at universities in obtaining the use of large-scale research facilities, including instruments and systems, and would present opportunities for graduate students to engage in thesis research. The University Corporation has energetically moved to establish the Center, and funds have been allocated by the National Science Foundation for the initial steps, including the appointment of a director and his planning and organizing staff.

Other centers, both regional and national, devoted to major research endeavors, are under construction, such as Kitt Peak and Green Bank observatories. Still others will follow, like those designed to carry on the great research programs in oceanography recently proposed. The centers will not be hampered by barriers among sciences and departments, for their mission is to search out answers, not to maintain the straight lacing of straight-laced disciplines! They will provide freedom to communicate, and encouragement to cooperate, throughout their scientific staffs. In all of them, the common factor is dependence on instrumentation.

Instruments, the Unifying Element

Little doubt remains that in our burgeoning activities, research in instrumentology and in instrumentation for research are gaining recognition and respectability. Paradoxical as it may be, the application of research to the furtherance of research is as basic as the research itself. Without it, much research would be vastly more difficult if not impossible. To illustrate: astrophysical studies depend on expert knowledge of the latest in lens design, on image converters, and on maximization of signal-to-noise ratio in microwave receivers; study of the atomic nucleus could not proceed without the application of research to the constant improvement of high-energy accelerators and on their proper design, construction, and management.

The scientist whose research deals with instruments for research is finding his place among his fellows. Since the scientist's reward is, in large part, recognition by scientists, this is important to science, for it will encourage able students to select an interesting and satisfying career. The trend is gratifying. More persons with unusual talent will be needed to assume the planning and execution of the experimental attack as the problems and the instruments for probing them become more involved and complex.

Science faces a bright future. So does man, in his enjoyment of the fruits of science, if he can become and remain a rational being in his relations with his fellow-occupants of this planet. Theory and experiment will continue, as they have in the past, to work hand-in-hand to advance knowledge, and the greatest advances will occur where self-centered and ingrown disciplines shed their isolationism and work cooperatively in exploring dark areas of broad interest. In all such efforts instruments, the indispensable tools of science, are the unifying element; hence they must and will play a vital part.

Note

1. I am greatly indebted for guidance, in my endeavor to appraise the notables in the light of modern historical research, to Dr. Duane H. D. Roller, associate professor of the history of science at the University of Oklahoma, Norman.

Research on Handling Scientific Information

Improvements in communication and information handling contribute to scientific progress.

Helen L. Brownson

Research on new and improved methods of handling scientific information received its initial impetus from an imaginative and stimulating article by Vannevar Bush entitled "As we may think," which appeared in *The Atlantic Monthly* in July 1945. He stated the scientific information problem succinctly:

"There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial.

"The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships."

Bush predicted a change in this situation "as new and powerful instrumentalities come into use." He indicated tasks that might be performed by existing and potential mechanical aids in adding to the record of accumulated knowledge and in consulting that record. He envisaged a possible future device for individual use, a sort of mechanized private file and library, for which he coined the name "memex." Resembling a desk equipped with slanting translucent screens on which material could be projected for reading and a keyboard with selection buttons and levers, it could store on microfilm a tremendous volume of material—books, periodicals, newspapers, correspond-

The author is program director for documentation research in the Office of Science Information Service, National Science Foundation, Washington, D.C.

ence, longhand notes, and photographs. The stored material would be coded by a technique of "associative indexing," which would tie related items together; and the user would continually build the trails of association while entering new material and consulting stored material.

Fifteen years have passed since this article appeared. We still do not have a "memex" or a technique of associative indexing ready for use. But we do have a whole new field of interdisciplinary research on the handling of scientific information. The phrase "research on the handling of scientific information" is an awkward name to apply to a field, but it seems to me to be the most descriptive and least confusing name that is general enough to cover all the activities that make up this field of research. (The term *documentation research* is often used, but unfortunately *documentation* means different things to different people.) A recent Soviet paper (*1*) indicates that in the Soviet Union, also, this new field has no satisfactory name as yet: "Searches for new ways of storage, dissemination, and processing of scientific and technical information have led to the creation of a scientific discipline (still unnamed), which can be considered a branch of cybernetics."

The lack of an accepted name for a new field of research is perhaps not unusual, particularly when the field is still so young that there is as yet no clear delineation of its scope and no widespread agreement as to research objectives and priorities.

The research that is the subject of this article is concerned primarily with the handling of information expressed in language and other nonquantitative forms, such as photographs and circuit diagrams, as distinguished from quantitative or numerical data. The latter can be quite readily coded, stored, and manipulated by widely understood electronic data processing techniques. But the mechanized handling of the ideas, concepts, and techniques embodied in the texts and illustrations of scientific documents presents difficult problems that have yet to be solved.

It is my purpose in this article to give an informative account of the scope of present activities in the field and brief descriptions of current research projects. It is not feasible in the space of this one article to cover every project, but I shall describe representative projects in each research area. I shall be able to be more informative about the research projects supported by the National Science Foundation, because I am more familiar with them. The foundation, however, has for some years compiled and published a semiannual report on *Current Research and Development in Scientific Documentation*, which contains descriptions of projects and lists of relevant publications and reports. The information contained in the latest number (*2*) of this series serves as the basis of my discussion of the field as a whole.

I shall not attempt to assess the merits of any particular effort or the relative importance of the various areas of research. Any attempt to review the field critically should be based on thorough study of all available publications and reports, a much larger task than any one individual could undertake. In order to make possible such thorough study of work in the field, the National Science Foundation provided funds in 1958 for the establishment at the National Bureau of Standards of a Research Information Center and Advisory Service on Information Processing. Some of the center's operating funds have been provided also by the Council on Library Resources and the Bureau of Standards. The center is assembling all reports and publications in its field of interest and is preparing a number of background and "state-of-the-art" papers on research activities.

Before discussing the various areas of research I shall touch briefly on three other topics in order to provide a certain amount of background information: the organizations supporting and conducting research on the handling of scientific information, the development of special-purpose information-handling equipment, and the extent to which nonconventional information-handling systems are already in use.

