The Coming of Age of Information Technology

By Mortimer Taube, Ph.D.

Chairman, Board of Directors, and Chief Scientific Advisor
Documentation Incorporated
Bethesda, Maryland

ABSTRACT

A discussion of the distinction between a scientific discipline and a professional activity and of the reasons why "information technology" is a professional and not a scientific activity. The application of computers to the improvement of information handling is discussed as a professional advance similar to advances in medicine made possible by new instruments. A major new instrument, like the computer in information technology or the X-ray machine in medicine, may sometimes drastically change the practice of a profession. Some of the implications of the computer for librarianship are discussed in these terms.

I CAN perhaps best introduce the subject of information technology to medical librarians by comparing it to the practice of medicine. Medicine is not a science. It is an integration of a number of sciences for a practical purpose; that is to say, it is a learned profession. A learned profession is a form of activity which involves internship, clerkship, or some form of controlled experience based on knowledge obtained from one or more scientific disciplines. Similarly, information technology is a professional and not a scientific activity. It will become a learned profession when it, too, involves controlled experience based on knowledge derived from one or more scientific disciplines.

Medicine as a profession is concerned with the diagnosis, treatment, and cure of disease. Information technology, as a profession, is concerned with the design, installation, and operation of information systems.

The parallelism is also evident with respect to a number of conditions which militate against advances in either medicine or information technology. Just as medicine has its lunatic fringe of faith healers, quacks, and purveyors of flashing lights for the cure of cancer, so information technology has its fringe of charlatans who promise that miraculous improvements will issue from little black boxes or that soon there will be discovered a mechanical substitute for human intelligence.

Secondly, medicine, especially in the person of the general practitioner, has an overly cautious resistance to change. New therapies and new drugs

sometimes have to fight their way against the ingrained habits of old practitioners. The same resistance to change and a defense of the status quo occurs in the field of information technology. It has been said with reference to medicine that if God wanted people not to have tonsils, they would be born without tonsils, and in the field of information technology it sometimes seems that librarians believe that 3 × 5 cards were handed down to the library profession along with the Decalogue from Mount Sinai.

There is a third difficulty which plagues information technology, which is much more subtle and more difficult to handle, because unlike the defense of the status quo and resistance to change, this third difficulty comes wrapped in the mantle of progress. I refer to the concept of "basic research." Every government agency in Washington pays lip service to basic research; and sometimes this attitude is an indirect defense of the status quo. In fact, it is just the kind of attitude which wins Brownie points for those who have it, while making sure they won't have to change what they are doing. In other words, if your research is concerned with something which will come to fruition twenty years from now, if at all, then the pressure to improve what you are going to do tomorrow is removed. One excellent example of this attitude can be found in the basic research program in information retrieval carried out by our Patent Office. For ten years the Patent Office has avoided the necessity of ameliorating the present situation and seems willing to face a breakdown of the patent system while a basic research program is carried out to find machines which will read patents and determine what is patentable.

Medicine today is mature enough so that we have very little pretension to basic research in medicine. It is generally recognized that such research, if it occurs at all, will occur in the scientific disciplines upon which medicine is based, and that medicine itself will be concerned with the application, testing, and proving of advances in the preclinical and clinical sciences.

One of the most unfortunate developments in the short life of information technology was the location within the National Science Foundation of the office concerned with developing information technology on a federal and national level. The National Science Foundation supposedly supports basic research as contrasted with the mission-oriented research of the other government agencies. Hence it seemed reasonable for the Office of Scientific Information within the National Science Foundation also to support basic research as contrasted with professional solutions of real problems. Unfortunately, information technology is not a science, and basic research in information technology is a scientific aberration. Nor is it reasonable to suppose that the mission of the Office of Scientific Research is to foster basic research in those sciences which underlie information

technology. There are such sciences, but they are within the competence and responsibility of other divisions of the Foundation, e.g., information theory, chemistry, mathematics, physics, engineering, etc. It is true that the Office of Scientific Information has found its own "basic" disciplines, namely, lexicography, semantics, and linguistics.

