ABSTRACT.

AS WE WILL THINK. Theodor H. Nelson. 458 W. 20th Street, New York, N. Y. 10011.

Bush was right. His famous article is, however, generally misinterpreted, for it has little to do with "information retrieval" as prosecuted today. Bush rejected indexing and discussed instead new forms of interwoven documents.

It is possible that Bush's vision will be fulfilled substantially as he saw it, and information retrieval systems of the kinds now popular will see less use than anticipated.

As the technological base has changed, we must recast his thesis slightly, and regard Bush's "memex" as three things: the personal editing and file console; a digital feeder network for the delivery of documents in full-text digital form; and new types of documents, or hypertexts, which are especially worth receiving and sending in this manner.

The present implementation of a partial memex system is described.

In addition, we also consider a likely design for specialist hypertexts, and discuss problems of their publication.

BEATING AROUND THE BUSH

Twenty-three years ago, in a widely acclaimed article, Vannevar Bush made certain predictions about the way we of the future would handle written information (1). We are not yet doing so. Yet the Bush article is often cited as the historical beginning, or as a technological watershed, of the field of information retrieval. It is frequently cited without interpretation (2, 3). Although some commentators have said its predictions were improbable (4), in general its precepts have been ignored by acclamation.

In this paper, an effort in counter-discipleship, I hope to remind readers of what Bush did and did not say, and point out what is not yet recognized: that much of what he predicted is possible now; the memex is here; the "trails" he spoke of--suitably generalized, and now called hypertexts--may. and should, become the principal publishing form of the future.

In July of 1945 an article entitled "As We May Think," by Vannevar Bush, was published in the Atlantic Monthly. It bristled with technical references but was actually fairly candid and simple.

It predicted many things. Bush, as director of the wartime Office of Scientific Research and Development, had seen the new ways in which technologies could be combined. In the urbane paragraphs of this article, Bush predicted a variety of useful future machines, including improvements in photography, facsimile systems, computers and miscellany. Depending on how you read it, he predicted, as well as you could hope, devices closely related to the Polaroid camera, the Xerox machine, computer transformation of mathematical expressions, and the telephone company's ESS switching system.

But the article is best remembered for its description of the new ways that scientists and scholars could handle and share their ideas, writing, reading and filing in a magical system at their desks. The system is the famous "memex."

A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.

It consists of a desk, and while it can presumably be operated from a distance, it is primarily the piece of furniture at which he works. On the top are slanting translucent screens, on which material can be projected for convenient reading. There is a keyboard, and sets of buttons and levers. Otherwise it looks like an ordinary desk. (106-7)

(The memex will hold all the writings its master wants to read, and he can read them easily.

If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him, projected onto one of his viewing positions. Frequently-used codes are mnemonic, so that he seldom consults his code book; but when he does, a single tap of a key projects it for his use. Moreover, he has supplemental levers. On deflecting one of these levers to the right he runs through the book before him, each page in turn being projected at a speed which just allows a recognizing glance at each. If he deflects it further to the right, he steps through the book 10 pages at a time; still further at 100 pages at a time. Deflection to the left gives him the same control backwards.... Any given book of his library can thus be called up and consulted with far greater facility than if it were taken from a shelf. (107, col. 1, para. 5-6)

(Moreover, he can compare and annotate them.

As he has several projection positions, he can leave one item in position while he calls up another. He can add marginal notes and comments,... (107, col. 1, para. 6)

(Not only ordinary documents need be held in the memex. The user may make connections between different parts of the things stored. He does this by

associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. The process of tying two items together is the important thing. (107, col. 2, para. 1)

(By this associative technique he may create "trails," new documentary objects that are useful in new ways.

The patent attorney has on call the millions of issued patents, with familiar trails to every point of his client's interest. The physician, puzzled by a patient's reactions, strikes the trail established in studying an earlier similar case, and runs rapidly through analogous case histories, with side references to the classics for the pertinent anatomy and histology. (107,

(These new structures, or trails, may be taken and given to other people.

Tapping a few keys projects the head of the trail. A lever runs through it at will, stopping at interesting items, going off on side excursions. It is an interesting trail, pertinent to the discussion. So he sets a reproducer in action,... and passes it to his friend for insertion in his own memex, there to be linked into the more general trail. (107, col. 2, para. 5)

(And they may be published.