Organizations Supporting and Conducting Research in the Field

The National Science Foundation was given a broad directive in its enabling act of 1950 "to foster the interchange of scientific information among scientists in the United States and foreign countries." From the beginning, those responsible for the foundation's scientific information programs believed that one very important means of fostering such interchange would be the encouragement and support of studies of information problems and research directed toward improved methods and techniques for handling scientific information. The National Defense Education Act of 1958 made the responsibility of the foundation in this field more specific. It directed the foundation to establish a Science Information Service to, among other things, "undertake programs to develop new or improved methods, including mechanized systems, for making scientific information available." The foundation has supported studies in the field since 1952, the first year funds for grants were available; but the program, known as the Documentation Research Program, has been expanded considerably since 1956. Insofar as possible it places emphasis on projects of a fundamental or general nature that may produce new insights, knowledge, or techniques applicable to information systems and services.

Like the foundation, the Office of Naval Research and the Air Force Office of Scientific Research support research on information handling in the interests of science generally.

Certain other federal agencies have tremendous internal information-handling problems of their own and therefore have a compelling need to seek better techniques and devices for their own use. The intelligence agencies fall in this category and are making a considerable effort to devise new techniques. In addition to the work within the intelligence agencies, the Intelligence Laboratory of the Rome Air Development Center, an activity of the Air Research and Development Command, supports under contract a good many research and development projects. The U.S. Patent Office also handles a tremendous volume of information and has an extensive research program of its own on the mechanization of patent search procedures. The National Bureau of Standards is collaborating in the research program of the Patent Office and is also conducting independent research on various aspects of mechanized information handling.

Research on mechanical translation is currently supported by the Air Research and Development Command, the National Science Foundation, the Central Intelligence Agency, the Office of Naval Research, the Office of Army Research, and the Army Signal Corps. An Interagency Committee on Mechanical Translation Research has been established by the National Science Foundation to coordinate the administration of these various research programs.

In addition to these federal agencies,

some private organizations provide funds for work in the field. The Council on Library Resources, established in September 1956 with a grant of $5 million from the Ford Foundation, has initiated and supported many studies of library problems, including a number of projects concerned with the development of new information-handling equipment and new indexing techniques. The American Society for Metals for more than three years has financed a mechanized literature-searching project. The Chemical Abstracts Service and the American Institute of Physics have active research programs on information handling and dissemination, supported in large part by the National Science Foundation.

Much of the more fundamental research in the field is being done by a dozen or so universities with the aid of grants or contracts from federal agencies, and their number is gradually but steadily increasing. Private firms and research institutes are also doing a good deal of research in the field, most of it under contract with federal agencies or the Council on Library Resources.

Organizations abroad that are actively conducting research include the following: several governmental organizations, a society of librarians and information specialists, and a private language research group in England; a university laboratory of social psychology and a center for nuclear research in France; Euratom, an international organization with headquarters in Brussels, which has an internal research program in this field and contracts for additional work with organizations in several European countries; the Patent Office of the Netherlands; a library research group in India; and a number of research institutes in the Soviet Union that are concerned broadly with the applications of electronic machines and with the field of linguistics.

Development of Special-Purpose Information-Handling Equipment

Although standard punched-card equipment and general-purpose computers are used in most mechanized information systems and in most of the research in this field involving the use of machines, a number of devices designed especially to handle linguistic and graphic information have been and are being developed. The following are examples.

Perhaps the earliest development was the Rapid Selector (*3*), conceived by Vannevar Bush before World War II. The first model was built by the Engineering Research Associates, of St. Paul, Minnesota. It scans very rapidly a large reel of microfilm containing textual material on one side of each frame and coded information on the other and selects and reproduces the frames for which the codes meet prescribed specifications. The latest version of the Rapid Selector embodies some modifications built into it in recent years at the National Bureau of Standards and is in use in the Bureau of Ships of the Department of the Navy.

Minicard equipment (*4*) has been developed for the Air Force by the Eastman Kodak Company, and three sets of the equipment are now in use either operationally or experimentally. A set of Minicard equipment includes a special camera for photographing documents to be entered into the system, a film processor, storage magazines, a Minicard duplicator, a sorter, a selector, a viewer, an enlarger, and so on. The Minicard itself is a discrete piece of microfilm measuring 16 by 32 millimeters and containing up to 12 pages of text and an area for codes.

The AVCO Corporation, with partial support from the Council on Library Resources, has developed a high-density, rapid-access information storage device to provide a means of automatically gaining access to large collections of documentary material. As with the Minicard system a microphotographic medium is used for information storage. A model based on a 100:1 reduction of page size is under construction.

Filmorex equipment, a French development, stores and searches information recorded on "*microfiches*," microfilm cards measuring 35 by 60 millimeters. Each *microfiche* holds up to two pages of text and up to 20 codes or indexing terms. The cards pass through the Filmorex selector at the rate of 700 per minute, and the textual material on the selected cards can be enlarged and reproduced automatically.

Magnacard equipment (*5*), which has been developed for the Air Force by the Magnavox Company, uses individual magnetic cards as the basic storage medium and pneumatic techniques for selective transport of the cards to achieve any desired grouping or rearrangement.

Another development, which has been supported in part by the Navy and the Air Force, is the photoscopic disk and associated equipment at the IBM Research Center. It is a large-capacity, random-access device that manipulates coded, photographically recorded information. It was designed to be particularly well suited to the handling of natural languages for such purposes as automatic translation and information storage and retrieval.

The Rabinow Engineering Company is also developing a large-capacity, rapid-access device in which information is stored as small holes etched in metal plates. It is believed to be suitable for use as an automatic dictionary in translation or retrieval systems.

The Laboratory for Electro-Modeling in Moscow has developed, for use in information-handling systems, a storage device that uses sheets of paper on which small capacitors have been preprinted (*6*). Information is stored by punches in the sheets, each punch disconnecting a capacitor. The sheets are electrically interconnected to permit very rapid electronic searching of the stored information.

Quite a number of organizations are developing character-recognition devices that will automatically "read" alphabetical and numerical characters. Reading machines that are versatile enough to recognize different styles and sizes of printed type will be needed to provide the "input" to translation and information retrieval systems.

