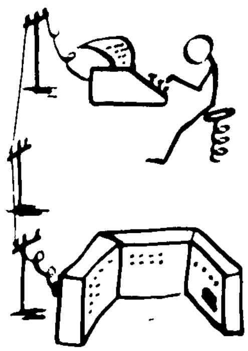
AFOSR-2711

ZTB-142



WANTED: A REACTIVE TYPEWRITER

Calvin N. Mooers

OCTOBER 1962

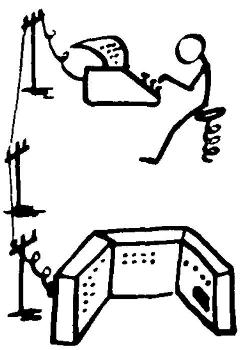
FINAL REPORT, CONTRACT 49(638)-376

SPONSORED BY

DIRECTORATE OF INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
WASHINGTON 25, D. C.

ZATOR COMPANY
140½ MOUNT AUBURN STREET, CAMBRIDGE 38, MASS.

AFOSR-2711



ZTB-142

WANTED: A REACTIVE TYPEWRITER

Calvin N. Mooers

OCTOBER 1962

FINAL REPORT, CONTRACT 49(638)-376

SPONSORED BY

DIRECTORATE OF INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
WASHINGTON 25, D.C.

ZATOR COMPANY
140½ MOUNT AUBURN STREET, CAMBRIDGE 38, MASS.

WANTED: A REACTIVE TYPEWRITER

Calvin N. Mooers

ABSTRACT

Some critical remarks are directed at many current and fashionable topics in information retrieval. The information crisis is attributed to the challenge of new machines, not to an unexpected surge in literature. The concentration on retrieval selector machines is discounted as misplaced emphasis. Programs in which information retrieval is an activity for independent effort at each documentation center are described as dated and leading to unnecessary duplication of costly effort. Thesaurus building is analyzed as a neo-uniterm aberration undertaken in the mistaken belief that thesaurus word lists can discover meanings in texts. Compound descriptors and their algebra are mentioned. The very important future use of remote computers by means of typewriters in the library with wire connection to the computer is described. Such wire-connected typewriters are called "reactive typewriters." The impediments to getting reactive typewriters are enumerated. Appropriate wire-connecting typewriters having both upper and lower case characters are only available at exorbitant cost at present, i.e., about $10,000 for each typewriter terminal. Arguments are given to show that the cost should be in the order of $500 to $1000 for the small library users. Omissions of necessary hardware in contemporary computing machines are listed. Programming languages for the reactive typewriter are discussed, and the features of the TRAC language are mentioned. Some of the broader capabilities and consequences of the reactive typewriter are sketched. A plea is made for cooperation by potential users to insist that the manufacturers make available reactive typewriter equipment at reasonable prices.

ACKNOWLEDGMENTS:

Specific acknowledgment is made to the research support from the Air Force Office of Scientific Research, U. S. Air Force, and to Dr. Harold Wooster of the Division of Information Sciences, which has permitted the freedom of speculation that has led to the numerous results mentioned in this paper. Contributory research support has also come from the National Institute of Health under grant RG-7589, from contracts with the Rome Air Development Center on specific tasks, and from the Gmelin Institute, Frankfurt, Germany.

Thanks for specifically stimulating encounters are due to Dr. Louis Schreiber, Brandeis University; Dr. Ralph Shaw, Rutgers University; Dr. Eric Pietsch, Gmelin Institute; Mr. Ed Fredkin who has some of his own ideas about reactive typewriters; Mr. Stuart Fergusson who has unearthed teletypewriter facts; to the people at ASTIA for having a very large problem and sticking with it; and to many others in many places.

Acknowledgement is made to Mrs. Rowena Swanson for editorial suggestions and changes.

TABLE OF CONTENTS

WANTED: A REACTIVE TYPEWRITER

Calvin N. Mooers

THE LIBRARIAN ASKS A QUESTION

One of the interesting kinds of encounter that I have these days is when I meet a librarian -- a real librarian -- and he asks me the question, "What should we be doing at our library on information retrieval?"

That begins the discussion. Before long his next question comes. It often comes in approximately the following form. He'll say, "I've been doing some reading in your field. Of course, you understand, I'm not an expert. However, I have the strange feeling as I read that either I don't understand, or they don't. Could it be that I am missing something?"

At this point, I reassure him that he isn't missing something, and that his assessment is essentially correct. Our field is pretty strange. It has been for a long time. Unfortunately, we seem to have gotten so used to the state of affairs, that it is hard to look candidly at our current problem to see just where we are, and where we should be going from here. We need some current assessments of our field. We need to look at the present and future states of our equipment and methods. This is what I shall try to provide in this paper. At the same time, I may be able to answer some of the questions of my librarian friends.

THE INFORMATION CRISIS

The first item to consider is the "information crisis" and the consequent intense current interest that centers about methods of machine documentation.

