AFOSR 492

"Implications of Basic Research in Information Sciences to Machine Documentation"

(To be presented by Dr. Harold Wooster, Chief, Information Sciences Division, Directorate of Mathematical Sciences, Air Force Office of Scientific Research, at Third Institute on Information Storage and Retrieval, The American University)

There are at least three talks that I could and should have given this morning: the one Mrs. Bohnert asked me to give--an updated version of a paper of mine on the possible effects of current research in automatic information handling on technical writing and publishing; the talk I had in mind when I wrote the abstract you have before you; and, one based on some long-range plans I have been recently making for basic research in the information sciences.

It would have been, to say the least, convenient, if I could have fed the content of these three talks, together with information on the educational background and interests of the audience, into a computer, and have had the computer produce an optimal, or at least a mini-max, synthesis. To date, however, computers have not been used for belletristic composition. Turning out what my musical friends assure me are medioore musical compositions is no trick at all; as you may have seen on a recent AMF/MIT television show, Grade B Westerns can be plotted, even though the machine occasionally doesn't know any better than to have the villain shoot the sheriff; and at least one major chemical company uses its computer to write routine laboratory reports.

About two hundred and thirty-four years ago today--the year is a little vague, but the day is specific--the first use of a computer for

AD610592

COPY / OF /
HARD COPY $.1.00
MICROFICHE $.0.50

DDC FILE COPY

ARCHIVE COPY

PROCESSING COPY

DDC
FEB 3 1965
DDC-IRA C

original literary composition was discovered and described. You will recall that on Gulliver's Third Voyage, on the 17th day of February (about 1727), he left the Flying Island of Laputa and descended to Lagado, the Metropolis of Balnibari. There, he visits:

"A Professor, with 40 pupils, employed in a Project for improving speculative Knowledge by practical and mechanical Operations. Everyone knew how laborious the usual Method is of attaining to Arts and Sciences; whereas by his Contrivance, the most ignorant Person at a reasonable Charge, and with little bodily Labour, may write Books in Philosophy, Poetry, Politicks, Law, Mathematicks and Theology, without the least Assistance from Genius or Study. He then led me to the Frame, about the sides whereof all his Pupils stood in ranks. It was about Twenty Foot square, placed in the Middle of the Room. The Superficies was composed of several Bits of Wood, about the Bigness of a Dye, but some larger than others. They were linked together by slender Wires. These Bits of Wood were covered on every Square with Papers pasted on them; and on these Papers were written all the Words of their Language in their several Moods, Tenses, and Declensions, but without any order. The Professor then desired me to observe, for he was going to set his Engine at work. The Pupils at his Command took each of them hold of an Iron Handle, whereof there were Forty fixed round the Edges of the Frame; and giving them a sudden Turn, the whole Disposition of the Words was entirely changed. He then commanded Six and Thirty of the Lads to read the several Lines softly as they appeared upon the Frame; and where they found three or four Words together that might make part of a Sentence, they dictated to the four remaining Boys who were Scribes.

This Work was repeated three or four Times, and at every Turn the Engine was so contrived, that the Words shifted into new Places, as the square Bits of Wood moved upside down.

Six Hours a-Day the young Students were employed in this Labour; and the Professor showed me several Volumes in large Folio already collected, of broken sentences, which he intended to piece together; and out of those rich Materials to give the World a compleat Body of all Arts and Sciences; which however might be still improved, and much expedited, if the Publick would raise a Fund for making and employing five Hundred such Frames in Lagado, and oblige the Managers to contribute in common their several Collections."

Apparently the idea of union catalog collections, and, for that matter, the request for lots of money to promote a particular scheme for information processing, did not originate in the Twentieth Century.

Swift, like any good satirist, usually wrote with a specific target. In this case he was satirizing what I am willing to contend was the first recorded attempt at the mechanical coordination of index terms.

Not quite 700 years ago a Catalonian mystic, Raymond Lull, after many days of fasting and contemplation, had revealed to him the basis of his Ars magna--the earliest attempt in the history of formal logic to employ geometric diagrams to discover nonmathematical truths, and the first attempt to use a logic machine to facilitate the operation of a logic system. With Lull's device, sets of index terms were placed on as many concentric circles as there were sets; rotating the circles formed tables of combinations, or logical products and sums. The Model 1270, or Figura Universalis, could handle 14 sets of terms.

It would not have taken more than a few holes in these disks to make this into a polar coordinate Peek-a-Boo system--you can buy a two-disk model today for the mechanical translation of, say, French into English, if you ask for a "verb wheel"--but, like many other inventors, Lull preferred to concentrate on the brochures rather than the hardware. For example, he even produced a book on how preachers could use his Art, complete with 100 sample sermons produced by his computer.

If Lull is ever canonized, he would make an ideal patron saint for documentalists--for the very same reasons that the Church, although it has approved his beatification, will probably never canonize him. Namely, that his martyrdom seems to have been provoked by such rash behavior that it takes on the coloration of a suicide, and that his insistence on the divine origin of his Art and its indispensability raises serious questions about his sanity.