Wholly new forms of encyclopedias will appear, ready-made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified.... There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record. The inheritance from the master becomes, not only his additions to the world's record, but for his disciples the entire scaffolding by which they were erected. (108, col. 1, para. 2)

(It is strange that "As We May Think" has been taken so to heart in the field of information retrieval, since it runs counter to virtually all work being pursued under the name of information retrieval today. Such systems are principally concerned either with indexing conventional documents by content, or with somehow representing that content in a way that can be mechanically searched and deciphered.

This is paradoxical. On the one hand, Bush did not think well of indexing.

The real heart of the matter of selection, however, goes deeper than a lag in the adoption of mechanisms by libraries, or a lack of development of devices for their use. Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. ... [between documents] one has to emerge from the system and re-enter on a new path. (106, col. 2, para. 3)

On the other hand, Bush merely hinted about the use of structured-data representations and calculi in storing ideas (105, col. 2, para. 3), and did not plainly relate them to his main exposition. While we might argue scripture about such matters, the fact is that Bush's most extensive concern has had few successors in the field called "information retrieval."

TRANSPOSITION

The memex was to be a single screen console for handling the user's notes, writing and correspondence, for reading books and other writings created by others, and for creating new associative text structures, which may in their turn be read and distributed. All this I take to be the heart of Bush's prediction. This will happen. Such systems exist; they are approaching cost feasibility; and the world is readier than it thinks.

Bush's machine will not, of course, be built exactly as he foresaw. The complete description, which I omitted, involves microfilm cassettes, a photographic copying plate for adding new images to the microfilm file, and a telautograph stylus. Other

machines he describes, such as the forehead camera and the direct-dictation typing machine, might or might not have been coupled to it as well. In the revised version of the article (5) his emphasis shifted to videotape. These impediments we ignore. There exist microfilm cassettes, copiers and so on, but the hardware ready to support a memex-class system will be something else.

The system will be built from existing computer equipment and peripherals. Physically it will be a computer display, with a keyboard, at the user's desk; a support computer system (at the desk or elsewhere) for handling various technical chores; and a library network of digital feeder machines. The written materials, when not shown on the screen, will be stored and sent digitally, as telegraphic symbols. They will be sent back and forth among these systems automatically, as programmed in the various devices. The trails, or associative text structures--more generally called "hypertexts"--will be stored in coded form, along with the more conventional documents in the various devices. The user will be billed automatically for the services and the delivery of copyrighted materials. The publisher, who maintains these copyrighted materials in the feeder machines, will be duly paid for their use.

Prototype units exist now. Appropriate console hardware can be purchased now for about $15,000.

Supply systems, however, are not quite ready. The best supply system may be a special-purpose computer or a general-purpose time-sharing system; which is better is not clear. While costs of either are presently in the hundreds of thousands, they are coming down fast, and the use of a well-tailored system by many people at once should bring the cost of such ser-

vice down considerably. To name a figure arbitrarily, let us say that a service cost of $100 a month per user (exclusive of telephone lines and copyright) would be sufficiently low to draw many users. Such service at such a cost will surely be generally available between one and five years from now.

Various preoccupations have delayed us psychologically. We do not need direct dictation, optical scanning or the availability of vast libraries for such systems to be immediately practical and important to us.

THE CONSOLE

We are speaking of a single console to handle notes, writing, much correspondence, much reading, and the creation of new kinds of texts. On it the user must be able to view, edit, file, and otherwise manipulate.

Let me now describe an existing system. This system was created for the purpose of experimenting with various kinds of text handling, and to aid in the production editing of documents which require team study and revision.

The Brown University hypertext editing system, created by Andries Van Dam, Walter Gross and me, has a number of facilities which turn out to be those of Bush's memex. Conscious as we were of the Bush article, we were not paying attention to these parallels during the design deliberations.

Many of these features exist in other text handling systems, notably that of Douglas C. Engelbart at Stanford Research Institute (6). The purpose of this description is to show parallelisms between memex and this general type of system, not to distinguish this system from its relatives.

I will speak of the Brown hypertext system simply as "the system," "the current system" or the like, to distinguish it from the memex. As its implementation is not yet complete, this description applies to the present specifications and not only the parts that are working.

The user sits at a foot-square screen with a desk-top, typewriter keyboard, and a special set of pushbuttons. (This is the IBM 2250, but lesser facilities, in this setting, would be capable of comparable service.)

