP D         64   87   29
       -------------59-J CLIN INVEST-------     41  693
      {ALBANESE AA     NY ST J MED         64   64 4000
      {MERIGAN TC      ARCH IN MED         64  110  391
      {WESSON LG       J CLIN INV          64   43 1969
       -------------61-J AM CHEM SOC-------     85 5253
       ENGEL LL        ANN R BIOCH         64   33  501
       PECHET MM       J BIOL CHEM         64  239 PC70
       ULICK S         J AM CHEM S         64   86 4484
       -------------64-J CLIN INVEST-------     43 1072
       KRANE SM        J AM MED A          64  181  474
TUFFET R-----------*27*ANN PHYSIK----------     86  429
```

Reference

Source

I. R. Sandon's article in Phys Rev 5:331 (1913) was cited by H. Schwarz in Rev Sci Ins 35:196 (1964)

Reference

Sources

*before and after cited year identifies earliest paper cited for that author

code indicates type of source item
E = editorial
L = letter
A = abstract
C = correction
etc

non-journal entry (lozenge symbol)

Fig. 13. Portion of a page in the Science Citation Index.

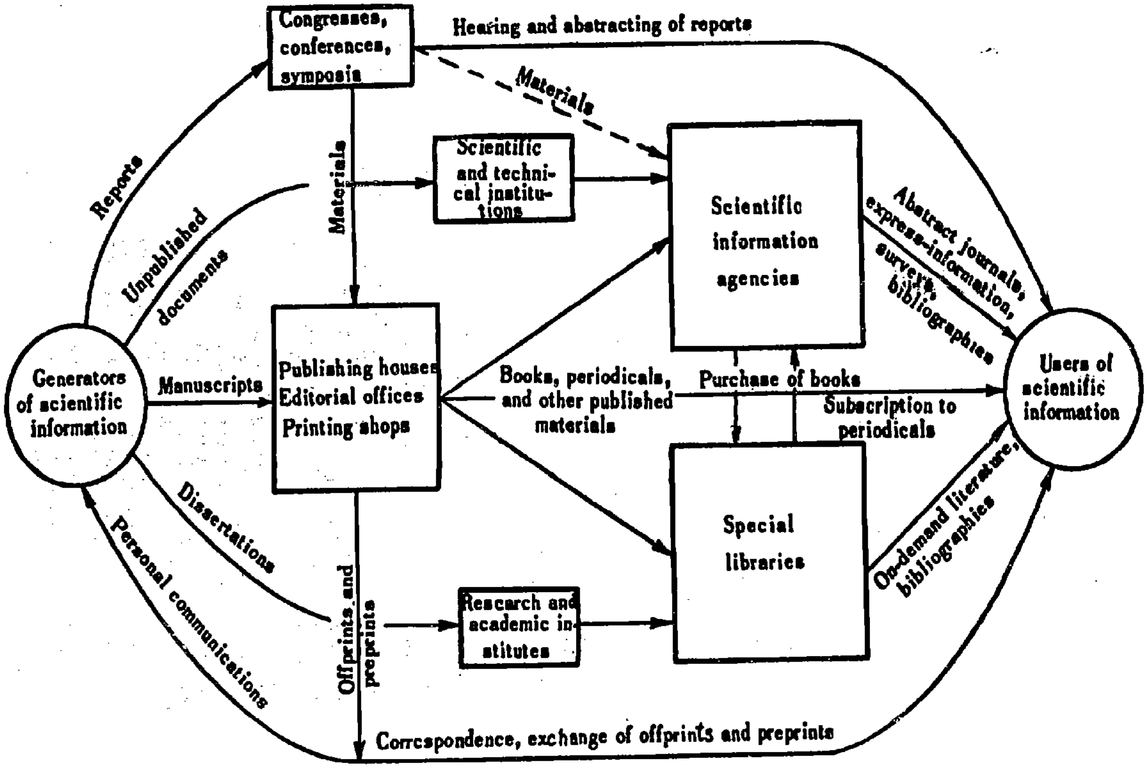


Fig. 14. Channels of dissemination of scientific information from the generators to the users.

The diagram is based on the lecture "Effektivnost' deyatel'nosti organov nauchnotekhnicheskoy informatsii" (Effectiveness of activities of information agencies, by M.L. Kolchinsky (Moscow, VINITI, 1966, 29 p.).

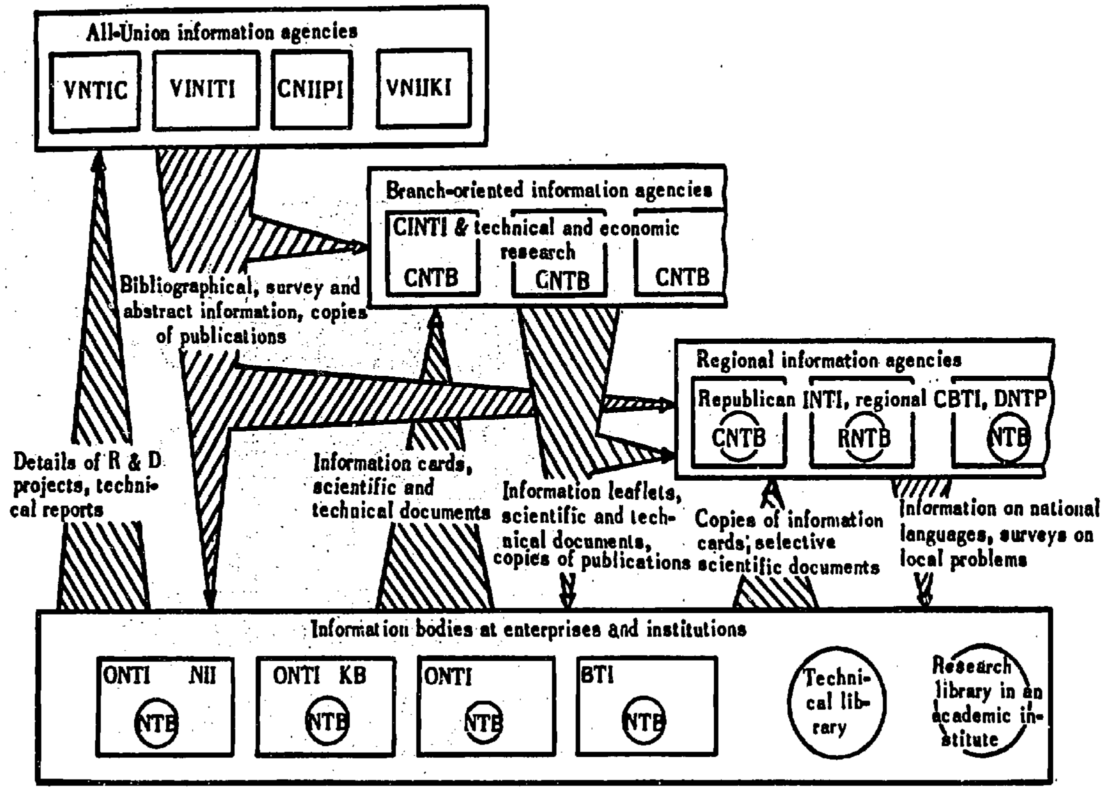


Fig. 15. Distribution of scientific and technical information flows among the information agencies in the USSR.

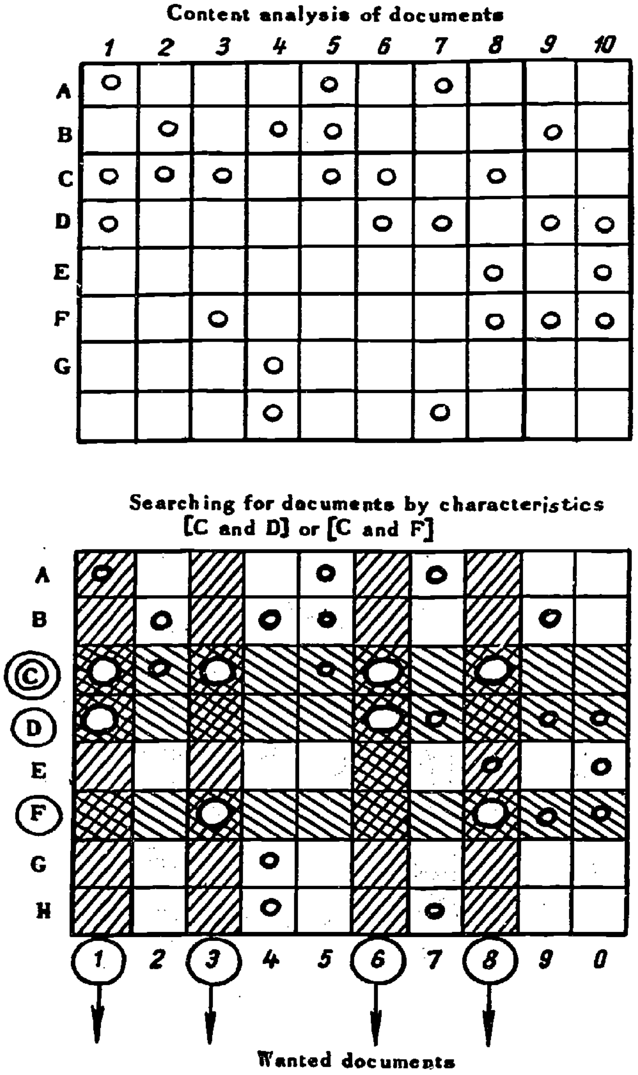


Fig. 16. Diagram of information retrieval.

Used on the paper by B.-A. Lipetz "Information storage and retrieval" in "Scientific American", 1966, vol. 215, No. 3, p. 226.