Inevitably, Lull's claims brought counter claims. Francis Bacon, for example, wrote in De augmentis scientiarum words you might expect advocates of hierarchical indexing to apply to coordinate indexing right now, if documentalists today could write as well as Bacon:

"And yet I must not omit to mention that some persons, more ostentatious than learned, have laboured about a kind of method not worthy to be called a legitimate method, being rather a method of imposture, which nevertheless would no doubt be very acceptable to certain meddling wits. The object of it is to sprinkle little drops of science about, in such a manner that any sciolist may make some show and ostentation of learning. Such was the art of Lullius:

such the Typocosmy traced out by some; being nothing but a mass and heap of the terms of all arts, to the end that they who are ready with the terms may be thought to understand the arts themselves. Such collections are like a fripper's or broker's shop, that has ends of everything, but nothing of worth."

The whole point of the foregoing, aside from displaying what Bacon might call ostentatious erudition, is to point out that much of what we are doing in documentation today is not really intellectually novel.

In Bernal's book on history and science there is an interesting passage on the scientist and the engineer; I suggest that where Bernal says "scientist" you might like to think about theoretical studies in the information sciences, and where he says "engineer" you might like to think of conventional, or even advanced, good documentation handling practices. Bernal writes:

"The functional aspects of the scientist and the engineer are radically different. The scientists's prime business is to find out how to do things; the engineer's is to get them done. The responsibility of the engineer is much greater in the practical sense than that of the scientist. He can not afford to rely so much on abstract theory; he must build on the traditions of past experience as well as try out new ideas. In certain fields of engineering science still plays a subsidiary role to experience. Ships today, although full of modern scientific devices in their engines and controls, are still built by men who have based their experience on those of older ships, so that one may say the building of ships, from the first dugout canoe to the modern liner, has been one unbroken technical tradition." (This reads equally well if you substitute "library" for ship.)

"The strength of technical tradition is that it can never go far wrong. If it has worked before it is likely to work again. Its weakness is, so to speak, that it cannot get off its own track. Steady, accumulative improvement of technique can be expected from engineering, but notable transformations only when science takes a hand. As J. J. Thompson once said, 'Research in Applied science leads to reforms. Research in pure science leads to revolutions.'"

One caveat should be entered to Thompson--applied science can produce developments that look like revolutions, but aren't, by making feasible the previously impractical. For example, there is almost nothing on the modern automobile, from column shifts through automatic transmissions to transaxles, that wasn't invented in the early efflorescence of the automobile in the 1900's, but the technologies of metals and machine tools just weren't ready for them. Similarly, for example, the posting and storing of books or documents in serial order of accession is not new--even though I can still make scientists shudder by telling them about the Crerar system of storing books. The Vatican Library got along nicely for 500 years with chronological accession numbers. Classification systems were invented so one wouldn't have to go through every book in the library to find the one wanted; only recently has the speed of the computer made just this process an attractive brute force method for literature searching.

There is, then, at least one question that should be asked about any piece of work in the field of documentation--will it lead to qualitative or quantitative improvements?--will it produce differences in kind, or differences in degree?

If it will produce differences in degree it is, in last analysis, an engineering problem. In talking engineering, one should bear in mind two definitions of a good engineer:

A. A man who can do for a dollar what any damn fool can do for five dollars.

B. A man who can tell whether a system will work before it's plugged in.

My choice of figures in A is deliberate. You will notice that I have stayed comfortably within one order of magnitude. The only engineering advance that I know of that has cut costs by two orders of magnitude is the ball-point pen. I am getting used to computer types saying, "Computers don't cut your costs; they raise your standard of living." I have no great faith in statistics of actual cost per unit of effort--be it abstract, index term, or what have you--being reduced by a factor of 5 over what old-fashioned people can do, nor, for that matter, the time it can be done in.

I readily admit that total cost is only one factor affecting purchase, be it of information systems or automobiles. When people go looking for a new car, they don't always buy the minimal set of wheels for their actual needs. It's nice to have the latest automatic model to impress the neighbors, or to think that a car just like yours lapped Daytona at 150 mph, even though you never go over 60. Both of these same factors seem at times to apply to the purchase of information systems, plus at least three additional ones:

A. It's nice when you're spending someone else's money, especially if it's Federal money. About a year ago I participated in a symposium for those using IBM equipment in actually operating information systems. After a few papers it was easy to tell which systems were operating out of the company's own pocket, say on fixed-price contracts, and which were charged against cost-plus contracts, or used in computing the general overhead.

B. Sometimes--and I'm afraid this applies to private industry just as to the Federal government--it's a lot easier to get money to buy machinery than to hire another person. Somehow, to some management types, a computer is a lot sexier than a cataloger!

C. Sometimes--and here we're usually talking about race cars rather than passenger cars, about military systems rather than civilian systems--it's penny wise and pound foolish not to buy the best and latest, to pay for speed which you need, must and will use.