We are told by eminent authorities that there is such a crisis, and we must do something about it right away. It is explained that the reason for this crisis is that the increasing number of scientists and engineers are writing more and more technical reports, and that the amount of technical literature is growing exponentially.

We can't disagree with these superficial facts. However, the number of scientists, engineers, and technical reports has been increasing during at least the past 150 years. Consequently, if this now constitutes a crisis, there has been a crisis for a long time. Those who read the library science literature know that we also thought there was an information crisis soon after 1900 -- at the time the Universal Decimal Classification was developed. As for "exponential growth", this is merely an expected mode of behavior. If the literature growth rate were to drop substantially, we should have real cause for alarm! Therefore, I am not greatly impressed with these stated reasons for our present great concern for machine documentation.

Neither am I impressed with arguments based upon a presumed universal demand for stored information on the part of all our engineers and scientists. My experience in working with many retrieval systems and their users has shown that it is not hard at all to get a retrieval system to work successfully. However, one of the really fatal things that can happen to many retrieval systems is to have the system finally put into successful operation. Then the plans and discussions on system design come to an end. Now the users in lab and office are faced with the task of reading the many pertinent documents that are so easily available. Planning and discussion are stimulating; reading is drudgery. Interest drops, and the system dies a slow death. I have seen a number of retrieval systems terminated this way. However, not all users are on this negative side of "Mooers' Law." (Ref. 1) There are actually quite a large number of people who do read, and who insist upon reading the literature. This is the group I am concerned with.

My reason for believing that there does exist a very real and valid reason for the present interest in machine documentation and information retrieval is the advent, since about 1945, of some extraordinary new machine capabilities for solving the old and serious library problem. Some of us who have been in the field since the pioneering stages of electronic computers knew that a new class of machine robots was due. Given this new challenging capability, we asked ourselves in what way could these machines be properly applied to library problems. What machines should be built for these problems? Even more fundamental, what rules or methodology might we propose for the solution of the problems? Were the problems properly stated? Should new machines be built to implement old systems which had grown up within the tight constraints of the catalog card, the linear shelf arrangement, and the decimal classification? Would the forthcoming machines, which were being built for numeric and arithmetic computation, be suitable for, or capable of, non-numeric problems of sorting and selecting upon ideas -- the kind of capability that is necessary for information retrieval?

Therefore, in my estimation, the most justifiable basis for our current interest in machine documentation stems directly from the challenge of emerging new capabilities. We now have a new class of gadgets to try in the solution of the serious problem of scientific and technical information handling.

If we accept this explanation for our interest in machine documentation, then to use the term "information crisis" is to use a slogan manufactured after the fact. It is not accurate; perhaps worse, it misleads us in many ways from knowing where we really are, and where we should be going. If we realize that our interest began with the challenge provided by emerging new electronics, we can then ask questions about whether the proper machines have been designed and built in the intervening seventeen years. We can also ask whether our current plans, for using the machines we have now, make sense.

I contend that such an examination shows considerable deficiencies. I must agree with the librarian. The present documentation and machines picture is confusing.

SYMPTOMS OF THE MALADY: THE PROFUSION OF RETRIEVAL SELECTOR MACHINES

One only needs a sense of perspective to see some symptoms of the malady. Two instances can be mentioned.

Both Eastman Kodak and International Business Machines have built giant retrieval machines, the "Minicard" and the "Walnut." They have attracted much attention. However, not even the giant government information center, the Armed Services Technical Information Agency (ASTIA), can afford these machines. Moreover, the machines are capable of only a part of the overall job of servicing the customers, e.g., they cannot keep track of document inventories, check on "need to know," or do many other essential tasks. If ASTIA cannot find a place for the machines, how much more likely is it that a college or university library, a municipal library, or a special library in a medium-size or small company could afford such machines?

Selector machines are being built "in profusion." Yet, talk with librarians indicates no primary worry about retrieval. Retrieval gets done, even if the customer has to do it himself -- as he usually does by using the card catalog and cited references from his reading. The painful problems for the librarian are those of internal library operations. One particular problem is that of card catalog maintenance -- card typing, filing, and even replacement on deterioration. Other problems include book ordering, charge outs, and many other necessary "housekeeping" jobs. Selector machines don't touch these problems. Yet is is in these areas that the librarians are looking for and need help.

DUPLICATION OF EFFORT IN INDEXING

In many special libraries there is a wide-spread interest in selector devices, usually for use with an independently generated retrieval system. Each unit thinks in terms of doing its own indexing of large amounts of material, card punching, mechanization and searching. Does this make sense?

It doesn't to me. It did make sense ten years ago, when the potential for operating a high-performance information retrieval system was much less than now. At that time, an organization had to go it alone. This is no longer necessary or advisable. Duplication of brain power in reading documents, in indexing, and in typing at each location is a sheer waste.