Documents	Relevant	Non-relevant	
Retrieved	a	b	$a+b$
Non-retrieved	c	d	$c+d$
	$a+c$	$b+d$	

Recall $\frac{a}{a+c}\cdot 100\%$

Precision $\frac{a}{a+b}\cdot 100\%$

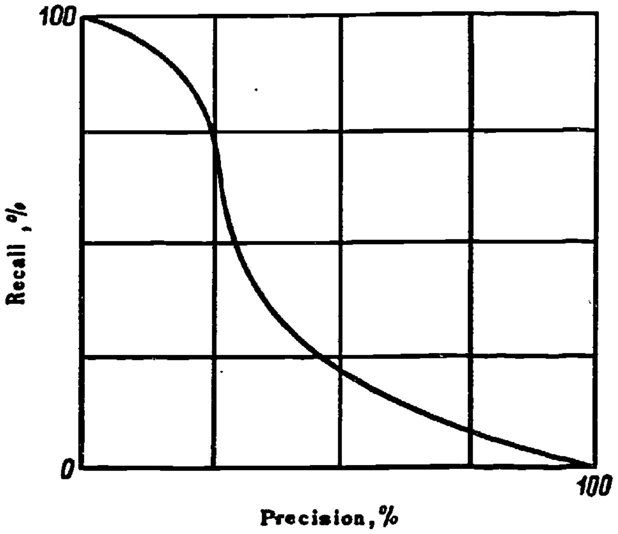


Fig. 17. Dependence of recall on precision (after C. Cleverdon).

C.W. Cleverdon and J. Mills. The testing of index language devices. - "Aslib Proceedings", 1963, v. 15, No. 4, p. 106-130.

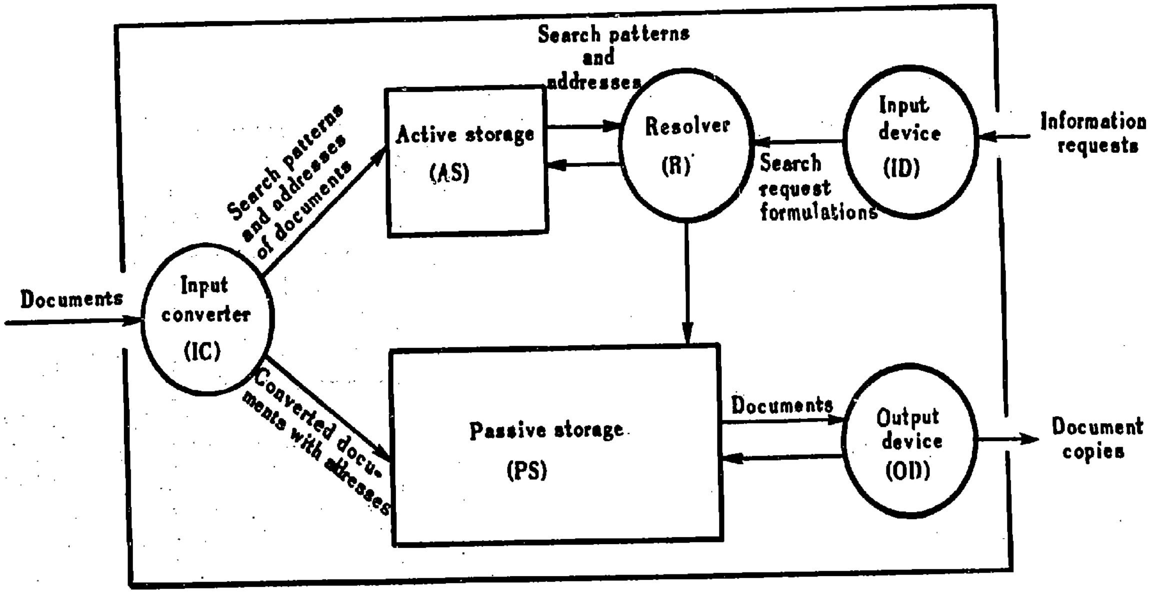
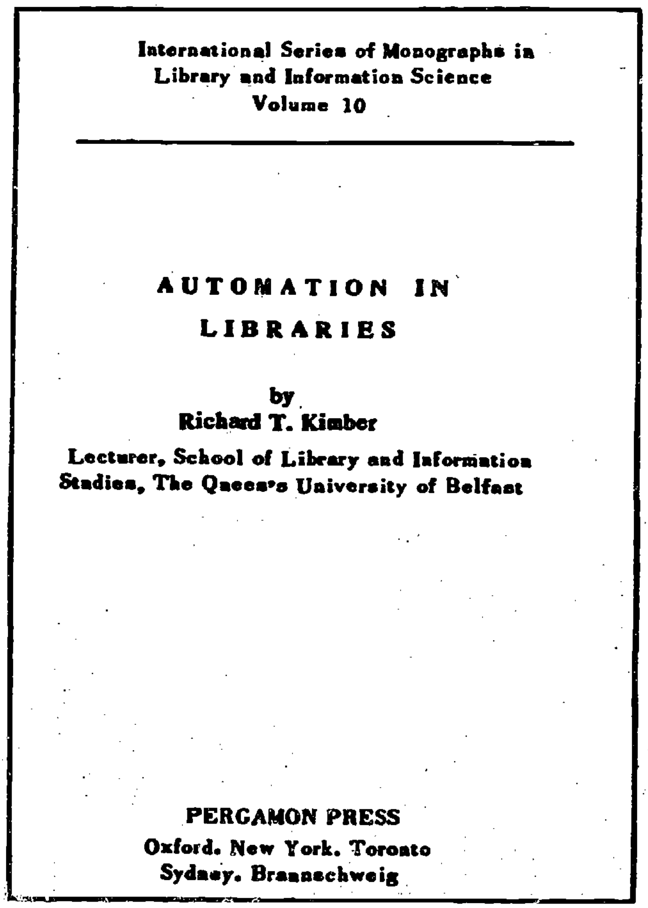


Fig. 18. General block-diagram of an information retrieval system.

Based on the paper, by G.E. Vleduts, "O nekotorykh storonakh issledovaniy po sozdaniyu informatsionno-poiskovykh sistem" (Some aspects of research in the development of information retrieval systems) in "Nauchno-Tekhnicheskaya Informatsiya", 1961, No. 1, p. 32.



Specimen title-page	Bibliographical particulars
International Series of Monographs in Library and Information Science Volume 10	Information before the title
AUTOMATION IN LIBRARIES	Title
by Richard T. Kimber Lecturer, School of Library and Information Studies, The Queen's University of Belfast	Subtitle information
PERGAMON PRESS Oxford. New York. Toronto Sydney. Braunschweig	Imprint

Specimen entry

Kimber R.T.
Automation in libraries. By Richard T. Kimber
Oxford /a.o./, Pergamon, /1970/.
VIII 140 p., fig. (Int. ser. of monographs in libr. and information science. Vol. 10).
Bibliogr. p. 134–138.

Fig. 19. Arrangement of bibliographical particulars on the title page and in the entry.

Publication types			Entries	
			Main	Added
Authorship		1. Works of one, two or three authors	Surname (forename) of first author	Added entries under second and third authors
		2. Works of four or more authors	Title	Added entry under first author
		3. Collections of works of different authors		Added entry under compiler (editor)
		4. Official publications and publications with standard title	Name of institution (corporate author)	Added entry under title
Number of volumes frequency of appearance	Books	5. Multivolume publications	Series entry for the whole publication	Entries for individual volumes
		6. Serial publications	Entries for each issue in a series	Series entry for the whole publication
	Periodicals	7. Continued publications (proceedings, etc.)	Series entry for the whole publication under the general title	Entries for separate volumes under the author and particular title
		8. Journals, magazines		Entries for individual issues under the general title
		9. Newspapers		
Polygraphical autonomy		10. Parts of publications (articles from collections and periodicals)	Analytical entry under author or title of the article	-
		11. Parts (chapters) of publications	Analytical entry under author or title of the whole work	-
		12. Reviews	Entry under author or title of the reviewed work	Entry under the author or title of the review

Fig. 20. Scheme of rules for description of the basic types of publications.

1	Tripathi S.M. Modern cataloguing. Theory and practice. Agra, Agarwala, /1969/. XII, 289 p. Bibliogr. p. 287-289.
2	A Handbook of the mosquitoes of the Southeastern United States. Washington, U.S. Gov. print. off., 1960. 188 p., ill. Aut.: W.V. King, G.H. Brandley, C.N. Smith and W.C. McDuffie.
3	Computer based information retrieval systems. Ed. by B. Houghton. London, Bingley, 1968. 136 p., ill.
4	Unesco. General catalogue of Unesco publications and Unesco sponsored publications. 1946-1959. Paris, UNESCO, 1962. XVI, 217 p.
5	International forum on informatics. Volume of papers. Vol. 1-2. Moscow, 1969. (All-Union Inst. for Scient. and Techn. Information). Vol. 1. 656 p. Vol. 2. 603 p.
6	Kimber R.T. Automation in libraries. Oxford /a.o./, Pergamon, 1970. VIII, 140 p., fig. (Intern. ser. of monographs in libr. and information science. Vol. 10)
7	Royal Society of Edinburgh. Transactions... Edinburgh-London, 1954-. Vol. 62 No. 1-2 1954 Vol. 63 No. 1-3 1955-1959 Vol. 64 No. 1-18 1959-1961
8	Unesco Bulletin for Libraries. Vol. 23-24. Paris, 1969-1970. 1969 Vol. 23, No. 1-6 1970 Vol. 24, No. 1-3
9	Trud, Moscow. 1967- 1967 1968
10	Price D.J. de S. Networks of scientific papers. - "Science", 1965, vol. 149, No. 3683, p. 514.
11	Coates E.J. Subject catalogues. Headings and structure. London, The Libr. Assoc., 1960. Chap. 13. Use and search strategy. p. 159-173.
12	On theoretical problems of informatics. Moscow, All-Union Inst. for Scient. and Techn. Information, 1969. Review: Brookes B.C. - "Journal of Documentation", 1969, vol. 25, No. 3, p. 262-264.

Fig. 21. Specimen index entries.