A principal point to be made in discussing equipment is that we do not yet have procedures for handling information that enable us to use existing equipment to best advantage. When effective procedures have been developed and tested, still other special-purpose devices may be needed. But it would appear that the state of our computer technology is sufficiently advanced that any such device can be readily built if the engineers can be told in very precise terms just what it must be able to do.

Current Use of Nonconventional Information-Handling Systems

The National Science Foundation issues a series of reports on *Nonconventional Technical Information Systems in Current Use* as a service to individuals and organizations interested in scientific-information handling. The latest number in this series, and its supplement (*7*), describe in general terms 52 systems that employ either nonconventional indexing techniques or automatic equipment, or both. The reports also list available publications describing the

systems in more detail. These 52 systems are by no means all the systems of this kind in existence, but they exemplify the various types of systems now in use. Brief descriptions of a few of the larger partially mechanized systems will illustrate the success with which new techniques and machines are being applied in systems now in operation and giving service to users.

The Merck Sharp and Dohme Research Laboratories use an IBM 101 Electronic Statistical Machine and punched cards for storing and searching index entries for publications in the fields of pharmaceutical chemistry, biochemistry, microbiology, and so on. The system was initiated in 1950 and covers about 60,000 documents. A panel arrangement, on which code numbers are indicated with switches, facilitates the setting up of searches for code patterns. This panel was developed at the Merck Laboratories for use with the 101.

The Socony Mobil Oil Company and Smith, Kline and French Laboratories both have punched card systems employing also the IBM 101. Each system covers upwards of 100,000 documents. The Socony system, in operation since 1952, covers journal articles and patents on petroleum chemistry. The SKF system, in use since 1953, covers the chemical, biological, and clinical data in published papers and internal reports.

The General Electric Company in Cincinnati has combined a system of key word indexing for about 50,000 documents on flight propulsion and related subjects with a program for searching the coded information on an IBM 704 computer. A machine search can be made to produce either a list of the accession numbers of documents identified as being relevant to the search prescription or printed citations and short abstracts of the documents identified.

The Office of Basic Instrumentation, National Bureau of Standards, has designed a "peek-a-boo" system, which covers some 30,000 papers and reports on instrumentation and is probably the largest operating system of its type. It makes use of 5- by 8-inch cards, one for each indexing term in the system, and punching and viewing devices. Each card has provision for 18,000 holes, and each hole punched on a term card identifies by its location a document that has been indexed by that term. To search for documents relevant to combinations of terms, the appropriate term cards are stacked in front of a light source, and the holes in alignment that permit the light to shine through all of the cards identify the documents relevant to the selected terms. Since the capacity of a card is 18,000 holes, another set of cards must be started after 18,000 documents have been indexed.

Among the foreign systems of particular interest is that of the documentation center of the Centre National de Recherche Scientifique, in Paris. Employing Filmorex equipment, the system has been in experimental operation for four years and stores abstracts and codes for some 200,000 papers on the biological effects of chemical compounds.

Of the 52 systems described in the foundation reports, 34 deal primarily with chemical or biochemical information and 35 are operating within industrial organizations. Whether these proportions would hold true for all existing nonconventional systems is not known, but at least they suggest that the incentive for devising information-handling systems and for attempting to use machines and new techniques for this purpose is greater in industry than elsewhere. Similarly, it might be inferred either that chemical information is more readily coded for mechanized handling than other information or that chemists and chemical literature specialists are to some extent more interested than specialists in other disciplines in trying out new information techniques. Perhaps both inferences are correct.

Even the larger systems that have been described are small in comparison with the task of placing indexes to the content of all scientific literature under some sort of mechanized control. It is not at all clear that procedures that work satisfactorily for limited collections of information in relatively homogeneous fields could be adapted to handle very large volumes of material cutting across a number of scientific disciplines.

A recent report of the Senate Committee on Government Operations (*8*) contained a brief description in general terms of a very large punched card system for intelligence information—the "Intellofax" system of the Central Intelligence Agency. The system was instituted in 1947 and has undergone some modifications since then. It now comprises over 40 million punched cards; those prepared in recent years bear both coded information and microfilms of documents mounted in apertures in the cards. Quite understandably, the agency has not made available detailed information about the system. For intelligence purposes, however, it may not be necessary to index or classify the subject matter of scientific and technical documents to anything like the the depth that would be required for a system designed to make literature searches for scientists.

One last observation about the existing partially mechanized systems is that most of them appear to be based on some form of "coordinate" indexing, which makes use of individual indexing terms or short compound terms that can be combined in a variety of ways in the process of searching. Conventional indexing or subject-heading systems include phrases as index entries whenever several terms are required to describe a concept. The indexing terms of a coordinate system may be called key words, descriptors, Uniterms, or simply terms. By now, a good deal of experience has been gained with systems of this type, some of which incorporate techniques for handling various types of relationships among terms. The National Science Foundation, therefore, has contracted with Documentation, Incorporated, of Washington, D.C., for a state-of-the-art study of all aspects of coordinate indexing, including operating experience, theoretical work, and experimentation. The report of this study should be available early in 1961.

Current Research Activities

In spite of rapid and continuing progress in computer technology and programming techniques, continuing development of special-purpose devices for the handling of linguistic and graphic data, and accumulating experience with mechanized techniques in searching limited collections of information, the essential problem of applying machines to the handling of scientific information on a large scale has yet to be solved. This unsolved problem has to do with means of analyzing the subject content, meaning, and relevance of documents for mechanized handling. Research directed toward this end is making progress but is still in its infancy.

Only a few short years ago, most people in the field were thinking in terms of attempting to mechanize only the more routine or clerical aspects of information handling, with reliance on human judgment for the analysis and indexing or classification of the subject content of documents. All of the present operating systems that employ automatic searching techniques are of this type. Much of the current research work

is directed toward the development of improved indexing or classification systems to be used by human analysts in combination with automatic techniques for coding, storing, and searching the index data and printing out the results of a search. Continued work in this area is important because improved systems of this type will undoubtedly be useful for many purposes.

Within the past five years, however, more and more persons have become interested in exploring how far it is possible to go in devising ways of using machines to process and analyze automatically the actual language of documents for a variety of purposes—translation, abstracting, indexing, selective dissemination, and storage and retrieval. Some of the current research efforts on the processing of information for dissemination and retrieval and all of the mechanical translation efforts are in this area.