A large body of "standard" information could be indexed (up to a certain point of specialization) as a common project, and the results made available in machine readable form to all cooperators or subscribers. ASTIA had begun efforts in this direction for government contractor reports. "Standard scientific journals" constitute another body of information. Bibliographic information for all the papers and articles in such journals could be available in machine readable form from appropriate centers. Elimination of only the typing load would be a great gain. (More about this problem below.)

WORDS VS. IDEAS IN INDEXING

Indexing of the content of a document -- particularly of a scientific document -- pertains to the ideas the document embodies. Accordingly, the indexing terms are useful in proportion to their ability to reflect the ideas. Again on the current scene we have confusion. We are told that, with use of a "thesaurus," what are called "uniterms" suddenly become "descriptors." We are told that "uniterms" have always been "descriptors." Already these precise terms have lost much of their meaning. "Thesaurus" refers to a subject-classified compendium of words in the manner of Roget's Thesaurus. "Uniterms" are words extracted directly from the title or text of a document in the methodology popularized by Taube. (Ref. 2) "Descriptors" are idea units with arbitrary labels (usually the labels are mnemonic words or phrases) from a controlled list of such idea units, in the methodology I introduced. (Ref. 3, 4)

I am thoroughly familiar with both of these methods. In particular, I made use of the "extracted word method" as the demonstration indexing method when I was widely showing the Zatocoding System in 1948. I used this indexing method because it was sufficiently simple-minded so that a businessman, who was seeing for the first time a sorted card system with selection by correlation of index terms would not become confused by too many new techniques at once. In each case, I also demonstrated the "idea unit method." (The two methods received their present names somewhat later.) I introduced the term "descriptor" for the idea unit method in 1950, and Taube began calling the other "Uniterms" in 1951. These two methods are substantially different in concept and in practice.

Confusion of the two seems to arise because words of the text of a document convey ideas or meaning, and both uniterms and descriptors are themselves words, or words and phrases. In poetry, belles-lettres, history, literature, law, and other humanistic studies in general, the exact words used in the text may be important. In science, an idea can be (and usually is) expressed in text through widely different sets of words and symbols. In mathematics, the same message can be written in many different mathematical languages (all English) -- to say nothing of dialects within the languages. I have recently come to believe that there are apparently significant qualitative differences between humanistic texts and scientific texts, and the approach of people who work with them. These differences are reflected in quite different indexing preferences.

Uniterms seem to appeal to those in literature and allied fields, where the exact words of the text are often appropriate for indexing. It is my conjecture that persons in these fields do not appreciate the situation in science. In science, the exact words of the text can be both unimportant and very misleading out of context, and are not a direct guide for indexing. In science, the ideas embodied by the words must be extracted and indexed. We cannot safely use the words by themselves. The descriptor methodology was intended for the generation of the idea units discovered in the text.

If scientific material is to be indexed to its idea content, there cannot be a mere juggling of words extracted from the text for an index useful in retrieval. To do an acceptable job, the indexer must know what the words mean, what ideas they contain, and then he must index to these ideas. It must be emphasized that in the scientific area, an idea is not defined or marked out by a list of "related" words, or by lists of words generic to, or specific to, a given word. Calling the set of such lists a "thesaurus" does not accomplish the definition of the ideas. Yet in the current misplaced enthusiasm over the "thesaurus method," this is the foundation of sand upon which the resulting index stands.

An accurate characterization of the current "thesaurus method," such as that shown by the Chemical Engineering Thesaurus, (Ref. 5) is that this method is a "uniterm system with complete cross referencing to all other listed uniterms." Most specifically, this is not a descriptor method. Unfortunately, the thesaurus method still retains some of the most serious intellectual shortcomings of the original uniterm method. Since this "thesaurus method" has attracted much uncritical attention, and is thus an aberration of some consequence, let me state the matter in another way: no amount of listing of words and related words, and related-related words, and generic words, in a thesaurus is capable of transforming words simply extracted from the text (uniterms) into descriptors. Any assertions to the contrary indicate a lack of understanding of these methods of indexing.

The descriptor method has not been static since its introduction in 1947. In a revision that I did of the entire set of 7000 ASTIA descriptors (from the ASTIA Thesaurus of Descriptors, May 1960) (Ref. 6), some new descriptor techniques were developed and displayed. Some of the highlights are: elimination of repetition, redundancy, and fuzzy statements. I have reduced the earlier version to a concise set of 1600 "atomic" descriptors. (It should be noted that the 1960 descriptor set was published in a half-completed form, for administrative reasons.) The new descriptor set covers all of ASTIA interests except chemistry, for which another technique is appropriate as an intermediate to the descriptor formation. Since the atomic descriptors tend to be rather broad, controlled specificity is attained by allowing the ad hoc formation of compound descriptors. These are new descriptors created by joining two atomic descriptors in a binomial compound. The binomial may be commutative or noncommutative. The new descriptors are systematic because they are formed from the atomic descriptors as a basis. Documents are delineated in the indexing process by sets of both atomic and compound descriptors. Document selection is prescribed by requests to sets of compound and atomic descriptors. Usually, selection can be to any descriptor which is a part of a compound. However, in some cases this is not desired, when it is desired to prevent certain of the atomic descriptors from operating in selection. The indexer makes this decision. He is able to "suppress" either one or both of the atomic descriptors in any compound. By means of these few new postulates for an "algebra" of descriptors, an enormous new range of selective capability becomes available. A report on the 1600 descriptor set, with scope notes, and on the method of compounding descriptors, is in preparation. (Ref. 7) (The interest and help of ASTIA personnel is here noted with thanks.)