(Serial numbers correspond to the chart in Fig. 20)

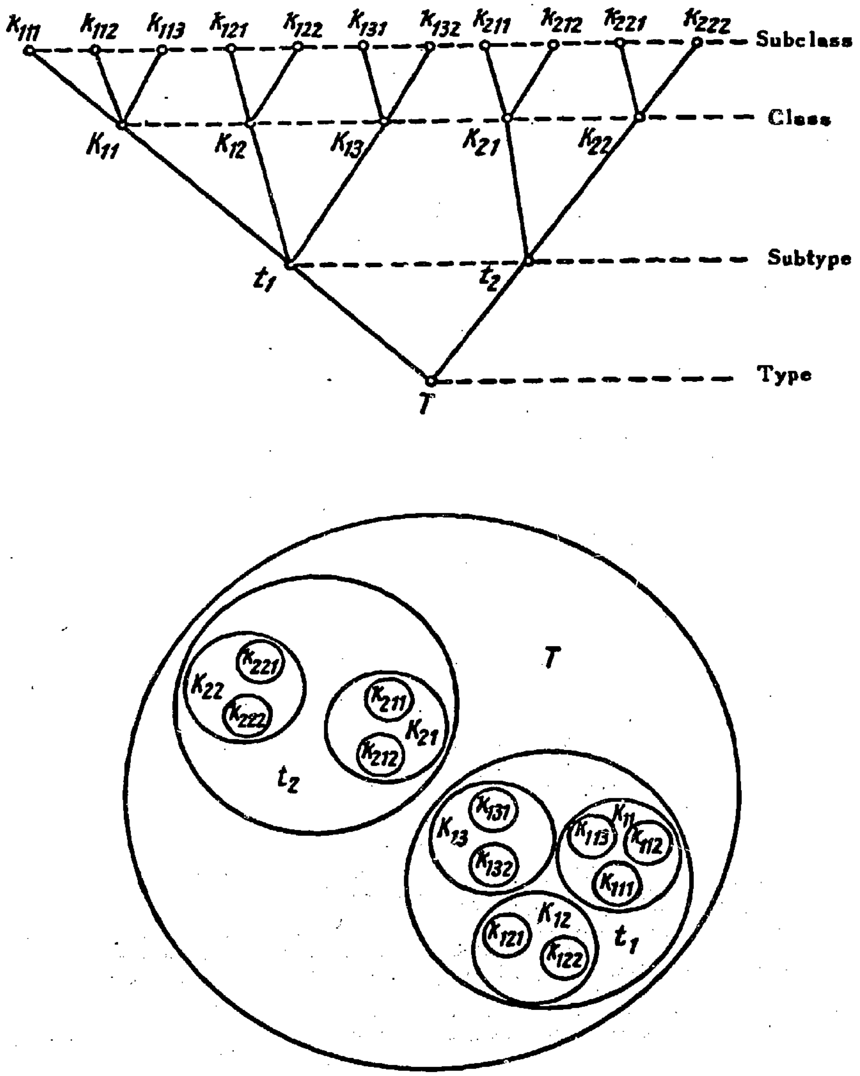


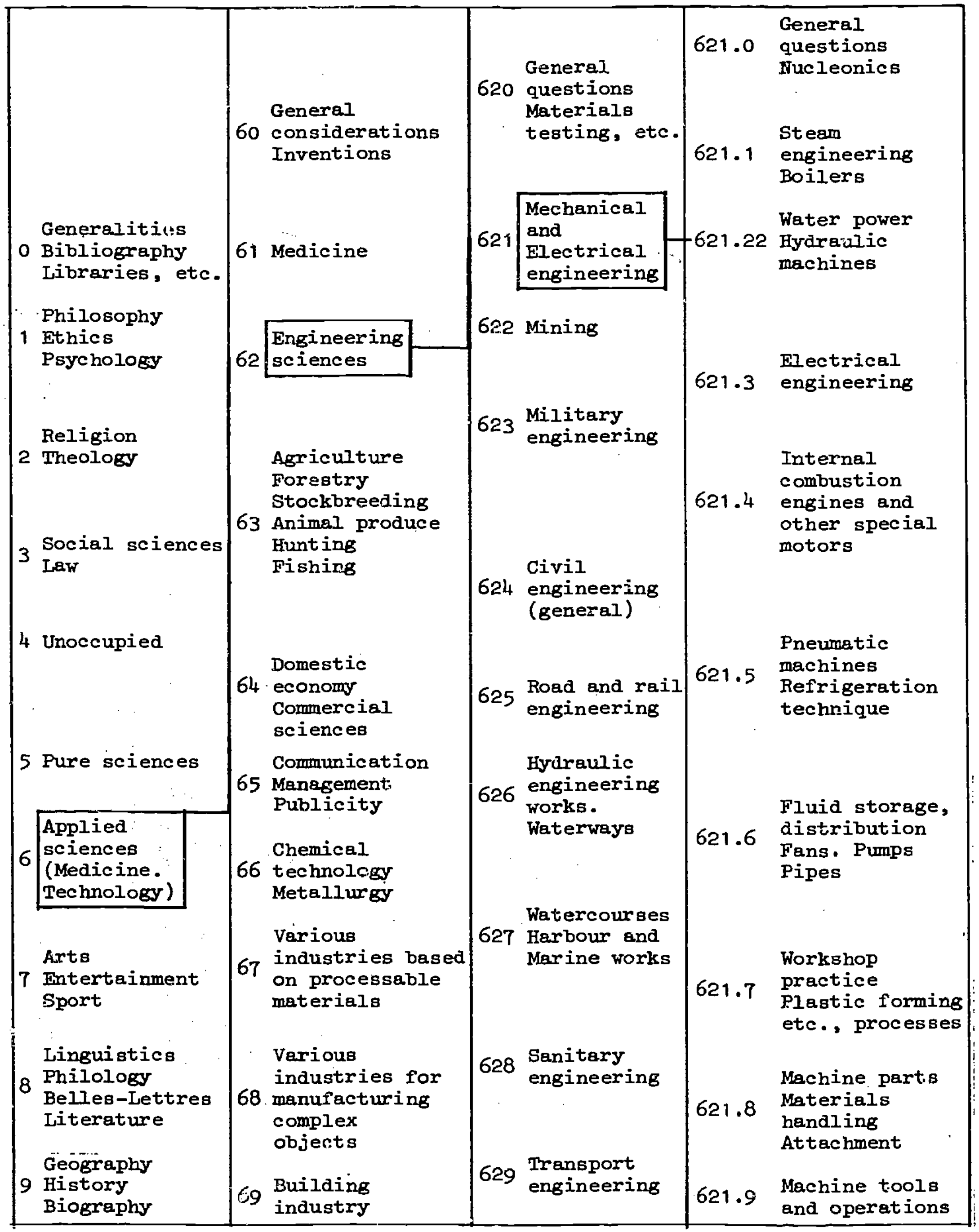
Fig. 22. Graph of an hierarchical classification and the relationship between its divisions (represented by means of Euler diagrams)

0	GENERALITIES OF KNOWLEDGE
00	Prolegomena. Fundamentals of knowledge and culture
01	Bibliography. Catalogues and booklists
02	Librarianship and reading
03	Encyclopaedias. Dictionaries. Reference books
04	Pamphlets, cuttings, contributions, summary publications
05	Periodicals. Reviews. Serial publications
06	Organizations. Associations. Conventions. Museums
07	Newspapers. The press. Journalism
08	Collective publications and documents of special form
09	Manuscripts. Rare, remarkable books and publications
1	PHILOSOPHY. METAPHYSICS. PSYCHOLOGY. LOGIC. ETHICS
11	Metaphysics and ontology
12	Special metaphysics
13	Spiritual life. Cultural and moral systems, spheres
14	Philosophical systems, speculations and theories
15	(Unoccupied, except for 159.9)
16	Logic and reasoning. Theory of knowledge
17	Ethics. Morals. Convention
18	Aesthetics and taste (in general)
19	History of philosophy
2	RELIGION. THEOLOGY
21	Natural theology
22	Holy Scripture. The Bible
23	Dogmatic theology. Apologetics
24	Practical theology. The religious life. Christian living
25	Pastoral theology
26	The Christian Church in general
27	Christian Church history
28	Christian Churches, worshipping bodies and sects
29	Oriental and other non-Christian religions and cults
3	SOCIAL SCIENCES
30	Sociology. Sociography
31	Statistics. Demography
32	Political science. Politics. Current affairs
33	Political economy. Economics
34	Jurisprudence. Law. Legislation
35	Public administration. Government. Defence
36	Social relief, welfare, movements. Insurance
37	Education. Teaching. Training. Schools
38	Commerce. Trade. Communications and transport
39	Ethnography. Customs and rites. Folklore
4	(Unoccupied, since 1962)

Fig. 23. Main divisions in the Universal Decimal Classification (UDC).

5	MATHEMATICS AND NATURAL SCIENCES
51	Mathematics
52	Astronomy. Geodesy and surveying
53	Physics and mechanics
54	Chemistry
55	Geology and associated sciences. Meteorology
56	Palaeontology. Fossils
57	Biological sciences
58	Botany. Plant biology and taxonomy
59	Zoology. Animal biology and taxonomy
6	APPLIED SCIENCES. MEDICINE. TECHNOLOGY
61	Medical sciences. Health and safety
62	Engineering and technology generally
63	Agriculture. Forestry. Livestock. Fisheries
64	Domestic science and economy. Household management
65	Management. Organization of industry, business, communication and transport
66	Chemical industry and technology
67	Industries and crafts based on processable materials
68	Specialized trades and industries for finished articles
69	Building: materials, construction, trades
7	THE ARTS. RECREATION. ENTERTAINMENT. SPORT
71	Physical planning. Landscape, parks and gardens
72	Architecture
73	Sculpture and the plastic arts
74	Drawing and minor (decorative) arts or crafts
75	Painting
76	Engraving, prints and (commercial) graphic art
77	Photography and cinematography
78	Music
79	Entertainment. Pastimes. Games. Sport
8	LINGUISTICS. PHILOLOGY. BELLES-LETTRES. LITERATURE
80	General linguistics and philology
82/89	Literatures of individual languages
9	GEOGRAPHY. BIOGRAPHY. HISTORY
91	Geography. Exploration. Travel
92	Biography
93	History
94/99	Mediaeval and modern history
	SYNOPSIS OF AUXILIARIES (IN ORDER OF FILING)
a.	Aggregation signs + and / (preceding the simple class number)
b.	Relation signs : and [ ]
c.	Language auxiliaries = ...
d.	Form and presentation auxiliaries (0...)
e.	Place, region, country auxiliaries (1/9)
f.	Race, people, nationality auxiliaries (=...)
g.	Time, date, period auxiliaries "..." (alternatively before e.)
h.	Alphabetical and numerical (non-UDC) auxiliaries A/Z, No. 1 to ...
i.	Point of view subdivisions .00...
k.	Special auxiliaries -... and .0...

Fig. 24. Main divisions (continued), and synopsis of auxiliaries in the UDC.