The following descriptions of research activities are grouped under several headings: background studies of the communication practices and information needs of scientists; research on partially mechanized information retrieval systems, with mechanization applied to the more routine processes subsequent to document analysis; research on the automatic analysis of texts; evaluation and comparison of information-handling systems and procedures; and studies in other areas.

Communication Practices and Information Needs of Scientists

Although most of the research described in this article is directed toward the development of faster and more reliable techniques for handling scientific information, it goes without saying that the information needs of the scientific community should determine the character of new information services and techniques. Although there is a good deal of intuitive and subjective knowledge about the various ways in which scientists communicate the results of their own work to others and learn about the work of others, there is very little precise, objective knowledge of the inadequacies in the present flow of scientific information and of the cost to scientists and to society of those inadequacies. A deeper understanding of the role and the mechanics of communication within the sciences and of the information problems and needs of scientists would be of assistance to scientific societies and all other organizations involved in planning and maintaining publication programs and scientific and technical information services. Such understanding is also needed for the design of information-handling systems and procedures that will provide the sort of help scientists can use to best advantage.

In an excellent discussion of the problem of "technical communication in psychology" (*9*), the Board of Scientific Affairs of the American Psychological Association presented the thinking of a group of psychologists on the need for work in this area of communication practices and information needs:

"BSA considers the problem of efficient and effective communication of scientific information to be perhaps the most critical problem faced by scientific psychology today. . . . The techniques of scientific communication now employed by psychologists have evolved from past experiences, conditions, and needs, without much self-conscious analysis or deliberate planning.

"The interdependence of the functions of different media of communication suggests that psychologists should self-consciously design a communication system in which the inherent capabilities and limitations of different media are matched with the specific needs or functions to be served. Only by the analysis of such a total system will it be possible to specify the required media, their principal functions, and the criteria in terms of which they should be evaluated.

". . . traditional roles of our communication media need to be reexamined in the light of present and anticipated information loads and needs. Further, these needs should be those of the user of the information rather than those of the originator of the information."

During the past 10 to 15 years, some 30 or more studies and surveys have been made of the use by scientists of scientific publications and reference tools and of their views on information problems and services. Most of these studies have been made by means of questionnaires or interviews, in which scientists were asked to recall or to estimate their use of publications and services. A few investigators have used other techniques for gathering data, and some have attempted to delve deeper into the communication problem and to identify actual situations where research suffered because certain existing knowledge was not available, and thus to throw some light on specific failures of our present communication system. By and large, however, the studies thus far have been concerned with what scientists do in the present situation, which is admittedly far from perfect. They should be regarded as merely a first step toward better understanding of the information problem, to be followed by experimentation with courses of action likely to lead to improvements in the situation.

Studies in this area prior to the summer of 1959 were analyzed and compared in a *Review of Studies in the Flow of Information Among Scientists,* prepared for the National Science Foundation by the Bureau of Applied Social Research, Columbia University, and issued in January 1960. This review contains a discussion of completed research and a synthesis of the findings of the various studies wherever the data are at all comparable. The many obstacles to comparison included the diversity of populations studied, differences in the units of observation or recording, differences in classification of communication media and channels, and paucity of analyses in depth and of interpretations of the collected data. Some of the studies were designed for the purpose of guiding the activities of a single establishment and were useful for that purpose but added little to the knowledge needed for decisions on a more general level. Techniques of data-gathering have included the analysis of library withdrawal records (with or without special questionnaires attached); study of records of inquiries at information centers; use of questionnaires, interviews, and diaries; and observations during specified time intervals.

The choice of technique, however, is believed to be secondary in importance to careful delineation of the area of inquiry and of the units of observation. The review states the belief that the time has come to devise new research strategies and suggests some neglected approaches that seem both feasible and promising, such as studies focusing attention on the functions of the scientific communication system and the ways in which each is being met and studies designed to determine how individual units of information percolate through the scientific community.

Among the studies in progress at present, two are being conducted in in-

dividual industrial laboratories. A group at the Bell Telephone Laboratories is analyzing the responses to questionnaires sent to their 4700 technical and administrative personnel to obtain data on the use of the library and of information; a paper on the principal results and implications of the study will be published. The Technical Information Division of the Esso Research and Engineering Company is analyzing some 2000 requests for technical information, covering both current reference use of materials and retrospective searches of published literature and internal reports. It is hoped that the study will help to determine the types of indexes most useful for answering the questions put to the Division.

The Institute for Advancement of Medical Communication is conducting two very similar studies of the fate of papers given orally at scientific meetings. The purpose is to discover what proportion of them are eventually published, the factors determining whether or not a given paper is published, and the time lag between oral presentation and publication.

The American Institute of Physics is experimenting with various forms of printed indexes to see how they are received by physicists and how useful they prove to be in practice, and is studying several aspects of current publication practices in physics.

The Operations Research Group at the Case Institute of Technology is obtaining data on the use of published scientific literature by chemists and physicists. A technique has been devised whereby the participating scientists, by using a random alarm device, make notes at random intervals on what they are reading, if they are in fact reading when the alarm sounds. This study is a continuation of an earlier operations research study of the scientific activity of chemists (*10*), in which both observations by others and self-observations away from the place of work were used. The data collected on a large representative sample of chemists indicated that the average chemist working in industry seems to spend more time in scientific communication (reading, writing, listening, and talking about scientific matters) than in all the rest of his activities concerned directly with science. Of a total of 32.4 hours per week (both during and after working hours) devoted to scientific activities, 16.5 hours were spent in scientific communication, as compared with 10.4 hours in working with equipment, 3 hours in data treatment, and 2.5 hours in thinking and planning. Comparable figures are not available for university chemists, because the "after hours" portion of the study covered only industrial chemists. Although the 25,000 observations made during the first study resulted in a clear picture of the prominent role of communication in the activities of chemists, there were relatively few (some 900) observations of reading acts, not enough from which to draw conclusions. Consequently the present phase of the study is aimed at gathering more data on the use of publications.