MEASURING RETRIEVAL EFFICIENCY

Quite another symptom of the malady in retrieval work is the absence of a measure of retrieval efficiency. The purpose of a retrieval system is to furnish appropriate answers to questions. An appropriate answer is usually in the form of a list of documents, or the documents themselves. But, "How well does a given system perform? What is its retrieval efficiency?" Tests and measurements have been performed. Some can be excluded because they were performed on trivially small systems (e.g., one hundred documents) or because the main purpose of the tests was to prove some other point. The rest of the measurements -- some on large collections of documents -- must be dismissed as questionable or misleading in measuring efficiency because of substantial inadequacies in experimental design. Several current projects may correct this situation. But for the moment, no quantitative reliable measures exist.

MACHINES BUILT BY COMPUTER ENGINEERS -- PROBLEMS WITH TEXT

Let us now look at the related field of computing machines and their auxiliary equipment. We who deal with bibliography and information systems must use the electronic robots furnished by the computing machinery industry. Therefore, from our viewpoint, what is the state of affairs with respect to the products and services available?

My answer, in a few words, is: primitive, out of perspective, and lacking in essential components." I will now substantiate this.

Bibliography and documentation activities deal with vast amounts of text. The text is to be read by someone. Therefore, readability of the type face (technically, legibility) at all stages is important -- not only for the text of a report, but for listings of titles of reports, for abstracts, and for indexes.

Perhaps engineers don't read. Perhaps they don't know what makes for good readability of type -- a topic that has received a good deal of study. At a national meeting of a computing machinery society, the proposal has been made in full seriousness to a plenary audience that all scientific publications be issued in a standard text composed solely of block, sans-serif, upper-case characters. It was stated that people could read such text just as well or better than that in present text type fonts (erroneous), and that such a uniform font would simplify the construction of machines for character recognition and text reading!

Type readability studies disclose that serifs and the lower case of the Roman alphabet are essential for good human readability. Yet, with but one exception, available computer print-out devices, and teletype machines are all restricted to the upper case. The one exception, a modification of the IBM 1403 printer, was forced on a reluctant company by insistent customer demand. This is currently the only high-speed computer output printer that can produce characters in both the upper and the lower case as well as numerals. Other printers are too primitive to be considered seriously for incorporation into a library system.

MEGABITS OR NOTHING

Let us now consider some of the computer matters which are out of perspective. In the computing world, a recent accomplishment is the capability of "data transmission." Instead of physical transmission of sets of reels of tape from San Francisco to New York by jet plane, the data can be transmitted by wire at very high rates of speed. Transmission rates range up to the order of megabits per minute (with suitable lines) over distances of several thousand miles and with extreme accuracy. "Core storage to core storage" high-rate transmission between the electronic memories of two computers has been accomplished. Such accomplishments are fine from the "gee whiz" standpoint, and large-volume numeric data workers should get good use from such facilities. However, this doesn't solve library problems.

While it is now possible to get such high-speed long-haul transmission (and it costs accordingly), it is not possible to get suitable low-speed and short-haul transmission of a kind really useful in library service planning. There is great need for an inexpensive keyboard and printing unit (really an electric typewriter), which can be inexpensively connected via a telephone wire to a remote electronic computer. Properly installed, one or several typewriters in a library could provide all of the advantages of having an in-house computer, without the disadvantages (such as the cost of the machine while it is not being used).

It seems a modest request to ask for the construction of such typewriters and local data links to the nearest computer. But such facilities do not exist! An unbelievable host of frustrations and unrealistic costs are conjured up at the thought. Furthermore, the current computers would be thrown into tizzies with the problems of serving an array of such typewriters. Through lack of perspective, or lack of perception of a real need, the computing machinery industry is currently unable to meet demands of this kind at prices realistic to library budgets. In data transmission, attention has been restricted almost solely to high-speed, high-quantity, high-accuracy, and high-priced numeric communications.

WANTED: A REACTIVE TYPEWRITER

Let the term "reactive typewriter" refer to the special computer-connected typewriter. The term derives from the capability of the typewriter to react automatically by typing back modified versions of the basic input information. Let me describe a typical facility use. If the problem of the day at a library is typing new catalog cards, each document may be represented by five to ten different versions or arrangements of the basic information. Via the reactive typewriter, the remote computer would be loaded with procedural descriptions of the catalog card formats and a description of the permutations on the data desired to be typed out. Next the basic required master information for each document (i.e., titles, authors, etc.) would be loaded, and then the "go" signal would be pressed. The remote computer would then take over, and direct typing of the desired sets of cards.