A. Text handling.

First let me briefly describe the text-handling and text-editing facilities of the system. While this is a sidetrack, since Bush did not discuss editorial work, it is a basic aspect of the system. (This is a great advantage of digital over pictorial storage: the system may manipulate the words letter by letter, rather than as a single image.)

The basic text-handling features of the system are the presentation of text materials on a screen; the ability to command basic editing operations by simple manipulations; the ability to make these editorial changes tentatively, on copies or alternate versions of your material; and the option of having your actions recorded automatically in a cumulative editorial log, in case it is later necessary to retract any of them.

The system is generally geared to tentative and thoughtful operation. The alternative versions and editorial log are the strongest examples, but there are others. An editing command, once enacted on the screen, must still be accepted by the user before it actually occurs in the file. Another example: a section of text being tried in a new place is shown with a

number of spaces to the left of its beginning and the right of its end, so you know its limits within the new setting.

B. Filing and manipulation.

Various techniques must exist for manipulating the file in such a system.

The user of Bush's memex called its contents by means of a "code book" (107 passim)--but this too was in the system. From the code book the user was to choose contents for viewing by looking up and keying in their listed codes--some mnemonic, some shortened over use.

In the current system it is possible for the user to draw something to the screen by typing its user-assigned name, or picking the name from a screen menu with the light pen. He may do this not just with whole units, but also summoning instances of things to be repetitively copied, and different versions--separate copies--of documents to be tentatively edited.

In the memex, the user could skip through a document at adjustable rates (107, col. 1, para. 5). In the current system it is possible to step forward through the text by an adjustable number of lines at a time.

The memex user looked at several documents simultaneously through several screens or panels of the same screen (107, col. 1, para. 6). This will be possible for up to three documents on the current system.

The memex user created annotations by hand, or links between the things that were being viewed simultaneously.

The user of the current system has two possible types of annotation he can create. The first, presently called the "tag," is a marker in the text to which a short note or explanation is attached. This may be typed in at the keyboard. From the tag there is no place to go except back. The second connector, called in this system the "link," joins two items of text. Either item may be anywhere in the user's file. A link may also be made to a nonexistent text item, which the user may then put in by the keyboard.

The current system assumes no structure, but creates links between text sections regardless of whether or not they are parts of the same unit or otherwise related.

In addition, the memex described in the revised version of Bush's article kept a log of manipulations by the user ("Memex Revisited," 95-6). The current system will also possess this capability.

Design problems arise in the richer operations, such as creating connective structures. It is taken for granted that the console must be easy to use. That is no design problem for a small set of operations, such as text editing. But the design of the overall file handling elements and actions is more complicated. The problem is providing conceptual unification for the system's filing structure and conventions.

For example, in the current system, the number of basic file connectors is held to two.

The basic idea is that the tag (signified by an @ in the text) is a machine-scannable annotation, the link (signified by an * in the text) a basic jump from one item to another. Tags may be of various types, and each type may be collected, listed and searched by the machine. Links, marked by an asterisk, may

also have a textual "explainer" telling what kind of a jump is available, or anything else the hypertext's creator wants to attach to the link.

This dichotomy, while simple, should be sufficient to support real hypertext experimentation.

Complex couplings are presently not defined for the current system. However, the textual labelling of links is adequate for various possible forms of discrete-jump hypertexts, including Bush trails, and may in principle be extended to computer responsibility for link behaviors and complex coupling maintenance.

HYPERTEXTS

While Bush's term, "trail," represents a very useful concept, we must generalize it. Bush's interest in microfilm led to his idea of the trail having a sequence.

By "trail," Bush appears to have meant a sequence of documents, document excerpts, and comments upon them.

For [the user] runs through an encyclopedia, finds an interesting but sketchy article, leaves it projected. Next, in a history, he finds another pertinent item, and ties the two together. Thus he goes, building a trail of many items. (107, col. 2, para. 4).

This sequence would be established by making paired couplings.

When the user is building a trail, he names it, inserts the name in his code book, and taps it out on his keyboard. Before him are the two items to be joined projected onto adjacent viewing positions. ... The user taps a single key, and the items are permanently

joined. (107, col. 2, para. 2).

Bush mentions two other types of trail. One is the "side trail," branching out from a main trail sequence.