0 Generalities Bibliography Libraries, etc.
1 Philosophy Ethics Psychology
2 Religion Theology
3 Social sciences Law
4 Unoccupied
5 Pure sciences
6 Applied sciences (Medicine. Technology)
7 Arts Entertainment Sport
8 Linguistics Philology Belles-Lettres Literature
9 Geography History Biography

60 General considerations Inventions
61 Medicine
62 Engineering sciences
63 Agriculture Forestry Stockbreeding Animal produce Hunting Fishing
64 Domestic economy Commercial sciences
65 Communication Management Publicity
66 Chemical technology Metallurgy
67 Various industries based on processable materials
68 Various industries for manufacturing complex objects
69 Building industry

620 General questions Materials testing, etc.
621 Mechanical and Electrical engineering
622 Mining
623 Military engineering
624 Civil engineering (general)
625 Road and rail engineering
626 Hydraulic engineering works. Waterways
627 Watercourses Harbour and Marine works
628 Sanitary engineering
629 Transport engineering

621.0 General questions Nucleonics
621.1 Steam engineering Boilers
621.22 Water power Hydraulic machines
621.3 Electrical engineering
621.4 Internal combustion engines and other special motors
621.5 Pneumatic machines Refrigeration technique
621.6 Fluid storage, distribution Fans. Pumps Pipes
621.7 Workshop practice Plastic forming etc., processes
621.8 Machine parts Materials handling Attachment
621.9 Machine tools and operations

Fig. 25. Deciphering of a UDC number.

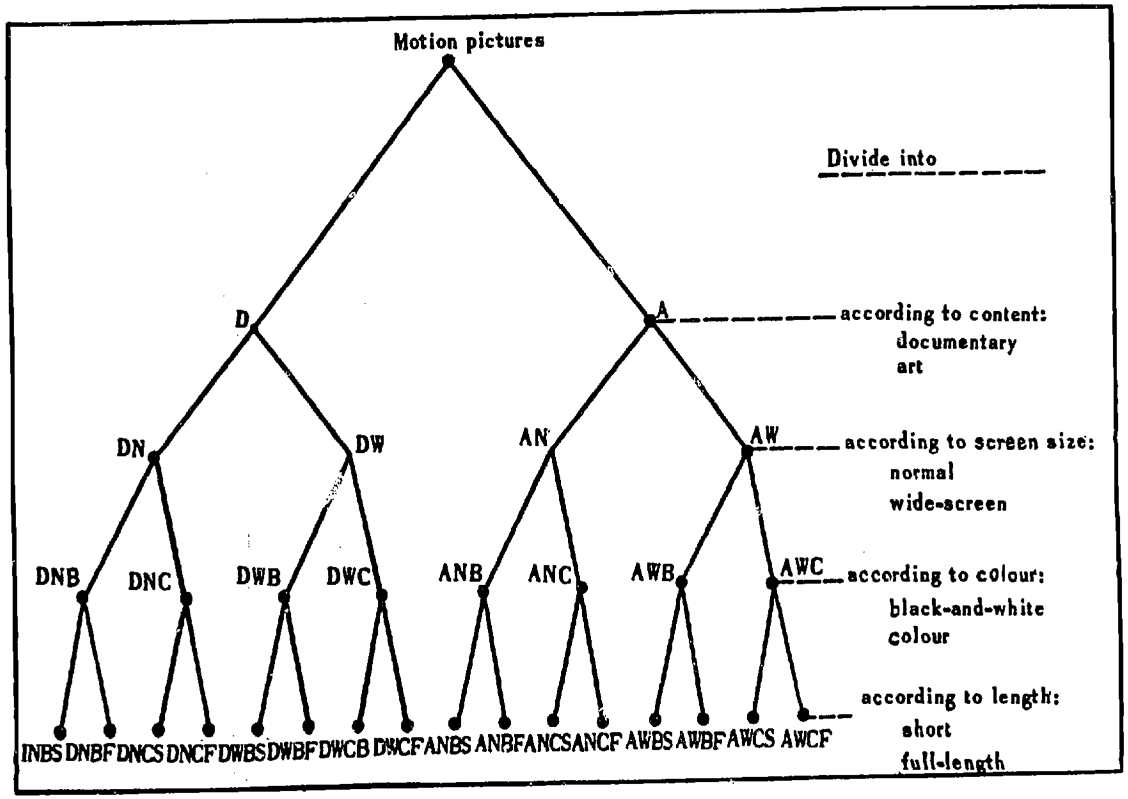


Fig. 26. A hypothetical classification of objects according to attributes which do not form a natural hierarchy.

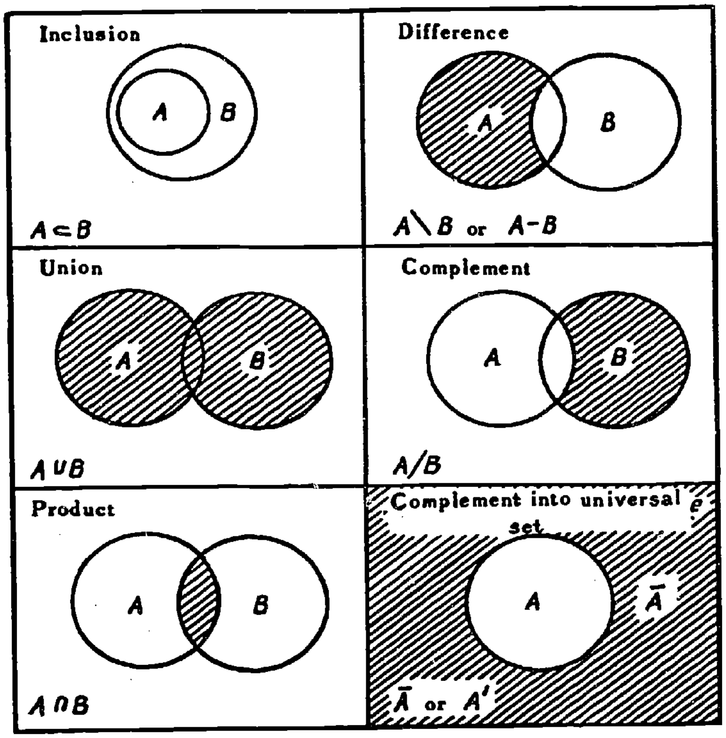


Fig. 27. Graphical representation of relations between sets (by means of Venn diagrams).

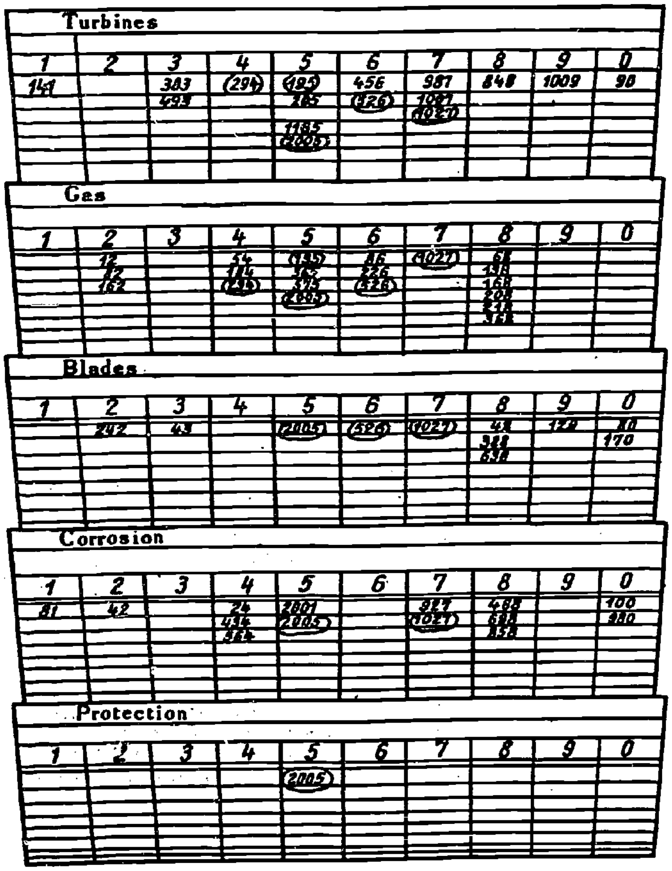


Fig. 28. Working principle of the Uniterm System.

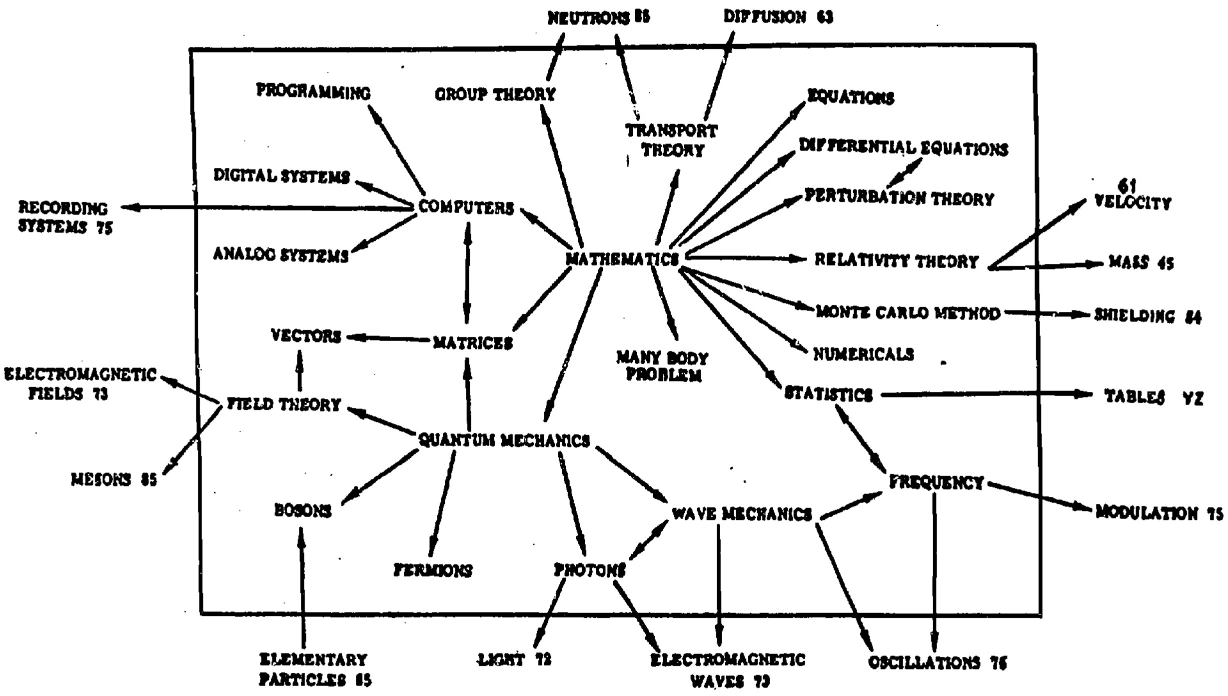


Fig. 29 Specimen layout of a subject class in a thesaurus.