Another study of the use of publications is being made by the University of Chicago Library to determine whether the present or potential value of publications can be established on the basis of frequency of use, and whether criteria that will predict their future use or obsolescence can be discovered. Browsing patterns, as well as circulation use, for both journals and monographs are being examined in several subject areas. The prediction of obsolescence functions will be tested against expert judgments, and the data collected at the University of Chicago Library will be compared with check samples taken at other research libraries.

The question of measurement of the value of recorded scientific information is to be explored in a new study just getting under way at the Case Institute of Technology. It will encompass a survey of the various measures of the value of information and of scientific productivity that have been used, a controlled evaluation to determine their usefulness and limitations, and an exploration of the feasibility of developing more reliable measures of value. The value or significance, and also the rate of obsolescence, of scientific papers and of the information contained in them are factors which must be taken into account, if possible, in designing improved means of scientific communication and mechanized searching systems.

It will be apparent from this brief summary that the current studies of communication practices and information needs of scientists are concerned with pieces of the problem. Plans have been discussed for a thorough "systems analysis" of communication within certain scientific disciplines, but no such study has yet materialized. Perhaps the next step in this area of research will be one or more broad studies covering all aspects of communication within a discipline with the goal of working toward the design of an improved total communication system for that discipline.

Partially Mechanized Information Retrieval Systems

Some of the current research projects on information retrieval systems are concerned with the development of improved procedures for the analysis, indexing, or classification of the subject content of documents for use by human analysts and with the development of mechanized procedures for the subsequent, more routine, steps, such as coding the index entries for machine storage, matching search prescriptions against the stored data, and printing the results of a search.

The Chemical Abstracts Service has a broad research program directed toward the mechanization of information services in chemistry (*11*). Using existing equipment and techniques, they are experimenting with ways of coding data on the structure and properties of chemical compounds for mechanized storage and retrieval and for automatic preparation of data compilations. An important part of the broad program consists of research on the semantic content of chemical literature in order to discover whether basic patterns of the terms and phrases used can be identified and to determine how best to handle, in a mechanized system, the concepts expressed in language. The research tasks include the preparation of concept dictionaries and their experimental use in the processing of abstracts and mechanized searches.

The Cambridge Language Research Unit, Cambridge, England, is attempting to devise a mechanizable information retrieval system based on a thesaurus-like lattice structure of terms taken from the documents indexed. The system is being applied experimentally to the Unit's collection of reprints. Since the number of terms tends to increase rapidly as documents are added to the collection, attention now is being devoted to research on methods of combining terms on the basis of their tendency to occur together in documents. The present work is largely theoretical, being concerned with the development of a general method of grouping, the careful choice of defining criteria for

groups, and the formulation of mechanizable procedures for finding the groups in a given set of terms.

The Ramo-Wooldridge Division of Thompson Ramo Wooldridge Inc. is investigating the utility of assigning weighting factors to index tags, indicating in an intuitive way the degree of relevance of a tag to the document being indexed (*12*). Through the use of such "probabilistic indexing," it may be possible to improve retrieval efficiency and effectiveness by listing documents identified in a search in order of degree of relevance to the question. The degree of relevance is determined by a computational procedure in which the weighting factors assigned both to index tags and to components of a search question are used.

Soviet efforts in this area appear to be concentrated primarily in the field of chemistry (*13*). The goal is to develop, for use with machines, a "machine language" for the fundamental forms of chemical information. The Electro-Modeling Laboratory in Moscow has developed a system of unambiguous linear notations for structural formulas and their components suitable for mechanized handling, and has also developed symbols for chemical reactions which can be combined with the linear notations. This special chemical information language is being extended so that it can be used to express information on the mechanism of chemical reactions, the technical conditions of their realization, and the properties and behavior of physicochemical systems so that eventually the language can handle all knowledge in the fields of chemical science and engineering. It is hoped that in time chemists will be used only to select information from the literature for incorporation in the mechanized information system. Means are being devised for automatically converting structural formulas, various types of names of compounds, and other information into the notations and codes of the machine language. Eventually, with automatic reading devices, it is expected that the coding of structural formulas can be done entirely without human intervention.

The two remaining programs described in this section are placed last because, although their more immediate results will be applicable to partially mechanized systems in which the analysis is performed by human beings, some aspects of the work are relevant to mechanization of analysis and thus relate to the following section as well.

The United States Patent Office has a broad program of research directed toward the design, construction, and experimental use of mechanized search systems. Techniques are already in use for mechanized searching of the steroid patents, and similar work is well under way in the additional fields of the polymer art, chemical processes, organic phosphates, and electrical arts. The several concurrent projects include studies of ways of characterizing and coding the essential elements and relationships that make up the patents in the fields mentioned; studies of various search techniques, with experimentation primarily in the field of chemistry; and studies of techniques for file preparation and organization. The specific projects within the Patent Office program that are directed toward the ultimate goal of automatic analysis of complete texts for mechanized searching are concerned with the linguistic interrelationships that occur in patents and related technical literature; the creation of a regularized unambiguous language, termed "ruly English"; and methods for converting natural language to a standardized form.

A research group at the Itek Corporation is endeavoring to systematize the operations of information searching systems (*14*). The ultimate goal is the mechanization of such operations wherever feasible, but some of the systematized procedures being deveolped may first be applied by human analysts. The operations under study include the selection of significant information items in a scientific text as index data, their conversion from natural language into a normalized language, the interpretation of search questions and their conversion into the same normalized language, and the planning and programming of searches. A method of representing natural-language expressions in normalized form is being developed. The "normal" grammar, which is based on concepts derived from logic, is used as a schema in the semantic organization of linguistic data, in the normalized representation of complex topics, and in the selection of index data. The project also includes the development and testing of tools, such as a dictionary of terms and operational rules governing the conversion of natural-language expressions into their normalized equivalents, and a thesaurus-type device that records and displays the semantic relationships among words and expressions.

Automatic Analysis of Texts

Current reseach on the automatic analysis of complete texts of scientific documents includes experimentation with indexing and abstracting techniques based on word-frequency statistics and tests of procedures for direct searching of entire documents for specified words and phrases (*15*). It also includes exploration of the feasibility of more elaborate techniques, beginning with automatic syntactic analysis of the sentences of the text.