Occasionally he inserts a comment of his own, either linking it into the main trail or joining it by a side trail to a particular item. (107, col. 2, para. 4)

The other type of trail is the "skip trail," a subset of a main trail sequence that contains the highlights.

The historian, with a vast chronological account of a people, parallels it with a skip trail which stops only on the salient items, and can follow at any time contemporary trails which lead him all over civilization at a particular epoch. (108, col. 1, para. 1)

In Bush's trails, the user had no choices to make as he moved through the sequence of items, except at an intersection of trails. With computer storage, however, no sequence need by imposed on the material; and, instead of simply storing materials in their order of arrival or of being noticed, it will be possible to create overall structures of greater useful complexity. These may have, for instance, patterns of branches in various directions. Such non-sequential or complex text structures we may call "hypertexts." (7)

"Hypertext" is the generic term; there are reasons, for which there is no room here, to rule out such other candidate terms as "branching text," "graph-structured text," "complex text" and "tree text."

The best current definition of hypertext, over

quite a broad range of types, is "text structure that cannot be conveniently printed." This is not very specific or profound, but it fits best.

As Bush pointed out in his own terms, we think in hypertext (106, col. 2). We have been speaking hypertext all our lives and never known it. It is usually only in writing that we must pick thoughts up and irrelevantly put them down in the sequence demanded by the printed word. Writing is a process of making the tree of thought into a picket fence.

Hypertext structures are varied. For instance, they may be free-branching with only one type of link and backing up; they may have modal links with different meanings in a free structure; or have modal links and repetitive structure.

Discrete-jump hypertexts are not the only kinds. There is, for example, "stretchtext." This is continuously variable text which never leaves the screen, but changes by small increments on user demand, growing longer and more detailed by a few words at a time, as required. Other continuous types are possible.

As items may be coupled, whole hypertexts may be coupled into books, or one another. An example of the first would be annotations coupling into the Bible. Such multi-couplings involve bundles of pointers between the texts, possibly with type codes or annotations. They may also involve alternative versions, which there is no room to discuss here. The structuring of these coupler types is a continuing design task for hypertext systems.

The creator of hypertexts may allow the user various options of jumping or branching. These options can lead the user to further reading in any pattern the author wants to make available to him. The only constraints on the author are usefulness, clarity, and art-

fulness.

There must, of course, be ways the reader may see, and choose among, possible branches from an item in the text. This problem was implicit in Bush's treatment. Since "any item can be joined into numerous trails" (107 col. 2 para. 3), there would have had to be some way of showing the user these options and their meaning. This is the case in general, and a standing aspect of hypertext system design.

Hypertexts may be casual rough notes, as described in Bush's extended example of the Turkish bow and arrow; or they may at the other extreme be finished units, editorially completed and organized. Such finished units would have many of the same properties as ordinary writing: intentional assembly, attempted clarity, and expository structure of enumerable "points," and an overall comprehensible pattern whose interrelated parts may be in some way remembered or visualized. Finally, the concept of authorship applies to hypertexts as much as it does to an ordinary book or article.

As with ordinary texts, too, the editorial properties and "feel" of hypertexts may be quite distinct and varied. For hypertexts these are of course largely unexplored. It is also very hard to anticipate their possible administrative and social settings, and this will greatly affect their character and the modes of their use.

THE TRANSMISSION NETWORK

A general transmission network will carry requested documents from libraries to users, new documents from users to libraries, and communications and documents between users.

The network will consist of several computers or

computer-like objects. In the user's own unit is digital logic, and possibly a small computer; this unit is serviced or managed by a computer which stores the user's files and communicates with the library network. The user's requests for documents that are not available locally go out to a library network. These requests are sorted out to the appropriate repository machine; the repository machine returns the document and a bill or fee schedule.

Various fees are logged up to the user. These will include various basic costs, such as membership in the system, rental of the terminal and hookup. Additional fees may include logged-in time, per-usage costs of various facilities (such as average memory area occupied and quantities of text moved), storage charges for materials kept locally, and royalties to copyright holders. It should be noted that various grades of service may also exist, in which the user gets faster service by claiming larger buffers and higher priorities, and pays for the privilege.