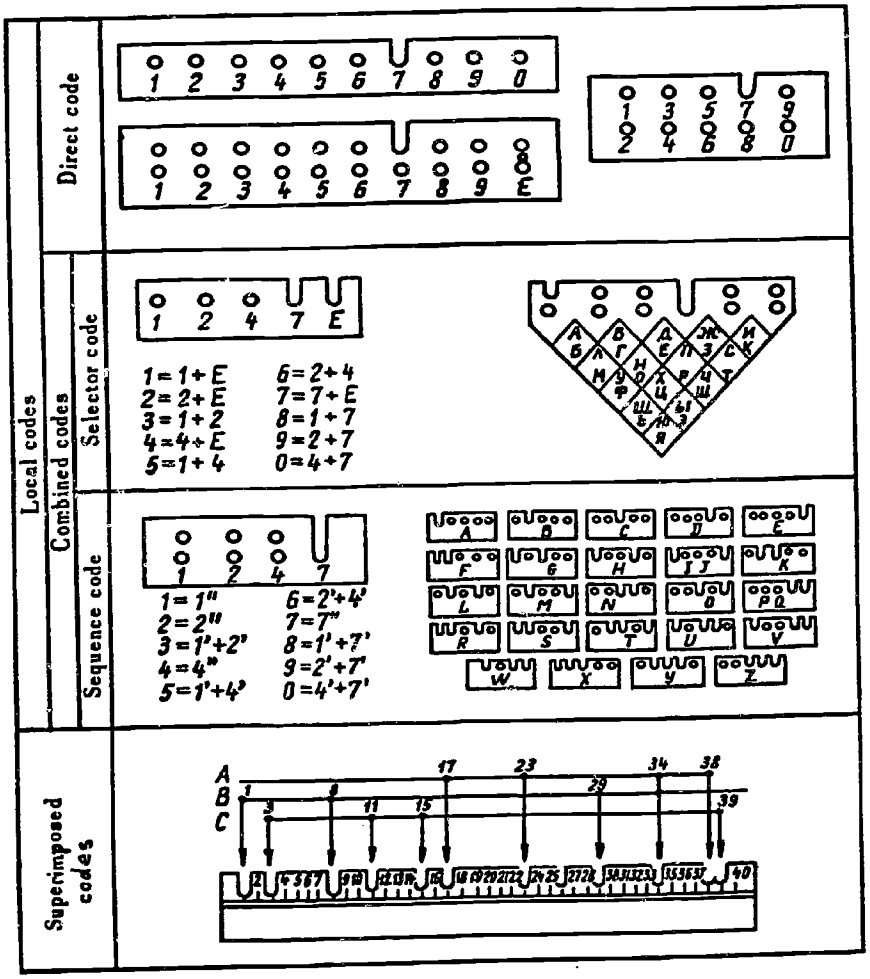


Fig. 30. Examples of codes for edge-punched cards.

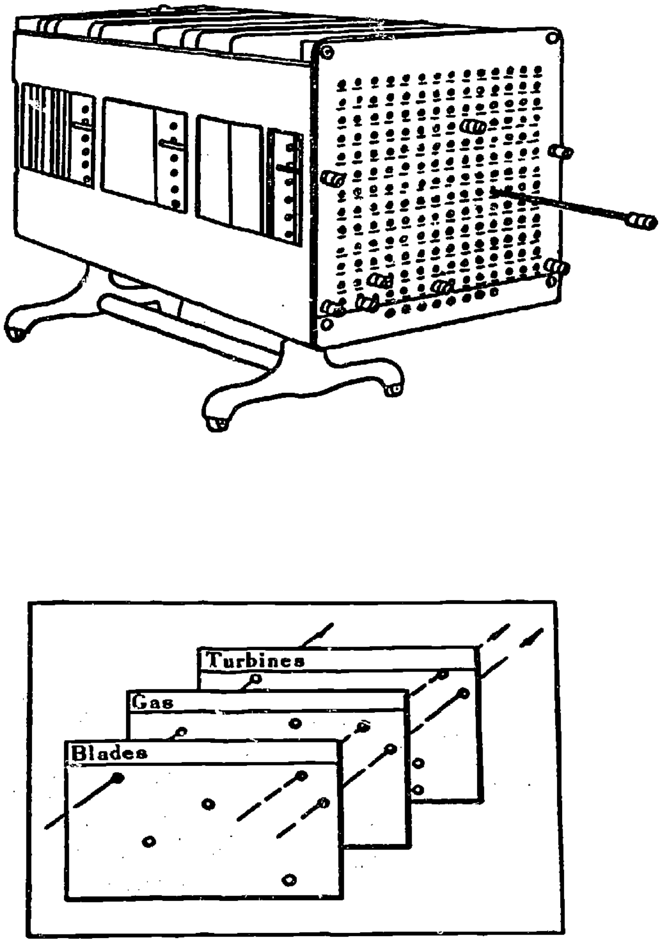


Fig. 31. The working principle of body-punched (top) and superimposable (bottom) cards.

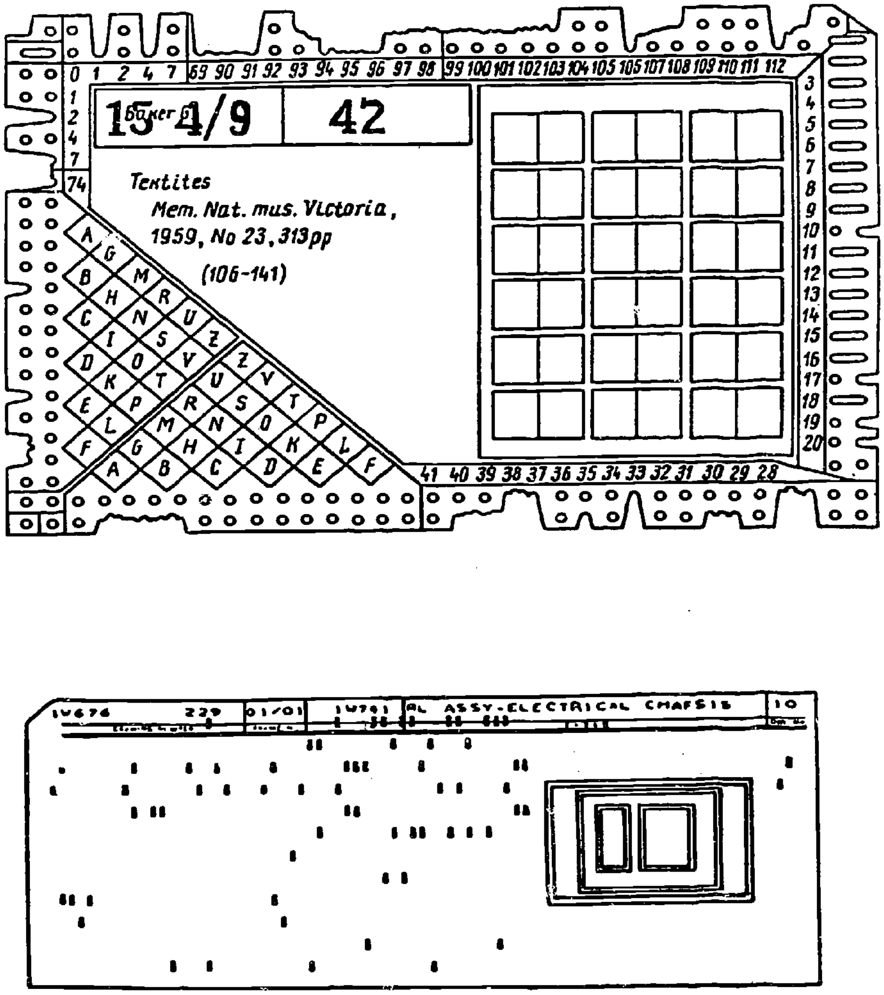
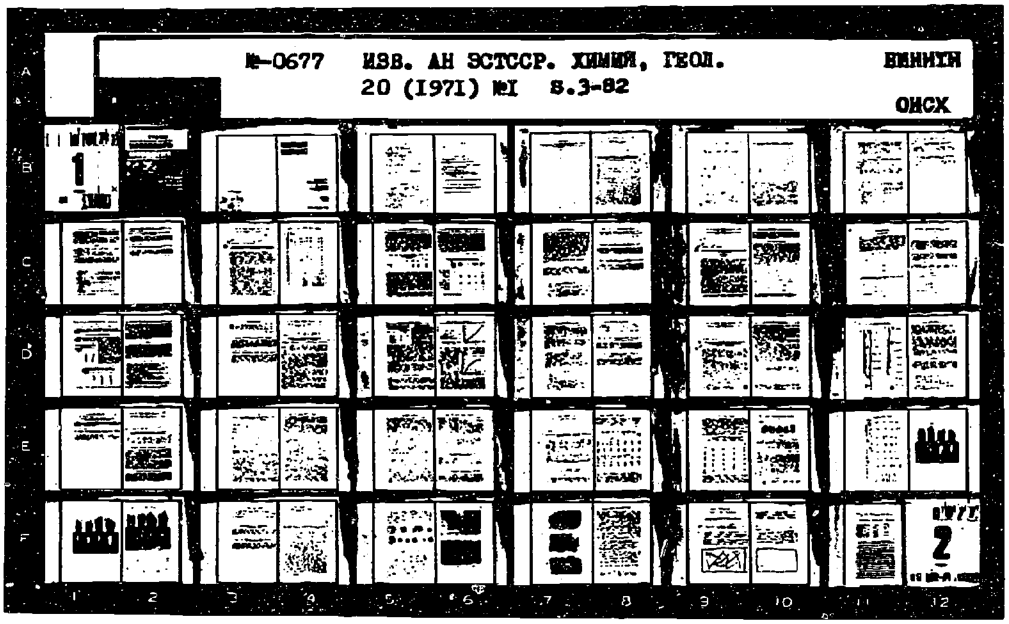


Fig. 32. Aperture cards (edge-punched card and mechanically sorted card).