Consider the matter of accuracy in transmission. In numeric computations, e.g., in accounting, an error in a single "bit" will cause an immense amount of nuisance. Therefore, extremely low error rates are mandatory for such service and costs concomitantly mount. In library work, extremely low error rates are usually not necessary. An error rate of one character per thousand would be tolerable for a very cheap data link (this corresponds to about one error per double-spaced typed page). A transmission error rate that was much better than one character per hundred thousand would be considerably better than the error rate found in most of the text supplied to the channel. At this point, cheapness of channel and equipment is preferable to a lower error rate. (Compare "typo" errors in newspaper text.) With text, the user can be his own error detector and corrector. Thus cheap equipment, and cheap transmission channels, can be used with the reactive typewriter.

The reactive typewriter could be a very versatile tool. In another use of it, the remote computer could be given the procedural description and the page format for composing a book list. Book titles could be listed in any format, corrections being made during input or subsequently. The reactive typewriter will then give a clean copy of the book list in the desired format. The reactive typewriter could also act as a selector machine or could be used in ordering and billing. It could draft copies of a technical report, or insert editorial changes and then type out a clean copy of the final form. It could merge book lists. It could make selective book lists or bibliographies, arranged by author, data, subject matter, etc.

I say "could be done" by a reactive typewriter. In fact, the lack of suitable equipment at an economical price, and the lack of certain computer features, prohibit present realization of reactive typewriters for library use. Certain omissions and economic hurdles determined thus far are discussed below.

COSTS THAT ARE IMPOSSIBLE

The first and major omission is the lack of a suitable economically priced typewriter unit. The first and absolute demand for any serious bibliographic work is that the typewriter be capable of handling both upper and lower case characters. This requirement completely eliminates all forms of commercial teletypewriter equipment -- except for mere experimentation on the principle of connection to a remote computer. Teletypewriter equipment is not outrageously priced (merely high priced) -- in the range of $1600 for the simplest of the page printing keyboard machines. The cost mounts with paper tape punching and reading capability -- but this is not really necessary to the operation of the reactive typewriter. *

The requirement of upper and lower case, with the capability of wire connection raises the cost prohibitively. There remains, thus, either a Flexowriter (or its equivalent) at about $2500 for the simplest unit, or a standard electric office typewriter with a unit to "feel and pull" the typewriter keys. The latter unit translates a set of electric signals to mechanical actions for page printing, and translates keystrokes to electric signals for operation in the other direction. "Feel and pull" units cost about $2000. The signals produced by a Flexowriter, or other machine combinations, come out as currents on six or eight separate wire pairs. A telephone connection is restricted to a single wire pair. Thus, a commutator is required to merge or multiplex the eight channels into one for sending, and to unmerge or separate from one into the eight for receiving. This sounds simple, and is simple in principle, but the cost of a suitable commutator seems to be about $7000. Thus, the total cost of the typewriter and associated equipment to this point is about $10,000.

The telephone company does not permit direct wire connections to telephone lines. The single wire pair from the commutator must go through a "black box" furnished by the telephone company before the typewriter signal

---

* We now learn that by the end of the year 1962, there will be two new Teletype keyboard page printers. These will have a price of less than $500. However, because of the continued lack of upper and lower case alphabetic characters, and because of the peculiar "feel," and lower action, of teleprinter keyboards (quite different from that of office typewriters), these machines at this stage are not the final answer.

can pass into the telephone system. The "black box" is called a "Data-Phone", the cheapest version of which rents for about $15 per month.

There are several versions of the Data-Phone for different transmission speeds and at different prices. It is a fair criticism to point out that typewriter terminal sets (like the Flexowriter) have been in active computer use for over ten years, and are widely known. Nevertheless, none of the telephone company Data-Phones was made or was offered in a form which would accept the standard eight-channel signal from this kind of equipment. The user was left to effect and pay for the necessary commutator hook-up for use with the Data-Phone.

INEXPENSIVE TYPEWRITERS CAN BE PRODUCED

Little electric portable typewriters are in the $180 range. Fully mechanized versions (with the power carriage return necessary for remote control operation) cost about $250. It would seem that an industry which can make such excellent machines at these prices could produce complete wire-connected versions for about $500. Further substantiation comes from a German device -- for about $700. A complete wire-connected keyboard with the necessary internal commutator, tape punching, tape reading, and typed output on a paper strip, is available in a single machine. This is durable, of heavy-duty type construction -- however, it doesn't have the lower case. The U. S. IBM "golf ball" typewriter (the Selectric) costs about $445 in the standard size. This machine is unusual in that it creates a six-bit code automatically in its internal mechanism. For this reason, the golf ball machine doesn't need feel and pull units for 44 keys. It would require only six micro switches and an equal number of pulling solenoids for operating the key mechanism. For machine controls (like carriage return or shift) it would require about six more sets. Nevertheless, these simple additions raise the cost to about $1200. Additionally, the modified Selectric requires the expensive commutator for signal transmission into any wire pair like the telephone line. The total cost of a two-way typewriter terminal would thus be about $8200.