Although this may sound like a formidable prospect, in general and with polishing there is reason to hope that the real costs of such a system will compete favorably with the real costs of the forms of publication and libraries we now employ. (Of course, in such "real cost" we must include the library services supplied "free" by various levels of government, including municipal libraries, grants to universities and the indirect subsidy of publishers by low postal rates. It is not unthinkable that similar encouragement will come about for this form of publishing and libraryship.)

Various technical design issues exist. These involve the feeder computers and their forms of memory. These hierarchies of memory are fairly clear.

They will generally include disk (for working areas and directories), magnetic card or data cell (for the corpus), and magnetic tape (for rarely-needed materials and safety copies).

A more difficult question is, what should the feeder computers be? Their job in this system is the lookup and shovelling of text, plus bookkeeping. One school of thought holds that a true general-purpose time-sharing system is necessary; another, that the correct machine is a dedicated computer with rich interrupts and comparatively little arithmetic capacity. The third school would point out the special character of the work and lean toward special designs and special tradeoffs, which could be anything from associative memory to the use of delay-line machines.

Similar issues exist for memory software and directory systems. There are complex technical problems of index and search techniques, and methods of their cross-tabulation. But they can be handled in some way or other.

A warning is necessary here, however. This area of console support is the area where things are not yet ready. The prediction of economic feasibility in five years, an eon in the computer world, is not the same as feasibility now. By devoting a whole computer, disk and tape to each user, the problem of console support can now be solved, in a manner of speaking; but the general problem of interleaved I/O and file management, with the efficient sharing of facilities, is another problem entirely, and the one that must be solved to make this whole thing go.

PUBLICATION: REDESIGN OF THE TECHNICAL LITERATURE

Bush regarded his new text structures as trans-

missible between individuals, and publishable. The same is true of hypertext units, the generalized form of trails. I think it likely that once such systems are available, the creation of branching and complex text will become recognized as far more natural than the structures in which we now must write.

This will all follow naturally from the existence of consoles which permit multiple couplings between texts. Having created for personal use a hyper-document on one's console, it will seem only reasonable to press a button passing this on to a colleague in its hyper-form, without chopping and aligning it into conventional writing.

Various interesting possibilities follow. Private "journals" in a field may be started among co-workers merely by the pooling of their hyper-documents. The rental of memory space on magnetic cards is inconsequential next to what have been the costs of printing and mailing.

When professional and technical societies become interested in sponsoring hyper-publication, one of the most straightforward ways to begin would be with the creation of society-sponsored review articles. These could be like the ordinary review articles sponsored by such societies, save that the review article would open directly into the various materials it was reviewing, and footnotes could be more extensive and slanted to different categories of reader. (Figure 1.)

The next step is, of course, the creation of hyper-magazines or journals under the sponsorship and supervision of professional societies. Here the problem of organization would seem to become thorny. But this collection could be much like the journals of today, except for the direct availability of previous liter-

ature, working papers and various odds and ends. (Figure 2.) It should be observed that any of the "documents" noted in the illustration can themselves be hypertexts.

In this magazine, an arbitrary conjecture, all material of the past year is considered "recent" and embraced in a common lookup structure. New material enters the collection surreptitiously, at whatever time of day or night the editors release it; material one year old is formally expelled to a different file on the first of each month. (It may be just as accessible, but its nominal status changes, rather like that of a book taken off the "recent acquisitions" list of a library.) This magazine would hold most of what people were talking about in a field, and it would all be right there.

There are numerous technical complications to hypertext publishing. There is no room here to discuss the more esoteric technical ones, such as facilities, billing and copyright conventions, or possible techniques of encryption and validation to prevent pirating of works and to authenticate expeditionary versions.

HOW TO BEGIN

How shall it begin?

Those contemplating massive retrieval systems commonly presume that they must begin with some massive corpus all accessible. The Library of Congress is often mentioned. Even Bush supposes regretfully (in the revised article, p. 100) that the personal system waits on large public establishments being automated first.

I do not believe this is so. It will be practicable and of considerable interest to begin on a small scale,

having no grand corpus available. The grand corpus will come soon enough, as requests emerge. (We have a precedent: the prowess of University Microfilms, Inc. in rendering texts available to scholars in microfilm.)

The way to begin is to furnish supported consoles to small communities of users: key members of a "small" discipline, or specialists among whose work there is close connection. Suitable groups might be "early Egyptologists," or just plain "everybody at Woods Hole."