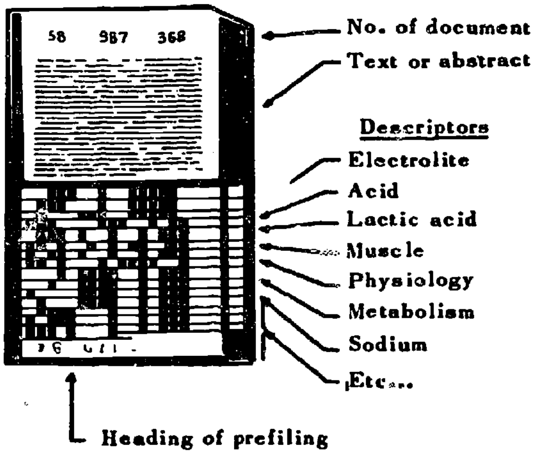


Fig. 33. Microfiches (specimens).
Top: Standard microfiche 105 x 145 mm
Bottom: Filmorex card

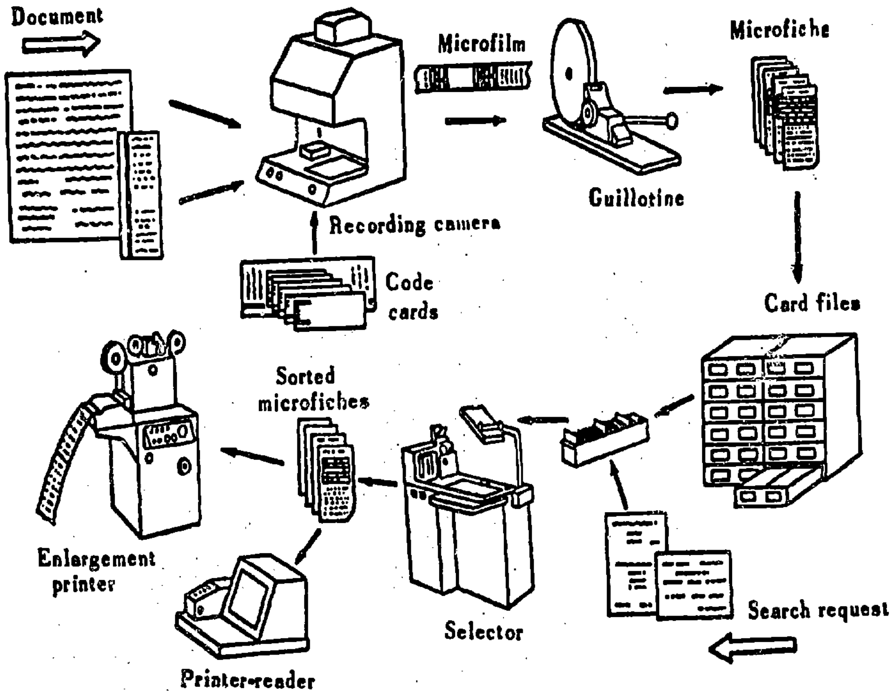


Fig. 34. Schematic operation and main units of Filmorex system.

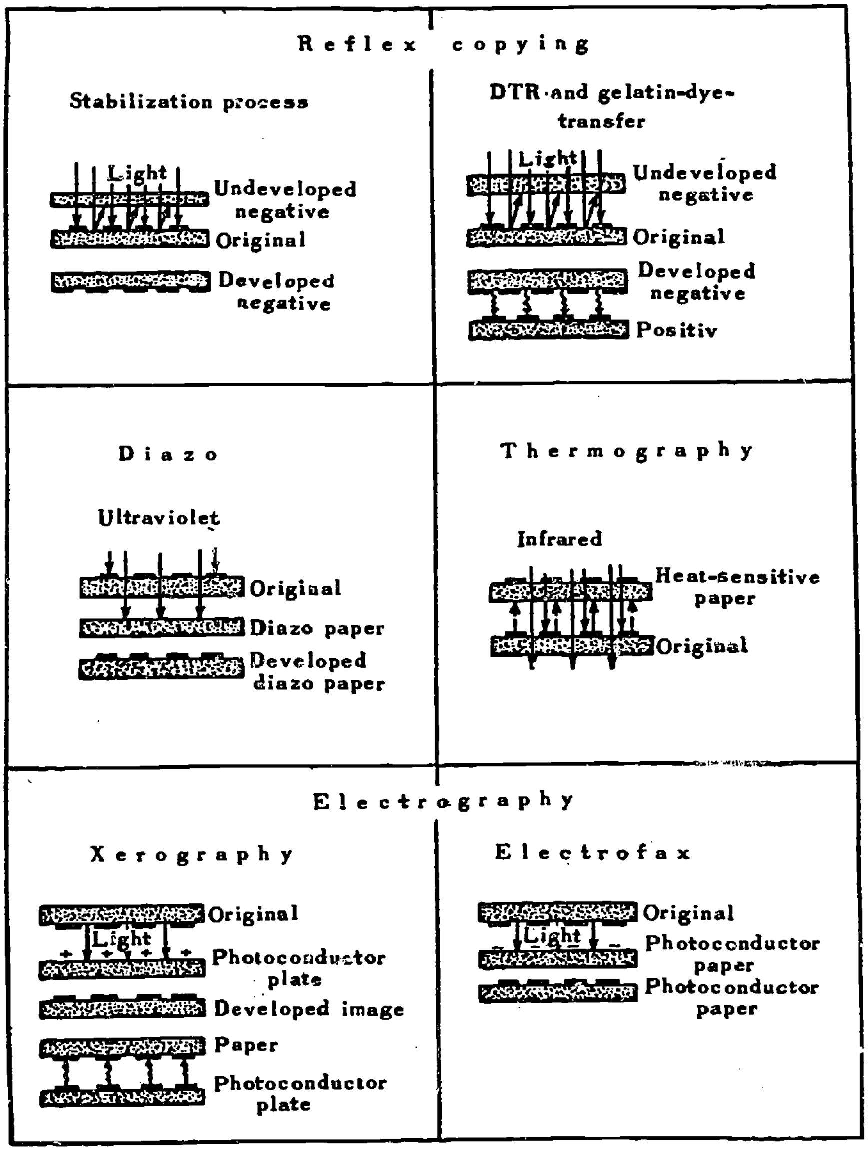


Fig. 35. Schematics of the basic copying processes.

Advantages	Disadvantages
Reflex copying	
High-contrast copies from line and half-tone (excluding Verifax) originals of any colour (except yellow)	Copies require drying up and tend to darken or fade out with time. Their cost is rather high.
Diazo	
Lasting and most cheap copies from originals of any colour (except blue and violet).	Requires translucent original with image on one side and involves the use of highly odorous chemicals.
Thermography	
The fastest and easiest method of copying of letterpress or typescript images and line drawings. Dry process in lighted room.	Does not copy inscriptions in writing ink and other organic dyes. Copies will darken and become brittle in time.
Xerography	
Lasting copies of excellent quality from textual and line originals of any colour on standard stationery and many other materials. Dry processes in lighted room.	Poor copies from half-tone illustrations and originals with large solid areas. Expensive equipment requiring skilled personnel to operate.
Electrofax	
The basic merits of xerography, plus quality copies from illustrations and originals with large solid areas.	Requires special stationery, from which the image may be easily scratched with metal objects. Some apparatus involve wet processes.

Fig. 36. Relative advantages and disadvantages of the different copying processes.

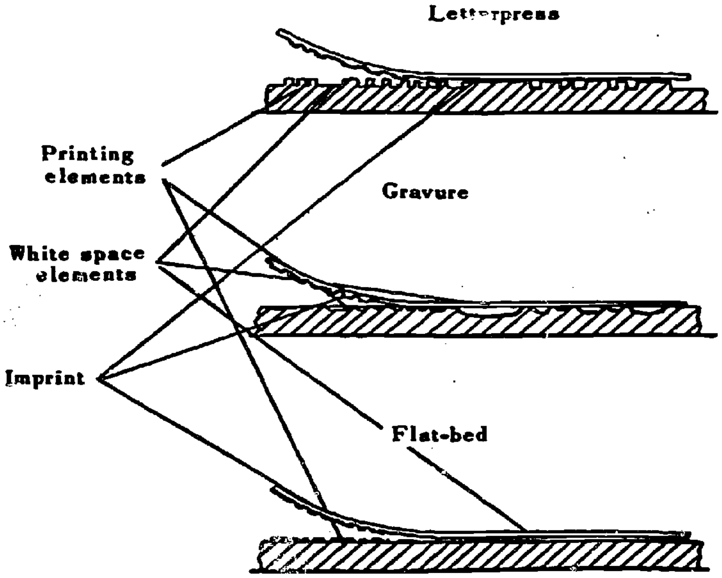


Fig. 37. Schematics of the different printing methods.