Prices of $8000 to $10, 000 for a typewriter with a telephone connection are excessive per se and for libraries. It is my contention, on the basis of the above data and other inquiries, that a typewriter-wire terminal outfit should cost less than $1000 complete, and closer to $500. Until we can get such equipment at such prices, it seems fair to assume that equipment manufacturers are not really interested in servicing the most important area of bibliographic operations -- the repetitive keyboard drudgery of the library.

The way to get such equipment is by consumer insistence and pressure for its production and availability at reasonable prices. We as consumers must first realize that such equipment can be made. We must then relate it to our uses and needs. We must insist upon reasonable price lines. There is no need for gross overcharging, as illustrated above.

THE COMPUTING MACHINES AREN'T READY

Computer manufacturers have been equally delinquent. Let us consider some of the situations at the computer end of the connection, given a battery of remote reactive typewriters operating over the telephone lines into the computer. Trying to handle the situation on current computers would give programmers nightmares. There are several necessary little hardware tricks that not all computers have. There are several other little hardware tricks that no computers have. Since all the computers are "general purpose" machines, one could compensate for part of the deficiencies by programming -- but only with tremendous inefficiencies in operation. The inefficiencies would be in the order of 100 to 1, and costs would increase accordingly. Let us see what some of these hardware tricks are.

The first is the "interrupt" capability. A computer with this capability can go about its regular computing activities while, across the town, the user sits at his reactive typewriter and thinks. Until he presses a key on his typewriter, the computer can be doing other things. On the striking of a typewriter key, the computer interrupts its other work, quickly records the character struck, and returns to its former work until the next key is struck. Action times are short. For all practical purposes the other work in the computer is completely unaffected. Thus, the user has immediate access to the computer, but the cost is only for the split second needed to record the typed character. The machine rental charge should be about 1/1000 of a cent per character. A line of typing should cost about 1/10 cent. A full page (double spaced) should cost about 2.5 cents. It should cost about the same to have the page typed back. Future generations of computers should be even more economical.

This is the basic idea. In practice, not all machines have the interrupt capability, though it has been known since 1959 from plans announced in France by Machines Bull for the "Gamma 60" computer. In IBM equipment, interrupt is an optional extra-cost feature available only on a few of the latest and largest machines. Most of the recently designed machines of other manufacturers have interrupt as a more or less standard feature.

However, more than interrupt is required. If action at a remote typewriter is to interrupt other work in the computer, the computer must be able to keep track of at least two actions at once -- the regular program, and a program for the typewriter when it is activated. Thus the computer must have a "multi-programming" feature. Again the French "Gamma 60" had it in announced plans in 1959. Now, a full computer generation later (computer "generations" are usually figured at three years), only the Minneapolis-Honeywell "MH-800" has a well-worked-out multi-programming feature.

If multi-programming is to be fully safe for all the programs, the computer must have internal markers in its store or memory, so that definite storage areas are reserved for each program. Without this provision, one user may inadvertently run over and erase or change someone else's work. Monitors built into the machine hardware are used for this purpose. The monitors set up the necessary fences between each of the numerous programs. Appropriate monitor circuits and registers do not exist in any current U. S. machine offerings, though they seem to be available on the British "Atlas" machine built by Ferranti.

Another essential feature for using low-cost interrupt in serving many customers is the ability of the machine to charge each user for only the split seconds that the machine is actually doing work for him. To do this, the computer should have counting registers, effectively one for each of the currently active subscribers. These registers would keep track of the exact number of machine operations performed for each subscriber, and would charge him accordingly. There would be an additional charge (or charge multiplying factor) for the amount of memory space the user had under option or control. To my knowledge, none of the available machines in this, or any other, country has this capability.

Another problem arises from the desirability that a large number (a hundred or a thousand), subscribers should be able to be served at their convenience by a single machine. To permit this number of input channels, without an immense cost in computer hardware, components will have to be arranged in the same manner as telephone switching equipment. Since not all (perhaps only fifty) subscribers would be active at one time, the input equipment should be arranged to accept only active subscribers -- and only as long as they continue to be active. The telephone company is able to accommodate thousands of subscribers in one neighborhood with only about a half dozen dial code digit receivers (technically called "senders"). Fast switching between characters necessitates only a very small number of input channels for a large number of subscribers; this is currently being used in transatlantic telephone communications. In such a connection, a different channel may be allocated for each spoken word. A channel may be used by Speaker B the instant A stops speaking, and fifty subscribers may be talking at full tilt on only thirty channels. Until computers are also able to use some of these equipment-sharing tricks, the costs of serving a large number of reactive typewriters will be excessive.