Such communicants, having been assured as well as possible of privacy and fail-safe design, will be encouraged to use the consoles fully. From the outset they may keep all their notes, manuscripts, articles, and copies of outgoing correspondence, on the system.

The rest will follow. I am fairly sure of the predictions so far, at least in broad outline, but I am just as sure that the first generation of hypertext users will invent twice as much as has been described and described so far.

Who will support these beginnings? We have a choice, at the outset, of universities, publishing companies, computer companies (including service bureaus), research organizations or the government. Any of these might take such an initiative. Though such an initiative would seem severally unthinkable, it somehow seems collectively plausible and, of course, historically inevitable. If you believe in manifest destiny.

References

1. Vannevar Bush, "As We May Think." The Atlantic Monthly 176:1 (July, 1945), 101-108.
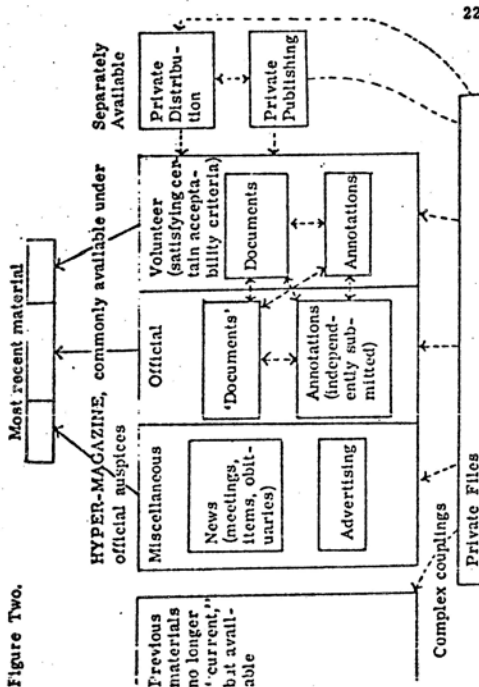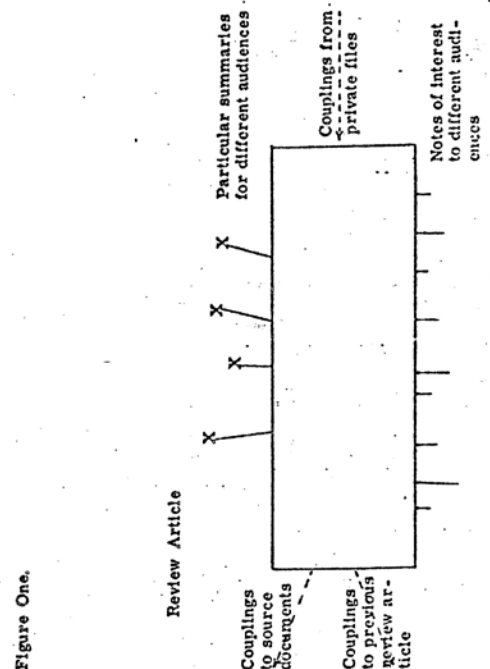
References to "Bush" cite this article unless

further specified, and all unexplained numbers and paginations in the text refer to it. Because the article is so tightly written and our interest in it so close, paginations are given to the column and paragraph.

2. Allen Kent, Textbook on Mechanized Information Retrieval, second edition. Wiley, 1966; pp. 7-9.
3. Joseph Becker and Robert M. Hayes, Information Storage and Retrieval: Tools, Elements, Theories. Wiley, 1963; p. 40.
4. John H. Wilson, "As We May Have Thought." In Progress in Information Science and Technology: Proceedings of the American Documentation Institute 1966 Annual Meeting. Adrianne Press, 1966, pp. 117-122.
5. Vannevar Bush, "Memex Revisited." In Vannevar Bush, Science Is Not Enough. William Morrow and Company, Inc., 1967, 75-101.
6. Douglas C. Engelbart, "Augmenting Human Intellect: A Conceptual Framework." AFOSR-3223 Summary Report, Stanford Research Institute, Menlo Park, California, October 1962.

   This expresses the philosophy, rather than the results, of the continuing project.

7. Theodor H. Nelson, "Getting It Out of Our System." In George Schechter (ed.), Information Retrieval: A Critical Review. Thompson Book Company, 1967; pp. 191-210.

Figure One.

Review Article

Couplings to source documents — