STEPHEN E. FURTH

Man-Machine Interaction in Publishing Technical Literature*

STEPHEN E. FURTH

IBM Corp., 112 E. Post Rd., White Plains, N.Y.   10601

Received August 12, 1968

Technology advances make it possible to assist authors in communicating with their potential readers faster and more effectively by a variety of media compatible with existing or planned information switching methods. Man-machine interaction in the creation of technical information will change both the format and content of the technical literature.

There has been an escalation in the number of papers dealing with the "Information Explosion" and in the number of suggestions on how to deal with it. Many of these suggestions have come from experts in the new technology—computers, communication, and micro recording. Very few have taken a look at the content and format of our technical information to determine if the new technology made a redesign of this literature desirable or possible and what effect this would have on our ability to cope with the volume of information being generated.

Study after study indicates that only a relatively small percentage of our information needs are satisfied by the published literature and that much of the information transfer takes place informally. The scientist and engineer rarely depends on the published literature in his field of competence to keep himself informed. The time lag between the occurrence of a significant event in the progress of an R and D project and the availability of information about it in the published literature is such that he must do without such information or obtain it elsewhere. In areas peripheral to his field of interest, there is a need for continuing alertness which is met in part by review articles, in part by abstracting or current awareness services.

The application of the new technology so far has been largely limited to the "post publication" processes of information transfer and these introductory remarks were intended to make the point that there is a need to examine the available hardware and software for its potential use in improving and accelerating the prepublication processes. Such an examination should extend beyond the format and distribution methods for our technical literature into the possible use of computers to improve the quality of the content by interaction of the machine system with the generator of information.

The 1963 report of the President's Science Advisory Committee on Science, Government, and Information[1] emphasizes that "transfer of information is an inseparable part of research and development." The committee made recommendations to the technical community which are relevant to the topic under discussion here.

Recommendation No. 1 states that the technical community must recognize that handling of technical information is a worthy and integral part of science. This recognition certainly implies the exploration of the technology available to generators of information in producing information that uses consistent terminology, that can be used to create text and images that can be published in a variety of media, and that can—without delay and human effort—become input to an information dissemination, storage, and retrieval system.

Third generation computers with their ability to store a large volume of data "on-line" are being used by writers using electric typewriters, connected by communication links to the computer, to create information. Today the communication link is telephone or microwave, tomorrow satellites may be available. Many remote stations can be connected to a computer and the internal speed of operation makes it possible for each user to consider the computer to be at his service only. By keyboarding the information in a normal manner, hard copy is produced at the remote typewriter while the same information is stored in machine-readable form in peripheral storage devices attached to the computer. By the use of on-line typewriters the entire process of creating the final report can be speeded up greatly. Corrections while typing a draft are simply a matter of backspacing and retyping the correct information. Subsequent corrections or changes resulting from an edit of the draft do not require retyping of entire pages, but merely the typing of the correcting information, the action code, and location of the word, sentence, or paragraph to be changed.

Retyping of the draft or preparation of the final report for publication is accomplished automatically by the computer under control of a program that will determine format of the page, do the page numbering, type continuation headings, etc.

The output of the computer does not have to be on a remote typewriter which, of course, is relatively slow. Text, in upper and lower case, with subscripts and superscripts and diacritical marks can be prepared for reproduction on a high-speed computer printer at the rate of approximately 250 lines per minute.

Where multiple typefonts are needed, the computer can prepare as output either punched paper tape or magnetic tape, which can be used to drive line casting or photo-

composing equipment. The computer program controls the line length as determined by the user, hyphenates and formats pages to provide space for illustrations.

Operating in this on-line manner permits great flexibility in inserting portions of text or data already available in machine-readable form or rearranging such text in any manner desired. Elapsed time from manuscript to final copy and effort to produce reports against deadlines are reduced.