There is another set of factors relating to my second definition of an engineer--a man who can tell whether it will work before you plug it in. I don't even want to talk about the people who know it will work because it's just like the last one they built, but rather about the builder of a radically different system. He can't always foretell, but the sounder the theoretical structure, the better the theoretical understanding of the principles involved, the better the chance of predicting success. Theoretical knowledge, gained in advance and at leisure, is far less expensive than applied knowledge, acquired on a crash and overtime basis to patch up a poorly designed piece of gear. Algorithms, the basic rules telling a computer what to do, are relatively inexpensive, but the progression from algorithms through flow-charts through de-bugged programs, to actual key-punching and experimentation raises costs by at least one order of magnitude. And, you can't build systems on unsound algorithms!

Eventually you do have to apply your algorithms--pragmatism, after all, is the peculiarly American philosophy--and build a pilot plant to get an estimate of actual production costs. Keeping in mind a cardinal tenet of basic research--the more basic the work, the broader the application--it may be legitimate and desirable to finance such experimental systems, even from Federal funds, remembering always that the test is not whether the consumer will accept the system as a gift, or even make the small convenient down payment, but whether he will keep up the payments when the novelty fades, and the chrome starts to rust.

About now, it would seem, I should take a deep breath and pay some attention to the nominal title of this talk. Even this title has a connotation that is not necessarily true, that there is always and inevitably a smooth unbroken deductive flow from basic research through applied research through development to operating systems, since there are at least two avenues to invention. Intuitive inductive inventions can be and are being made without obvious access to basic research. One of the most interesting methods of high-density information storage, for example, came from a crass commercial attempt to build better color television. As an administrative scientist, however, I cannot anticipate such inventions. I can only play probabilities, try to place intelligent bets on long-shots at the $2 window, and remember to keep cranking lots of elevation into my program to keep ahead of the thundering hordes of hardware builders.

Let me try briefly, then, to give you some idea of what basic research in the information sciences is about. The central problem can be stated very simply: "Can all knowledge, and the process of attaining and using it, be symbolically represented and, if so, how?"

Let's start off with some definitions: Information is the knowledge which man uses to operate on his environment. This environment may include material things, energy, people, or combinations of some or all of these. Information is characteristically compiled from a large number of elements or data which are collected, screened, ordered, combined, perhaps reprocessed after interpretation, and presented in a form appropriate for use. These events in the treatment of information imply the existence of elaborate apparatus or organizations to carry out and connect the steps in the creation process. These will be called information systems. Information systems may take many forms. They may be concerned with gathering, processing and interpretation of intelligence data; collection and evaluation of force dispositions to aid in command and control of operations; or inventory control in a widespread logistics complex. Such systems may be directly concerned with the information only, or they may be concerned both with the information and its use.

The information sciences, then, may be considered as those which are basic to the understanding and creation of information systems. They embrace the four functional areas of pattern recognition, lexical processing, decision making, and encoding for communications and control.

At this glance, there are perhaps thirteen research activities which seem most relevant to solving the central problem:

Self-organizing systems, or intelligent automata.

Multi-dimensional and nonlinear transforms and weighting function theories.

Research in the biological sciences pertaining to sensory perception, neural networks, and memory.

Research in psychological and social sciences pertaining to gestalts, universals, intelligence and values.

Research in heuristic and adaptive computers.

Research in encoding of basic information sources.

Research in linguistics and languages.

Research towards better quantization of value judgments.

Research in theoretical foundations for concepts, such as "information", "decision", "recognition", and "control".

Research in adaptive control systems.

Research in psychological and social sciences pertaining to concepts of control of other than "physical" things, such as attitudes, motivations, behavior.

Basic computer technology pertaining to high-speed, reliable content analysis, storage and retrieval, decision processes, encoding and decoding as a point of departure from high-speed arithmetic.

Basic technology pertaining to many-fold increases in component miniaturization.

This stakes out a pretty broad field of human intellectual endeavor. I wish I could be equally sure that this includes all the areas which will yield useful by-products for machine documentation.

I recently took part in a very heated debate of the program committee of an organization planning a forthcoming national meeting in the general area of documentation. I found myself very much in the minority with my contention that documentalists came to such meetings in the hope of broadening their intellectual horizons; the majority contended that people came in search of specific tricks, knacks or sleights they could take home and use immediately in their own libraries. These I cannot promise. Neither, I am afraid, can I reassure, except for the next few years, those who are happy with their routine intellectual skills. If a job can be formalized, if the steps can be described in such dull detail that anybody can do them, sooner or later a machine will be able to do them, and the competition with the machine will be quantitative, not qualitative. Man has become accustomed, since the Industrial Revolution, to the replacement of tools, set in motion with man's physical strength, by machines. The extension of man's physical senses by such devices as the microscope, or by machines, such as the electron microscope or the radio telescope, are also taken for granted. In the mental category, digital computers are used as tools. They can do very fast arithmetic, but need to be told in excruciating detail what to do. One of the implications of the Information Sciences research program is the possibility that some day computers can be used as mental machines, capable of relieving man of the burdens of routine decision-making.

The first "saboteurs" were French weavers, who threw their wooden shoes, or sabots, into the newfangled automatic looms they feared would replace them. It behooves us to act like sages, not saboteurs, and try to keep ahead of the machines. John Henry beat the steam drill down, but he died with his hammer in his hand. I suggest that, at the very least, we learn how best to use the steam drill until something better comes along, and be able to recognize it when it does.