A research group at the International Business Machines Corporation is experimenting with the use of statistical techniques to prepare abstracts and indexes automatically (*16*). A machine program is used to scan the entire text of a document and to identify significant words and their positional relationship to each other within sentences. From these data, a significance value is assigned to each sentence, and those sentences with the highest values are printed out as an "auto-abstract." For automatic indexing and encoding purposes, similar statistical procedures are used to establish lists and patterns of high-frequency words, and of pairs of words occurring together within sentences. The words may then be looked up in a special "thesaurus" that is stored in the machine and assigned code numbers designating the appropriate notional families. Variations on these techniques are being tried experimentally; these include an automatic system for the selective dissemination of information (*17*), based on comparison of the patterns of key words resulting from analysis of documents with the words characterizing the interests of scientists participating in the experimental program.

The Ramo-Wooldridge Division of Thompson Ramo Wooldridge Inc. is experimenting with automatic text searching with the ultimate goal of fully automatic indexing and word correlation. A recent set of experiments measured the effectiveness with which information responsive to specific questions can be recognized through machine search of the full text of documents. The experiments have been carried out within the framework of a "model" information system, consisting of a small collection of physics articles, a list of *ad hoc* questions, and a matrix of numbers representing the estimated degree of relevance of each article to each question, based on careful application of expert

human judgment. The questions were transformed into instructions to search for specified words and phrases within the texts. The texts of the model library, recorded digitally on magnetic tape, were then searched by a general-purpose computer. The responding documents were compared with those determined to be responsive through examination by subject specialists of the entire collection of articles. They were compared also with those identified by searching a conventional subject-heading index of the model library. The investigator has reported that, on the whole, retrieval effectiveness was rather poor, yet machine search of the texts produced significantly better results than human searching of the subject-heading index. Such exploration of the effectiveness of text-searching techniques is considered to be a prerequisite to the design of effective automatic indexing procedures, in which abbreviated representations of documents must be prepared by machine. The work is being extended to include a study of the feasibility of preparing abstracts automatically.

The Institute for Cooperative Research at the University of Pennsylvania is studying possibilities for mechanizing indexing and is attempting to devise sampling techniques to investigate the effectiveness of machine instructions for assigning indexing terms to documents.

A group of linguists, logicians, mathematicians, philosophers, and electronic engineers, working within the Department of Linguistics of the University of Pennsylvania, has for several years been engaged in an intensive program of research on the mechanization of syntactic analysis of English sentences (*18*). In their view, the automatic processing of the texts of documents for abstracting, indexing, and retrieval should begin with automatic phrase-structure analysis, or parsing, of the sentences of the text. A computer program for this type of analysis is being prepared. The second major step in the processing of documents, on which this group is working, will be a computer program for "transformational analysis"—converting sentences into simpler, more uniform "kernel" constructions (for example, passive constructions will be converted into simple, declarative statements). The planned third step will consist of procedures for the identification of significant words or "kernels" for storage in a retrieval system or for use in abstracts and indexes.

An investigator at Massachusetts Institute of Technology is studying the possibilities of using natural language for storage and retrieval in a mechanized literature-searching system. The work is based on the belief that it should be possible, by an appropriate research effort, to deduce the rules of natural language in explicit form so that machines can be explicitly instructed on the use of the languages. A step-by-step approach is used, each step bringing the artificial language, or "dialect," closer to English and incorporating within the language as many features of English as are thoroughly understood at the time.

The work on automatic syntactic analysis, or parsing of English sentences, that is being done in connection with information retrieval systems is closely related to much of the current research in the field of mechanical translation. Eleven research groups are active in this field in the United States, three in Great Britain, and at least six or seven in the Soviet Union; work is also under way, or is at least beginning, in Italy, France, Sweden, East Germany, Czechoslovakia, Yugoslavia, Japan, and Communist China. As one would expect, the groups in the United States and Britain are developing syntactic analysis techniques for foreign languages—primarily Russian, German, and French, with work just beginning on Chinese. The results of some of this work are beginning to be applied experimentally to English as well; for example, the "predictive analysis" technique, developed at the National Bureau of Standards and adopted also by the group at Harvard, is being tried on English at Harvard. Research in mechanical translation also includes studies of the semantic organization of languages. This work also, although oriented primarily toward translation procedures, may prove to be relevant to other information-handling procedures as well. In fact, some of the research groups and supporting agencies are interested in mechanical translation primarily as one aspect of the larger field of automatic processing of natural language for a variety of purposes. It is interesting to note that this view has also been expressed by a Soviet researcher (*19*).

"Today machine translation is regarded only as the first stage toward solving a more general and more important problem: by most fully using electronic machines as auxiliary tools of human thinking, to make the machine capable of performing the widest possible operations with texts written in different languages, to enable it not only to translate but also to edit, make abstracts, furnish bibliographical and other references, etc. All these operations boil down to extracting from the text required information and to recording that information in some other form."

Evaluation and Comparison of Information-Handling Systems

We are just at the beginning of serious efforts to evaluate and compare under controlled conditions the effectiveness and efficiency of information-handling systems and procedures. The Ramo-Wooldridge experiments with automatic text-searching procedures, described in the preceding section, included a carefully controlled comparison of their results with results obtained by human searching of a conventional subject-heading index. Several other current projects are concerned with evaluation and comparison of procedures and systems.

A project at the Cranfield Aeronautical College of England, administered by the Association of Special Libraries and Information Bureaux, involves tests of the comparative efficiency of four indexing and classification systems, which have all been applied under controlled conditions to a sample of 18,000 documents in the field of aeronautics (*20*). The systems are the Universal Decimal Classification, widely used in European libraries and publications; a "faceted" classification (*21*) devised for this project by members of the Classification Research Group of England; a conventional subject-heading system; and a system of coordinate indexing. The test searches are now in progress, and some preliminary results may be available by the end of 1960, and final results by the end of 1961. It is hoped that the results of this project will provide convincing evidence of the relative efficiencies of the systems for various purposes and will add to our knowledge of techniques for evaluating systems. It may also provide information that will be useful in the design of indexing and classification systems.

Herner and Company has developed an experimental classification and coding system for the field of atomic energy, suitable for use in either manual or mechanized information retrieval systems. The next phase of this project is

a comparison of the system's retrieval effectiveness with that of a subject-heading system and a coordinate indexing system now in use in the libraries of the Atomic Energy Commission.

At Western Reserve University, the procedures developed by the Center for Documentation and Communication Research for abstracting, encoding and searching technical information (*22*) are to be evaluated in a large-scale test program that got under way at the beginning of 1960. The procedures include the preparation by trained human beings of "telegraphic style" abstracts, consisting of terms and symbols for relationships and functions; the encoding of the resulting abstracts; and searching of the coded information by machine. The encoding procedure and a machine dictionary provide for control of synonyms and certain hierarchical relationships. These procedures are being applied to about 30,000 scientific and technical papers and reports per year in an experimental mechanized searching service for metallurgists. An *ad hoc* committee of metallurgists and information specialists has been named by the National Academy of Sciences–National Research Council to advise on means of evaluating these procedures, either in absolute terms or in comparison with other procedures. Its members can testify to the fact that such evaluation is a very complex undertaking for which there are as yet no tried and true techniques.

Studies in Other Areas

A number of other studies currently in progress do not fall within the research areas discussed thus far. The Electrada Corporation, for example, has undertaken a mathematical analysis of the structure of systems for information storage and retrieval with a view to developing methods for an economic analysis of system design. The corporation has also recently begun a program of exploratory research into file organizations, suitable for extremely large files, with a self-organizing capability and with a number of levels of increasing size and decreasing accessibility.

The University of Pennsylvania Institute for Cooperative Research is also studying file organization and the use of computers to organize files.

The National Bureau of Standards is engaged in mathematical research related to the efficient formulation of search questions for retrieval systems and to means of enabling a mechanized retrieval system to revise its own category structure. The bureau is also experimenting with a machine model of certain verbal selection and recall operations. The model consists of a small vocabulary of terms, stored records of interrelationships between the terms, and routines for various operations, such as "define" and "extend." The aim of the experiments is to determine ways in which machines may be used to speed the selection of stored knowledge and to reformulate search questions in accordance with results obtained.

With respect to the classification of knowledge, the Library Research Circle at Delhi University (*23*) and the Classification Research Group (*24*) which has been meeting in London since 1952 are continuing to study and develop techniques of facet analysis and faceted classification (*21*). In addition to this work, specifically on classification techniques, the various groups working on ways of handling relationships and grouping terms in retrieval systems are also clearly concerned with principles and techniques of classification.

In addition to the work related directly to the improved handling of scientific information, a great deal of research in other fields may produce results relevant to the design of information-handling procedures and systems—fields such as linguistics, logic, mathematics, psychology, neurophysiology, artificial intelligence, communications engineering, and computer programming techniques. It is relevant to note that some universities have established centers or programs in which the knowledge and talents of many disciplines and special fields are enlisted in research on communication problems of all kinds, including those touched on in this article. In discussing (*25*) the establishment of one such center, the Communication Sciences Center of the Massachusetts Institute of Technology, J. B. Wiesner has pointed out that "the fundamental difference between the communication sciences and many other sciences is that, though people in this field may be forced to study physical properties of systems, it is the organizational or structural properties which are of primary concern. Here there are new concepts to understand, and new mathematical tools are required and being created."

Conclusion

By describing current work in the relatively new field of research on the handling of scientific information, I have tried to indicate the present scope and objectives of that work. I hope that this descriptive account also conveys the impression that a great deal more research must be done before we shall know how best to use machines to aid in the handling of information. The possible consequences of such research are of the greatest importance for science. At the very least the research should result in increased understanding of the complex processes of communication among scientists and in improvements in the means for accomplishing such communication and for consulting the record of accumulated scientific knowledge. It may also, however, lead to completely new ways, only dimly foreseen at this time, of using machines to supplement human intelligence in information and communication processes (*26*).

References and Notes

1. G. E. Vleduts, V. V. Nalimov, N. I. Styazhkin, *Soviet Phys.: Uspekhi* (a translation of *Uspekhi Fiz. Nauk*) **2**, No. 5, 637 (1959).
2. *Current Research and Development in Scientific Documentation, Rept. No. 6* (National Science Foundation, Washington, D.C., 1960).
3. R. R. Shaw, *AIBS Bull.* **4**, No. 3, 20 (1954).
4. J. W. Kuipers, A. W. Tyler, W. L. Myers, *Am. Document.* **8**, 246 (1957).
5. A. M. Nelson, "The Research and Development of the Magnacard System," Rept. of Magnavox Co. Research Laboratories (1958) [*WADC (Wright Air Develop. Center) Tech. Rept. No. 58/421; ASTIA Document No. AD 211694*], distributed by Office of Technical Services, U.S. Dept. of Commerce.
6. W. H. Ware, Ed., "Soviet computer technology–1959" (account of a trip in 1959 of U.S. computer experts to the Soviet Union), *IRE Trans. on Electronic Computers* **EC-9**, No. 1, 99 (1960) [also in *Communs. ACM* **3**, No. 3 (1960)]; L. I. Gutenmakher, "Basic directions and trends in the development of digital computer systems in the problem of designing machines and instruments," in *Sovremennyye Napravleniya v Oblasti Konstruirovaniya Tekhnologicheskogo Oborudovaniya* (Contemporary trends in designing technological equipment) (Moscow, 1957) (JPRS translation No. 2276, available from Office of Technical Services, U.S. Dept. of Commerce, OTS 60-11, 319).
7. *Nonconventional Technical Information Systems in Current Use, Rept. No. 2* (National Science Foundation, Washington, D.C., Sept. 1959); suppl. (Mar. 1960).
8. *Documentation, Indexing, and Retrieval of Scientific Information* (Senate Committee on Government Operations, Washington, D.C., 1960).
9. *Am. Psychologist* **14**, 267 (1959).
10. "An Operations Research Study of the Scientific Activity of Chemists," Rept. of the Operations Research Group, Case Institute of Technology, Cleveland, Ohio (1958), submitted to the National Science Foundation.
11. G. M. Dyson, *Chem. Eng. News* **37**, 128 (1959); *ibid.* **38**, 70 (1960).
12. M. E. Maron, *J. Assoc. Computing Machinery* **7**, 216 (1960).
13. L. I. Gutenmakher and G. E. Vleduts, "The

prospects for the utilization of informational-logical machines in chemistry (USSR)," in *Problemy Vysshego Khimicheskogo i Tekhnologicheskogo Obrazovaniya* (Problems of Higher Chemical and Technological Education) (State Publishing House, Moscow, 1959) (paper presented at the 8th Mendeleyev Congress on General and Applied Chemistry; JPRS translation No. R-331-D, 19 Feb. 1960, available from Office of Technical Services, U.S. Dept. of Commerce); A. M. Zuckermann (Tsukerman) and A. P. Terentiev (Terent'yev), "Chemical nomenclature translation," paper presented at the International Conference for Standards on a Common Language for Machine Searching and Translation, Cleveland, Ohio, 1959 (Interscience, New York, in press).
14. T. M. Williams, "From text to topics in mechanized searching systems," "Proc. National Symposium on Machine Translation, Univ. of California, Los Angeles, Feb. 1960" (in press).
15. D. R. Swanson, *Science* **132**, 1099 (1960).
16. H. P. Luhn, *IBM J. Research Develop.* **2**, 159 (1958); T. R. Savage, "The Preparation of Auto-Abstracts on the IBM 704 Data Processing System," *International Business Machines Corp. Rept.* (1958).
17. H. P. Luhn, "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment," *International Business Machines Corp. Rept.* (1959).
18. Z. S. Harris, "Linguistic transformations for information retrieval," *Proc. Intern. Conf. Sci. Inform.* (National Academy of Sciences–National Research Council, Washington, D.C., 1959), pp. 937–950; A. K. Joshi, "Computation of syntactic structure," in "Proc. Intern. Conf. for Standards on a Common Language for Machine Searching and Translation, Cleveland, 1959" (Interscience, New York, in press); L. Gleitman, "The isolation of elements for grammatical analysis," *ibid.*
19. I. A. Melchuk, *Computers and Automation* **8**, 23 (1959).
20. C. Cleverdon, *Aslib Cranfield Research Project.* The report is available from College of Aeronautics, Cranfield, England (September 1960).
21. In a faceted classification (as distinguished from a hierarchical classification) terms are grouped into categories. B. C. Vickery, in a book entitled *Classification and Indexing in Science* [(Academic Press, New York; Butterworths, London; 1959), p. 12], offers the following explanation: "Instead of trying to construct from the 'original universe' one vast tree of knowledge, facet analysis first groups the terms into categories—kind, state, property, reaction, operation, device and so on—and then arranges the terms within each category into the form of a classificatory map." He points out, for example, that within the field of chemistry, "alcohol" is a *kind* of chemical substance, "liquid" is a *state* of that substance, "volatility" is a *property* of it, "combustion" is a *reaction*, "analysis" is an *operation* performed by man, and "burette" is a *device* for carrying out an operation.
22. J. W. Perry and A. Kent, *Tools for Machine Literature Searching* (Interscience, New York, 1958).
23. S. R. Raganathan, *Depth Classification* (Indian Library Association, 1953); *Ann. Library Sci.* **5**, 33 (1958).
24. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval," *Library Assoc. Record* **57**, 262 (1955).
25. J. B. Wiesner, *IBM J. Research and Develop.* **2**, 269 (1958).
26. The Russian transliteration system of the U.S. Board of Geographic Names is used in the Soviet references.

Book Reviews

Arms and Insecurity. A mathematical study of the causes and origins of war. Lewis F. Richardson. Nicholas Rashevsky and Ernesto Trucco, Eds. Boxwood Press, Pittsburgh, Pa.; Quadrangle Books, Chicago, Ill., 1960. xxv + 307 pp. $10.

Statistics of Deadly Quarrels. Lewis F. Richardson. Quincy Wright and C. C. Lienau, Eds. Boxwood Press, Pittsburgh, Pa.; Quadrangle Books, Chicago, Ill., 1960. xlvi + 373 pp. $12.50.

The English physicist and mathematician Lewis F. Richardson (1881–1953) was engaged for many years on the monumental task of constructing a purely objective, mathematical and physical theory of peace and war. These volumes, previously available only on microfilm, are now submitted to the public substantially as Richardson left them. During the author's lifetime only a small sample of his work in this field was published, but this sample included a notable essay entitled "Generalized foreign politics" (1939). *Arms and Insecurity* presents in full a model of armament races and wars outlined in this essay; *Statistics of Deadly Quarrels* takes up the problem of war from a different point of view, also adumbrated in short, early publications.

Arms and Insecurity traces the phenomenon of war to the breakdown of international equilibrium, as evidenced by runaway armament races. Richardson works here with an equilibrium model based upon the idea that changes in one power's military spending over a certain period of time are directly proportional to a potential enemy's actual military expenditures during the same period. The system of interacting powers is at equilibrium when the rate of change in spending is zero for each.

The model specifies the conditions under which such a state of equilibrium can be reached. In fact, the functional relationship between the *change* in *A*'s expenditures over time and *B*'s *actual* expenditures is not such that the former must be positive when the latter is; in other words, it is not the case that *A* is compelled to *increase* his spending as long as *B* spends anything, and vice versa. Rather, the function determining a power's spending differentials contains parameters that can bring the latter down to zero, even when the potential enemy is arming. For example, spending becomes onerous when it rises above a certain level. Having reached that stage, *A* will tend to slow down its armament efforts, and this will have a dampening effect upon *B*'s arming. Richardson postulates, moreover, that the relationship between potential enemies need not be one of pure hostility. While arming against each other, two powers or coalitions can also engage in cooperative activities, such as trade. This will counteract the hostile impulses ("grievances") that induce the powers to arm against each other in the first place. The question in each situation is whether the stimulus for increased spending, that is, spending by the other side, will be sufficiently counterbalanced, by the onerousness of spending and by the prevailing degree of cooperativeness, to result in the stabilization of military expenditures. The volume of trade, serving as a measure of cooperativeness, is crucial in this respect. If it is too small in relation to military spending, the latter will grow to infinity. Such an irreversible trend means war; stabilized military spending means peace.

Richardson's model is deterministic. When the crucial quantities, military spending and trade in particular, show