Another hardware problem arises from the need for information storage and automatic rapid availability. This is not now possible without manual change magnetic tape reels. It is not likely that any present computer accepts fifty tape drives connected to a single machine (though this is possible with some machines). The availability of such stores should be coupled with an ability (as in the present Ferranti "Atlas" computer) to transfer large blocks (called "pages") of information rapidly upon demand.

SOFTWARE FOR THE REACTIVE TYPEWRITER

Equally important in the purchase of computer hardware -- sometimes more important -- is the accompanying "computer software" package. This includes programs which help write programs (e.g., programming languages like COBOL, ALGOL, or FORTRAN) and programs for "housekeeping" functions. Programming languages are necessary because writing computer instructions in the language of the computer itself is difficult, tiresome, and usually a source of mistakes.

When instructions are written in a programming language such as COBOL, the computer first translates the COBOL words to its own computer language. Only then does it work through the problem, using its own language. A great advantage of the programming languages is that a few dozen statements in a programming language can be substituted for hundreds or thousands of statements in computer language. A more important advantage is that the programming languages can be specifically tailored to do a specific job. Thus, FORTRAN is good for some forms of mathematical computation, and COBOL is useful for business and accounting purposes.

THE TRAC LANGUAGE

Programming languages for text manipulation are rarer than for arithmetic work. Currently, COMIT, XTRAN, LISP, and FORTRAN-III (Refs. 8, 9, 10, 11) have capabilities in the required direction. Of these, COMIT is of special interest, since it was devised primarily for work on text for purposes of linguistic analysis and experimentation. However, none of these systems was designed for the requirements of reactive typewriters engaged in text manipulation. Therefore, I have designed such a programming language, which I call the TRAC System (an acronym based on "text reckoning* and compiling" system). (Ref. 12)

A few of the features of the TRAC System are that it allows text and programming statements to be intermixed; it allows programming statements to be treated as text; it has full capabilities for naming any of the entities and for storing them and recovering them by name; it is designed specifically to be operated with a reactive typewriter though it is not limited to such use; it allows recursive statements; it allows definitions and nested definitions; and it is especially capable of performing substitutions, rearrangements, and decisions.

In designing the TRAC language, it has been my intent to make it simple enough so that users with library backgrounds can apply it in defining formats for, and producing, catalog cards, book lists, and purchase orders. At the other pole, I wanted it to be sufficiently powerful so that advanced

---

*The reader should consult an unabridged dictionary to find the range of meaning for "reckoning" -- a set of meanings that is more appropriate than dictionary meanings of the usual words "data" and "processing."

problems in symbol manipulation could be performed with it. In the work on TRAC language to date (three versions with successive improvements have been made), it appears that these goals will be realized. A translator for TRAC is expected to be written after the language has been established and checked out.

SEQUEL: THE LIBRARIAN GETS SOME ANSWERS

Now let us try to give some answers to the question of the real librarian: "What should we be doing at our library on information retrieval?"

1. Take some cheer that there may be some techniques and equipment which will eventually help him to solve a set of meaningful library problems. 2. Learn how a reactive typewriter tied to a remote computer can help toward solution of these problems. 3. Begin a campaign with fellow librarians to try to get such facilities commercially at reasonable prices.

But in the interim, much else is to be learned. There are the new developments in descriptors for indexing to the ideas in a document. There are techniques for stringing descriptors together by machine to generate systematically sets of consistent subject headings and index entries for printed indexes to be compiled electronically. I call these "lattice indexes," and a paper on these is in preparation. (Ref. 13)

The basic principles and potentialities of machine operated typewriters can be learned now, before wire-connected reactive typewriters become available. Some of the more limited potentialities have been described. "The Tape Typewriter Plan -- A Method for Cooperation in Documentation." (Ref. 14) A familiarity with the basic principles of computing machines and their programming must be learned, since a reactive typewriter does everything a computing machine can do -- only more conveniently.

As to the large-scale implications of all of this to the library, I venture the following statements: the book will be with us for a long time. The book is too good a piece of human engineering to be discarded easily. Likewise, users will demand full-size copies of documents, even though the images may be stored in micro form. I expect the retrieval selector machines to move into the background, to be used only in special situations. For some applications, the selector mode of using the reactive typewriter will do the job. For others, machine-compiled lattice indexes or their equivalent will be used. Thus selection will be just a special case of the capabilities of the reactive typewriter connected to a TRAC-guided computer. Inter-library cooperation in bibliography via machine readable records (such as punched paper tape at first, then magnetic tape) will become important. This will soon be supplemented by wire links between reactive typewriters (and their computers) at various libraries. Central cataloging services, like those of the Library of Congress and ASTIA, will soon furnish machine readable bibliographic material on tape for incorporation into local library catalogs. Cooperation and services of this kind will cut down on the enormous duplication at each library.