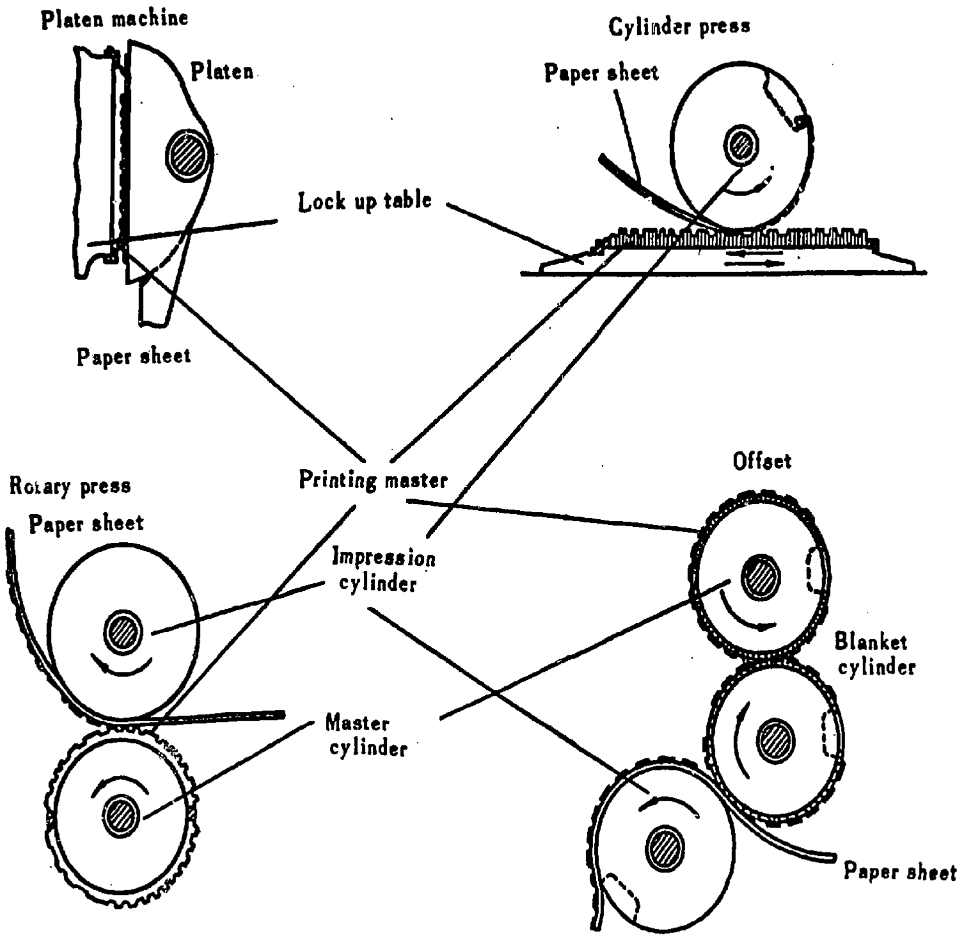


Fig. 38. Block-diagrams of the different printing machine types.

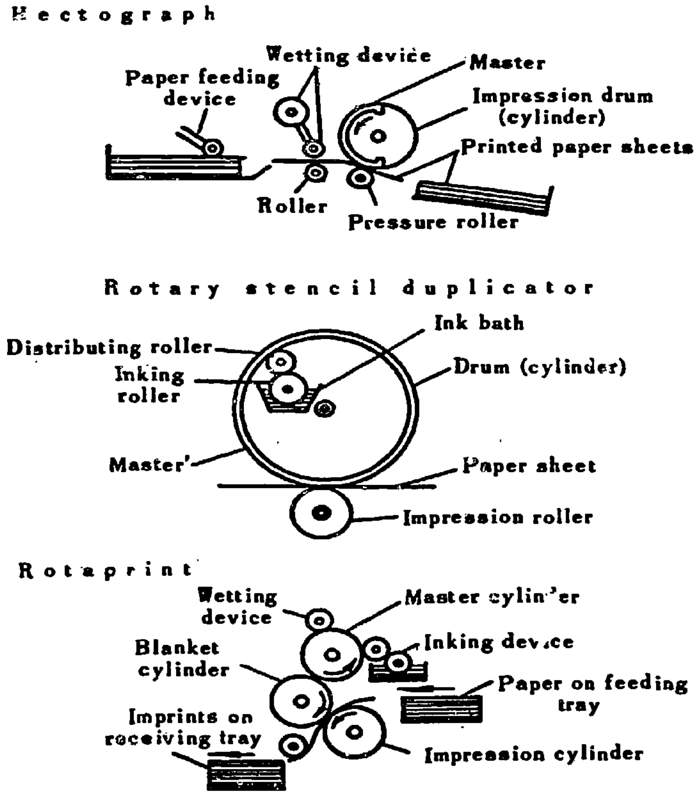


Fig. 39. Schematics of operation of the hectograph, rotary stencil duplicator and rotaprint.

Adopted from the book "Razmnozheniye tekhnicheskoy dokumentatsii" (Duplication of technical documents), by V.D. Zasov and V.N. Yurin. Moscow, "Mashinostroyeniye, 1968, p. 129, 135, 137.

Criterion for selection			Typewriter	Automatic typewriter	Hectograph (spirit duplicating)	Hectograph with changeable stamping mechanism	Rotary stencil duplicator with typescript "wax" stencil	Rotary stencil duplicator with electronically cut stencil	Rotaprint (offset duplicating)	Addressing machine
Imprint	Acceptable quality				▩	▩				
	Good quality		▩				▩			▩
	Excellent quality			▩				▩	▩	
Text	Typewritten		▩	▩		▩	▩			▩
	Any kind				▩	▨		▩	▩	
Illustrations	Line				▩		▨	▩	▩	
	Fine line (engineering drawings)							▩	▩	
	Half-tone							▩	▩	
Colour	One colour		▩	▩	▩	▩	▩	▩	▩	▩
	Multicolour	in one run			▩	▨				
		in several runs					▨	▩	▩	
Run	Up to 6 copies		▩	▩						▩
	From 6 to 12		▨	▩	▨	▨				▩
	From 12 to 100			▩	▩	▩				▩
	From 100 to 300			▨	▩	▩	▨	▨	▨	▩
	From 300 to 3000						▩	▩	▩	▩
Stationery	Writing		▨	▨	▩	▩			▨	▩
	Typographic		▩	▩			▨			▩
	Offset						▨	▨	▩	▩
	Note						▩	▩		▩

Fig. 40. Choosing means of document duplication.

Symbols: ▩ recommendable

▨ acceptable

Adopted from: Schöppenthau H. Rationelles Vervielfältigen. - "Rechentechnik", 1966, Jg. 3, No. 6, S. 27.

ALPHABETICAL INDEX

Abstract 47, 167
Abstract journal 47-49, 167
Active storage (AS) 82, 176
Addressing machines 141, 198
Afanasev, V.G. 145
Alembert, J.D. 42
Algorithm 118
Alisov, M.I. 137
All-Union Book Chamber 54
All-Union Institute of Scientific and Technical Information (VINITI) 48-50, 68
All-Union Research Institute for Technical Information, Classification and Coding (VNIIKI) 69
All-Union Scientific and Technical Information Centre (VNTI Centre) 69
Allen, G. 36, 162
Alphabetical filing of entries, rules 89-90
Alphabetical subject catalogue 51
Analytical reviews 46-47
Annotation 48
Aperture card 190
Artandi, S. 129
Article (in scientific journal) 28-30
AS (Active Storage) 83
Asheim, L. 74, 76
Asimov, I. 37
Author catalogue 52
Author systems 86-90
Author's certificate 31
Automatic Subject Citation Alert (ASCA) 61
Automatisation of IR 117
Azgaldov, E.G. 4

Barabanov, A.N. 30
Bartlett, H.A. 142
B.A.S.I.C. (Biological Abstracts Subjects In Context) 58
Batten, W. 107
Becker, J. 128
Beers, R.J. 162
Beilstein, E. 45
Bernal, J.D.12,29
Bernier, C.L. 113
Berry, M.M. 129
Besterman, T.A. 56
Bibliographical description, rules 86
Bibliographical entry 87
Bibliographie de la France 55
Bibliography 18, 52-56
Bibliography of bibliographies 56
Biographical (who's who) dictionary 46
Biological Abstracts 47, 58
Biological Abstracts Subjects In Context (B.A.S.I.C.) 58
Blondleau, R. 38
Body-punched card 121
Book 25-27
Book science 19
Borko, H. 21
Bottle, R.T. 56
Bourne, Ch.P. 21
Bradford, S.C. 35
Bradford's law of scattering 35-36
Branch-oriented information agency 69
Brandley, G.H. 179
Breyer, A. 131
Brillouin, L. 117
British National Bibliography 55
Brockhaus (German encyclopaedia) 44
Brogan, G.E. 76
Broglie, L. 12
Brookes, B.C. 179
Brown, Ch. 39
Brown, P. 29
Bryson, K.D. 154
Buck, J.T. 76
Burman, C.R. 56
Burton, R.E. 39,41,165
Bush, V. 117, 125

Callaham, L.I. 46
Callimachus 42
Card-se[illegible]ng machine 123
Carey, [illegible].P. 62
Carlson, C.F. 134
Casey, R.S. 129
Catalogue, library 50-52
Central Research Institute for Patent Information (CNIIPI) 69
Chandler, G. 62
Channels of dissemination of scientific information 63-64
Chemical Abstracts 49
Chemical Titles 58

Chernyi, A.I. 4,14,16,21
Citations
- bibliographic coupling of documents by 36-39
- indexes 58-60

Clapp, V. 76
Classification
- faceted 99-100,103-104
- hierarchical 92-93, 103
- Soviet scheme 96-97

Classified catalogue 51
Cleverdon, C.W. 81,175
CNIIPI (Central Research Institute for Patent Information) 67
Coates, F.J. 179
Codes for punched cards 120-121
Collection of papers 26
Colon Classification 99
Composing typewriters 137-138
Computer and IR 127-128
Continued publications (serials) 27-30,88
Coordinate indexing 105-107
Copying, document 130-136
Costello, J.C. 116
CRIS 126
Cumulative Book Index 54
Current awareness information service 60-61
Current Contents 61
Cybernetics 17
Cylinder presses 136,196

Daguerre, J.M. 131
Davidson, D.E. 62
Davis, W. 29
De Solla Price, D.J. 11,21,34,41,144
Dean, J. 74
Dean, H.H. 154
Dembowska, M. 76
Deposited manuscript 33
Descriptor 107
Descriptor IR language 107-108
Deutsche Bibliographie 55
Deutsche Nationalbibliographie 55
Developing countries 73-75
Dewey, M. 94
Diazo process 133
Dictionary 46
Diderot, D. 42
Diffusion-transfer-reversal process 132
Digiset 138
Dissemination of scientific information 14,63-64
Dissertation 33
Distillation process 133
Document
- definition 24
- stages of development 22-23
- types 24

Documentation 17
Dual Spectrum process 133
Duplicating, document 136-141

Edge-punched card 119
Edison, T.A. 137
Efficiency of
- IR system 80-81
- scientific information activities 149-150

Electrofax method 135
Electrography 134
Electrostatic process 134
Emin, I. 46
Encyclopaedia Britannica 44
Encyclopaedias 43-45
Engels, F. 10
Epstein, L. 59,162
Erastov, K.O. 4
Estienne, 113
Euler, L. 180
Euratom-Thesaurus 112
Express-Information 50

Faceted classification 99-100, 103-104
Fairthorne, R.A. 85, 154
Feature card 121-122
FID (International Federation for Documentation) 3, 16, 75,94
Filmorex system 124-125,191,192
Flat-bed printing 136
Flexowriter 138
FLIP 126
Flugge, W. 45
Foskett, D.J. 98,101
Fowler, F.G. 46
Fowler, H.W. 46
Fowler, M.J. 62
Fry, B.M. 56

Garfield, E. 37,60,62
Gelatin-dye-transfer process 132
Gelfand, M.A. 76
General bibliography 54-56
Giljarevskij, R.S. 14,21,76
Glidden, K. 137
Gmelin, L. 45
Goldberg, E. 125
Gorin, E.E. 134
Grammatical resources of IR languages 113-115
Gravure printing 136,195

Gray, P. 45
Gregory, W. 55
GSSSD (State Service of Standard Reference Data) 69
Gudovshchikova, I.V. 56
Guide to
- bibliographies 56
- general reference books 46

Handbook 45-46
Hayes, R.M. 129
Hectograph 138-139,197
Herschel, J. 133
Hierarchical classification 92-93, 103
Hollerith, H. 119
Holm, B.E. 114
Houghton, B. 41
Howerton, P.W. 116
Huygens, Ch. 13

ICSU (International Council of Scientific Unions) 75
IFIP (International Federation for Information Processing) 75
IFLA (International Federation of Library Associations) 75
Indexes, novel 56-60
Industrial catalogue 31
Informatics
- basic concepts 17
- definition 16
- interrelationships 17-19
Information 15
Information agency 65-70,173
Information carrier
- continuous 125-126
- discrete 119-125
- material 118-127
Information leaflet 32
Information needs
- study 146-147
- types 148-149
Information Retrieval (IR)
- basic notions 78-80
- efficiency 80-81
- essentials 77-84
- language 78-79,113-115
-technical means 117-128
Information Retrieval (IR) system
- definition 78
- efficiency 80-81
- examples 82-84
- general framework 81-82
Information theory 17
Input Converter (IC) 81,176
Input Device (ID) 82,176
International Council of Scientific Unions (ICSU) 75
International Federation for Documentation (FID) 3,16,75,94
International Federation for Information Processing (IFIP) 75
International Federation of Library Associations (IFLA) 75
International Organization for Standardization (ISO) 75
Invisible college 63
IR (Information Retrieval) 78
ISO (International Organization for Standardization) 75
Ivanov, V.V. 22

Jacquard, J.M. 119
Jerom, St. 42
Johnson, E.D. 41
Journal 28-30
Justowriter 138

Kaluzhnin, L.A. 22
Kapitsa, P.L. 13
Katz, W.A. 62
Kaye, G.W. 45
Kebler, R.E. 39,41,165
Kennedy, J.F. 146
Kent, A. 76,85,154
Kerker, A.E. 56
Keyword 105
Kimber, R.T. 177,179
King, W.V. 179
Kirk, R.E. 45
Klyachkin, Y. 126
Kolchinsky, M.L. 172
Kolmogorov, A.N. 17
Koltay, E. 55
Kosolapov, V.V. 148
Kozenko, A.V. 148
Krichevsky, G.G. 56
Kronick, D.A. 41
Kulagina, O.S. 22
KWIC indexes 57-58

Laby, T.H. 45
LaFontaine, H. 94
Lancaster, F.W. 116
Lange, N.A. 45
Larousse (French encyclopaedias) 44
Latini, B. 112
Lenin, V.T. 97
Letterpress printing 136,195
Le Versier, U. 151
Library catalogues 50-52

Library classification schemes 93
Library routines 71-73
Library science 18,65-66,70-73
Linguistics 18
Links indicators 114
Linofilm 138
Lipetz, B.-A. 174
Luhn, H.P. 58-62

McDuffie, W.C. 179
Magazine 28-30
Magnetic tape 127
Maichel, K. 46
Maizel, R.E. 55
Malclès, L.-N. 56
Manual 26
Manuscript, deposited 33
Marx, K. 10
Matching criterion 79
Mason, B. 56
Maxwell, J. 13
Mechanically sorted cards 122
Mechanization of IR 117-119
Memex 125
Mendeleyev, D.I. 151
Menou, M. 143,144
Merton, R.K. 41
Metcalfe, J. 101
Michaelis, L. 59,162
Microcard
- opaque 124
- transparent (microfiche) 124-125, 191,192
Microfiche 124-125,191,192
Microfilm 125-126
Microfilm reader 126,131
Microfilm selector 126
Microfilming 131
Mikhailov, A.I. 14,21
Mills, J. 175
MIRACODE 126
Monograph 26
Mooers, C.N. 108,152
Moore, C.K. 56
Moscow University (Lomonosov University of Moscow) 3
Muller, C.S. 133
Multivolume publication 27,88

Needham, C.D. 101
Nesmeyanov, A.N. 9
Newspapers 30
Newton, I. 13

Office reproduction techniques 136-137
Official publication 26,87
Offset lithography 140,197
Offset printing 136
Organization of scientific information activities in the USSR 67-70, 173
Otlet, P. 16,94
Output Device(OD) 82, 176

Paper (in scientific journal) 28-30
Passive Storage (PS) 82, 176
Patent literature 31
Peek-a-boo card 121-122
Periodical publication 27,88
Permuted-title (KWIC) index 57-58, 170
Perry, J.W. 129
Photography 131
Photon 138
Phototypesetting machine 138
Piróg, W. 150
Platen machine 136,196
Poisk-Ok microfilm selector 126
Polovinchik, A.N. 148
Preprint 33
Printing office operations 64-65,172
Proceedings of conferences 26,88
PS (Passive Storage) 82,176
Psychology 18
Publishing house operations 64-65, 172
Punched card 119-123,188,189,190
Punched paper tape 127,137-138
Putseyko, E.K. 135

Rabinowitch, E. 59,162
RADIR 126
Ranganathan, S.R. 71,99
Rapid Selector 125
Reference literature 43-46
Reflex processes 131-132
Regional information agency 69
Relevance 79
Remington 137
Reports, scientific and technical 32
Reprography 130
Research (special) library 70-71
Resolver (R) 82,176
Reviews 46
Roget, P.M. 113
Rohinsky, M. 55
Role indicator 114
Rotaprint 140,197
Rotary machine 135,196
Rotary stencil duplicator 139-140,196
Rules
- for alphabetical filing of entries 89

- for bibliographical description 86-90,177,178,179
RX-codes IR language 115

Samain, J. 124
Sarton, G. 45
Sayers, W.C.B. 101
Schaffert, R.M. 142
Scheele, M. 129
Schlundt, E.M. 56
Science, regularities of the development 9
Science Citation Index 58-60, 171
Science of science 19
Scientific document, definition 23-25
Scientific information 15
Scientific information activities
- basic concepts 15-17,160
- contents 150-152
- definition 15-16,143
- efficiency 149-150
- organization in the USSR 67-70, 173
- prospects 152-153
- role in science 143
- social demand for 12-13
- tasks and stages 13-14
Scientific publications
- patterns of the distribution 34-40,162,163,164
- regularities of the obsolescence 39-40,165,166
Scientific report 32
Secondary bibliography 56
Secondary documents
- definition 42
- stages in the development 42
- types 23,42-61,161
Selective dissemination of information (SDI) 61
Selenyi, P. 134
Semantic Code 115
Semiotics 18
Serial books 27,87
Serials (continued publications) 27,88
Series entry 88
Set-theoretic terminology 104-105, 185
Shamurin, E.I. 18
Sharp, J.R. 85
Shaw, R. 125
Shcherban, A.N. 148
Shepard's Citation 60
Shera, J.H. 76
Sholles, K. 137
Siegel, F. 55
Simon, K.R. 53,56
Smith, C.N. 179
Smith, R. 56
Sorting machine 122
Soviet Encyclopeadia, Great 43
Soviet scheme of classification 96-97
Special bibliography 55-56
Special library 70-71
Spencer, K.J. 56
Spirit duplicating 138-139,197
Stabilization process 131-132
Standard 30
Standard numbers for books and serials 89
State Service of Standard Reference Data (GSSSD) 69
Steiner, R.F. 162
Stencil duplicating 139,197
Stratton, G.B. 29
Subject catalogue 51
Subject systems 90-92,102-103
Superimposable card 121-122
SYNTOL 115

Taube, M. 108,114,116,129
Taublee, O.E. 154
Taylor, H. 108,122
Technical catalogue 31
Technical means of IR 117-128
Technical publications, special types 30-32
Technical report 32
Terenin, A.M. 135
Textbook 26
Thermo-Fax process 133-134,193
Thermography 133,193
Thesaurus 110-113,187
Thesaurus of Engineering and Scientific Terms 113
Thesaurus of Engineering Terms 112
Timiryazev, K.A. 44
Transactions 28,88
Translating dictionary 46
Tripathi, S.M. 179
Totok, W. 56
Tropovskij, L.N. 97
Typewriter 137
Typographic printing 136,195,196

UDC (Universal Decimal Classification) 94-98,181,182,183
Ulrich, C. 55

Underwood Typewriter Co. 137
Unesco (United Nations Educational, Scientific and Cultural Organization) 4,75
Uniterm System 108-110,186
Universal Decimal Classification (UDC) 94-98,181,182,183
Unpublished scientific documents 32-34
Users of scientific information, principal groups 144-147

Varityper 138
Vavilov, S.I. 12
Venn, J. 185
Verifax process 132
Verry, H.R. 142
Vickery, B.C. 35,85,101,107,116
Videofile System 127
VINITI (All-Union Institute of Scientific and Technical Information) 49,67,68-69
VINITI Abstract Journal 49,68-69
Virchow, R. 9
Vleduts, G.E. 176
Voigt, N. 148
VNIIKI (All-Union Research Institute for Technical Information, Classification and Coding) 69
VNTIC (All-Union Scientific and Technical Information Centre) 69

Walford, A.J. 46
Walker, J.R.A. 62
Weimann, K.-H. 56
Weitzel, R. 56
Western Reserve University Semantic Code 115
Who's who (biographical) dictionary 46
Wiener, N. 127-128
Winchell, C.M. 46
Wooster, H. 113,116
World Information System 73-75
Wyatt, H.V. 56
Wysocki, A. 154

Xerography 134-135

Yurin, V.N. 197

Zanker, V. 59,162
Zasov, V.D. 197
Zatocoding System 108
Ziman, J.M. 21
Zipf Law 35
Zlochevskij, S.G. 148