Another recommendation of the President's Advisory Committee is that the individual author must accept more responsibility for subsequent retrieval of what is published. Specifically, the report urges authors of technical papers to:

a. Title paper in a meaty and informative manner.
b. Index their contributions with keywords taken from standard thesauri.
c. Write informative abstracts.
d. Refrain from unnecessary publication.

Creating new information in an "on-line" manner as suggested above makes it possible to use the computer to assist in accomplishing some of these goals. Having the text in machine-readable form permits the application of computer techniques generally known as syntactical analysis.

One such program[2] analyzes text input and through grammatical analysis identifies descriptors. These words, checked against a dictionary stored in the computer, are then used in the creation of an abstract.

The analysis of the text is done in two phases: dictionary lookup and parsing. Stored in the dictionary are words and inflectional endings. Each word in text is checked against the dictionary. If a match is found, a part of speech is assigned. In case a word can be used in more than one word class, several parts of speech are assigned. In case of a non-match, the computer checks for inflectional endings and a part of speech is assigned. Approximately 95% of the words in a text can be identified this way.

Spelling errors and words not in the dictionary are isolated and a determination made by the writer. Use of such techniques will assure correct, that is, accepted word usage and spelling, and may indicate to the writer the richness or paucity of information-bearing sentences in his report.

The third recommendation of the President's Science Advisory Committee is that techniques of handling information must be widely taught. The report emphasizes that "our schools and colleges will have to do more than insist on proficiency in handling language." Proficiency in the techniques of information retrieval is essential. The user of the on-line retrieval system one day becomes the author of a technical paper the next day. His awareness of the needs of the technical community on the one hand and the potential of providing information by the use of modern technology will influence not only the format but the contents of technical papers in the future.

Most of our technical literature will continue to appear on paper, typed or printed on, in an increasing number of cases, by devices attached to a computer, but the availability of large storage capacity for the storage of text and images may make it desirable to print the full report only on demand. The availability of the report would be announced through publication of an abstract through current awareness media.

Having the text of the document stored in machine-readable form for reproduction on demand makes delivery of the document at the users' place of work by messenger, mail, or telecommunication a matter of choice based on urgency and cost. Telecommunication systems are capable of transmitting publications almost anywhere in the world.

Instead of output on paper, computers can also produce microfilm directly from machine-readable information. Microfiche can store many pages on one unit record and additional microfiche copies can be produced inexpensively and mailed to requesters or subscribers. Availability of microforms would also permit the operation of an "on demand" publication system.

For some type of publication, especially those involving line drawings representing three-dimensional models or charts showing the results of computation, such as histograms, frequency distributions, time series, it may be advantageous for the user to view these on a cathode ray tube on-line to a computer, giving him the opportunity to replicate the image using his own data or to change the parameters given in the publication to create new information. This, however, assumes that the tables and data contained in the publication are explained in a convenient manner.

It is in this latter area where the need for creative work in the design of future technical publications seems desirable. Multimedia for the dissemination of information will become available—publishing will mean marketing the information in printed form, or microfiche, or storing the information in a data bank from which it can be retrieved on demand. The tables and underlying data should be explicitly identified and cross-referenced.

The organization of the technical publication of the future should give greater consideration to the need to "re-package" the information in a variety of formats to meet the individual needs of the user. The man-machine relationship in the information transfer process should make it possible to a far greater extent than before to individualize the presentation of data. Required is the cooperation of authors, referees, and editors in organizing the material in a manner that each "part" is identifiable and self-sufficient, but cross-referenced.

The work already begun by professional societies in standardizing the identification of elements of a bibliographic record and standardization of references appearing in the published literature are essential prerequisites to the effective user–data bank interaction envisaged here.

LITERATURE CITED

(1) Science, Government, and Information, A Report of the President's Science Advisory Committee, The White House, January 10, 1963.
(2) Briner, L. L., and G. J. Carney, SYNTRAN/360, Technical Report TR 00 1599, IBM Systems Development Division, Poughkeepsie Laboratory, May 15, 1967.