A big change will be the disappearance of the card catalog. It will be replaced by book form catalogs compiled periodically and brought up to date by electronic machines. Current high-speed print-out machines for computers now can print from 36, 000 to 60, 000 lines of double column text (two pages at once) in an hour, with the material arranged in whatever fashion it will be the most useful. The biggest reason that this has not yet been done has been the lack of high-speed printers with a lower case--and this limitation no longer prevails. Now, the real impediment to such a change-over is the necessity of re-typing the entire contents of the card catalog which is to be put into a machine-compiled book catalog. However, if future cards for the card catalog are to be prepared by a typewriter that preserves a paper tape or other machine record, this machine record will furnish the grist for the automatic compilation of a book catalog for all the items that are so processed.

Machine derivation of index terms from the text of a document will come along slowly. It will be faster in some fields than in others. As we have seen, different kinds of library materials differ in the treatment that can be given to the superficial text word, as compared to the ideas behind the words. Where indexing must be to ideas, progress of an "artificial intelligence" variety for machines will be needed before a good machine derivation of the descriptor type of index terms can be expected. Luhn has shown (Ref. 15) that word extracts can be made which are useful for some kinds of indexes for some kinds of texts.

The real challenge and opportunity lies in the reactive typewriter. For the library, it will mean a great simplification and facilitation of a variety of clerical tasks. In the hands of the user of information -- the scientist or the engineer at his desk -- it will permit the machine filing and rearrangement of texts. And so on. In this paper, only hints of the full capabilities of the reactive typewriter could be given. I predict that -- in some version -- it will soon be with us, and will revolutionize office clerical procedures.

With a reactive typewriter now, as I am typing out my rough draft, my task of writing would be nearly complete. The reactive typewriter would make a clean copy of the paper, inserting all corrections and deletions, and would put it in a uniform page format for me. It would help me in locating and inserting references, and it would do so without requiring manual recopying of anything, since the pertinent references would be stored in an index in the computer store. If this report had tables, it would help me to set up the tables, and it would insert the numerical data in the appropriate places with correct alignment. If I needed a computation, the reactive typewriter, with its set of numerals, would operate like a calculator. Finally, if I wanted a nicer job of typescript than the reactive typewriter could produce, the paper could be typed on a special composing typewriter at the computer center, with the "repro copy" sent to me by messenger. Such are a few of the capabilities of the reactive typewriter.

I need a reactive typewriter. You need one. Let us work together to get reactive typewriters for us all!

REFERENCES

1. Mooers, C. N., "Mooers' Law: On Why Some Retrieval Systems Are Used and Others Are Not," American Documentation, p. ii (facing p. 205) v. 11, No. 3 (July 1960); also "Seven System Models," pp. 31-33, ZTB-133-II (Aug. 1959).

2. Taube, Mortimer, "Studies in Coordinate Indexing," (first volume) Documentation Incorporated (1953).

3. Mooers, C. N., "Zatocoding Applied to Mechanical Organization of Knowledge," American Documentation, v. 2, No. 1, pp. 20-32 (Jan. 1951).

4. Mooers, C. N., "Zatocoding and Developments in Information Retrieval," ASLIB Proceedings, v. 8, pp. 3-22 (Feb. 1956).

5. Morse, Rollin, et. al., "Chemical Engineering Thesaurus," American Institute of Chemical Engineers, New York (1961).

6. Heald, J. H., et. al., "Thesaurus of ASTIA Descriptors, First Edition," Armed Services Technical Information Agency, Arlington, Va., (May 1960).

7. Mooers, C. N., "Feb. 1962 Draft Revision of ASTIA Descriptors," photocopy available from ASTIA (1962); being prepared for publication.

8. Yngve, Victor H., "A Programming Language for Mechanical Translation," MT, pp. 299-314, v. 3 (July 1958).

9. Green, Julien, "Symbol Manipulation in XTRAN," Communications of the ACM," v. 3, No. 4, pp. 213-214 (Apr. 1960).

10. McCarthy, John, "Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I," Communications of the ACM, v. 3, No. 4, pp. 184-195 (Apr. 1960).

11. Anonymous, "Reference Manual Fortran II for the IBM 704 Data Processing System," (C28-6000) (1958) plus reference sheets J28-6087, J28-6082, J28-6081 (1960) which give it "Fortran III" capabilities; International Business Machines.

12. Mooers, C. N., "The TRAC System for Text Recoding, Arranging, and Compiling," Preliminary version, V-138 (Sept. 1961), final version in preparation.

13. Mooers, C. N., "The Lattice Index, a Machine Compiled Form of Subject Index," Preliminary version V-144 (Sept. 1961), final version in preparation.

14. Mooers, C. N., "The Tape Typewriter Plan, a Method for Cooperation in Documentation," ASLIB Proceedings, v. 12, No. 8, pp. 277-291, (Aug. 1960).

15. Luhn, H. P., "The Automatic Creation of Literature Abstracts," IBM Journal of Research, v. 2, No. 2, pp. 159-165 (Apr. 1958).

* * *