The first of these, lexicography, Webster defines as "the art, process or occupation of making a lexicon or a dictionary." Those of you who have seen and enjoyed the recent reviews of the Webster-Merriam Third Edition will certainly agree that lexicography is anything but a science.

Semantics is defined by Webster as that portion of philology which is concerned with the science of meanings, but here the word "science" is a euphemism. If there is any subject of interest to man which least resembles a science, it is the concern with what words do or should mean. As a matter of fact, linguistics made an attempt to become scientific during the nineteenth century by asserting that it was not concerned with meanings, but only with physical sounds and expressions. In any case, linguistics is a humanistic, historical discipline and not a natural science. There is scarcely a single attested case, after ten years of linguistic research fostered by the National Science Foundation and similar organizations, of an application of such research to the design, installation, or operation of an information system. Yet if the "basic research" reported in such compilations as *Current Research and Development in Scientific Documentation* is examined, there will be found no reports of mathematical advances or developments in logic, physics, or chemistry; and it may fairly be concluded that basic research in information technology is again a euphemism describing a great number of activities whose relevance to the solution of practical problems is either specious or confused.

If we turn now from a consideration of these three conditions which stand in the way of the coming of age of information technology, we still must make one more comparison with the practice of medicine before concerning ourselves with information technology per se. If someone were to stand before you to defend the Harvard system of medicine, as opposed to another speaker who defended the Princeton system of medicine, you would dismiss them both as charlatans; yet you are accustomed in meetings of this type to hear people expound and defend the Western Reserve system, or the DOC INC system, or the ASTIA system, or even the MEDLARS system. The point I would make is that there are diverse applications of information technology to resolve different problems and to serve different purposes just as there are different treatments in medicine for different diseases, different age groups, and different environments. In the field of medicine we would give a doctor short shrift who defended his practice in a particular circumstance by stating, "This might not be the best treatment, but the patient didn't die."

But within the field of information technology there is a vast amount of literature which begins with some such words as, "The system I am about to describe may not be the best system, but it works for us." The whole point of the coming of age of information technology is that there is now available, for those who care to employ it, professional skill to institute the best system for any given purpose. The only qualification I would accept to this statement is that as the state of the art advances, applications to particular problems will change and improve just as medical treatment changes and improves, but we expect from a competent physician the marshaling of the cumulated resources of his profession to treat any disease, and we should expect from the information technologist the marshaling of the cumulated technological know-how for the solution of problems in information.

So far I have emphasized the parallelism of information technology and medicine, but most of you here are concerned not with medicine but with medical librarianship, and you may wonder what is the relation of librarianship to information technology. Librarianship is a particular application of information technology to collections of books and periodicals. It has certain similarities and it differs from other applications—the organization and searching of patents, the organization of materials for publication in indexes and abstract journals, the organization and control of inventories, the classification of cargo and the loading and unloading of ships, etc.

The relation of librarianship to these other activities and its union with them in the profession of information technology is not generally recognized because librarianship developed as a practical art long before anyone was concerned with its theoretical or professional basis. The same, by the way, can be said of medicine, which emerged from barbering long before it was a profession based upon established scientific disciplines.

One of the principal factors contributing to the advance of medicine is the development of new instruments, ranging from the clinical thermometer and the stethoscope to electroencephalograms and radioisotopes. The same situation prevails in information technology. Verner Clapp and others have pointed out that it was the invention of stereotype printing which led to the first experiments with centrally produced card catalogs and the hope for eventual centralized cataloging and distribution of cards. The movable type used by the Library of Congress to print its catalog cards is a technological development which permits the production of cards and book catalogs from a single text composition.

Within information technology, the instrument which seems to offer the most promise for the future is the computer or the data processing machine which has an internally stored program. Before a new instrument can become effective within a technology, those who use it must be willing

to reconsider their current practices in light of new practices made possible by the new instruments. The new instrument will usually fail if it is forced to operate in an environment created by old technologies. I have remarked in other places that the first automobile was a smelly, noise-making toy until such time as there developed an environment of roads, gas stations, and spare part depots, in which the automobile could function efficiently.

By the same token, it is pointless to set a computer in the middle of a traditional library and expect miracles; but it is reasonable to inquire how the existing library technology and apparatus of a library can be modified in order to take advantage of the potentialities of a computer for improving library service.

In the remaining portion of this paper, there will be presented a number of generalized descriptions of the manner in which a computer functions in information technology, with illustrations drawn from the Scientific and Technical Information Facility which Documentation Incorporated operates for NASA, and the MEDLARS Project, about which you will hear in some detail tomorrow.

Computers were originally developed for mathematical operations. When they are required to perform a great number of repetitive operations, and this number may run into the millions, on a small amount of data, i.e., a set of mathematical equations, the operation of the computer attains maximum efficiency. On the other hand, when a computer performs a relatively simple operation on a great mass of data fed into it, the efficiency of the computer approaches a minimum. Consider the simple problem of looking up an author entry in a catalog. There is no existing or planned computer that can beat the facility with which a librarian can go to the card catalog or telephone book and find a single proper name, or, for that matter, a single subject entry. On the other hand, faced with the task of looking up a thousand names or subjects in a catalog, a computer is usually more efficient than a manual operation. This conclusion illustrates the principle of "batching," or cumulating requests. If a computer is to be used for information retrieval or literature search, in an information center, such searches will be efficient only if they can be batched.

There should be no mystery for librarians in the process of batching; for librarians practice it now in catalog maintenance, although they do not do so in reference work. When a cataloger at the Library of Congress finishes cataloging a single item, he does not immediately rush down to the main catalog and file in the cards. Such a procedure would be most inefficient. Rather, the cataloger or the catalog department accumulates enough cards to make interfiling a batch process.

One of the ways in which the computer may change library services is by providing reference service to customers in batches, the size of the batch

to be determined by the traffic in the library and by the amount of delay the individual customer will stand before going elsewhere. Nevertheless, in spite of the relative efficiency of batch look-up, I think it can be said that if the sole role of a computer in the library is to look up and retrieve information, then it will be some time before the use of computers in libraries will be justified in terms of cost and efficiency.

There is another obvious role of computers in information centers, namely, as printers. The tabulators and high-speed printers which are parts of computer systems can print out hundreds of lines per minute of references, bibliographies, abstract journals, etc. On the other hand, it is expensive to prepare and store material in a computer for such print-outs and as a pure printing device the computer does not match in either cost or quality of print-out a number of other devices which are readily available. These two conclusions, namely, the inefficiency of a computer treated solely as an information storage and retrieval device and the inefficiency of a computer treated solely as a printer, are implicitly, if not explicitly, recognized by most professionals in the information technology field.

It was natural, therefore, that the next step in computer utilization should be an attempt to secure efficiency by building a computer system which would function both as a storage and retrieval apparatus and as a printer. This type of combination is the essence of the MEDLARS Project and is also the fundamental concept behind the mechanization of the operations carried out in the Scientific and Technical Information Facility. In both organizations a regular abstracting or indexing publication is looked upon as a by-product of a computer store of information which is used for special bibliographical compilation and reference search. The MEDLARS Project is still in the planning stage, but the combination of computer look-up and abstract journal print-out has already been accomplished by the Scientific and Technical Information Facility.

One other contribution the computer has made, is already making in our operation, and will certainly make in the MEDLARS operation, is that it permits cooperation in two different directions. We not only put out abstract bulletins based on our own work—publish them from the computer—but we enter into our computer and cumulate in it indexes prepared by other organizations, so that the computer has increased the compatibility of indexes in different abstract publications.

This is compatibility *in*. So far as compatibility and cooperation *out* are concerned, in addition to distributing an abstract bulletin to those people who have no computer, we also distribute to a number of organizations the actual magnetic tapes from our computers, with a program, so that they can be used in computers by organizations large enough to have them.

Admittedly, this is not an easy problem. There are many difficulties in distributing such material, but I am sure that both concepts of compati-

bility are essential to the eventual development of the MEDLARS Project and, therefore, both are relevant to the medical field.

At this point, I am omitting a great many details concerning the actual preparation and integration of the bulletin and machine search system, but there are certain important points which should be made. By careful attention to the quality of paper, type of ribbons, and cleanliness of type faces, we are able to get a very satisfactory text from the computer. That seems to be a detail, but I think it is most important.

We have introduced one important innovation. By a programmed overstrike, that is, by getting the computer to "stutter," we are overstriking on the index heading and, therefore, are able to achieve a boldface for such headings on the computer. The total index page is composed inside the computer, including running heads, pagination, and column indication. In the publication each item is restricted to an average of five indexing terms. In the computer store, each item is indexed to an average of twenty-five terms.

In this task of devising a computer store of information which is at the same time the raw material for an *Index Medicus* or *Star*, one of the major intellectual problems is devising an appropriate vocabulary or system of headings for both machine search and bulletin, either *Star* or *Index Medicus*.

In devising such vocabularies or guides, one thing becomes clear—that the great amount of lexicographical work now going on in the development of thesauri, about which many of you have heard, is almost totally irrelevant to the operating problems of devising the required vocabularies for machine search and bulletin publication. Within the machine, the Scientific and Technical Information Facility does use, and MEDLARS will use, a coordinate indexing system consisting of a vocabulary of terms or descriptors which will be manipulated in the search process. It is not necessary to store within a machine the heading "Magnetic Amplifiers" nor the heading "Brain Neoplasms." In one case, the machine can be asked to compare items stored under "Magnetism" and "Amplifiers" in order to discover all items on magnetic amplifiers, and in the other case it can be asked to search under "Brain" and "Neoplasm" in order to find all items on brain neoplasms. The only time it becomes necessary to store combined terms or phrases in a machine is when experience discloses that the combination of individual terms results in ambiguity or too much noise. This is the standard problem of "blind Venetians" and "venetian blinds," or "lead coating for brass pipes" as opposed to "brass coating for lead pipes." As contrasted with the terms used in a machine store of information, the headings used in a bulletin will be determined in large measure by the volume of citations under each heading. To avoid the cumulation of

hundreds of items under a heading like "Amplifiers" or "Neoplasm," *Star* would use "Magnetic Amplifiers" and the *Index Medicus* would use "Brain Neoplasms." In brief, combinations of terms are used as headings in the machine only to eliminate potential noise. Combinations of terms as headings are used in the bulletin, as in a standard library catalog, to reduce the volume of entries. This means that the headings in *Star* or the *Index Medicus* will resemble to a marked degree the subject headings used in a standard library catalog, but the intellectual determination of the proper headings is identical, whether the headings are to be used in a machine store or published bibliographical service. The major difference will be that the machine preparation of such publications will facilitate the *systematic* use of subdivision in order to break up a mass of entries, the subdivisions to be applied according to a rule rather than perfunctorily or intuitively.

I should like to close by referring again to the potential role of a new instrument in revitalizing a profession. The great flowering of American librarianship occurred around the turn of the century when the giants of our profession—Hanson, Cutter, Dewey, Billings, and others—developed the rules for standard and book catalogs. Subsequently, cataloging in libraries became for the most part a technical effort designed to carry out the cataloging of individual books or periodical articles according to established rules of dictionary, subject, and class catalogs. The advent of the computer means that the total cataloging enterprise needs rethinking. There is an opportunity now to concern ourselves not with applying set rules to the cataloging of a single item, but with evolving new rules for the organization of total collections and new bibliographical services. If librarians take advantage of this great opportunity, theirs can emerge as one of the major intellectual professions of our time.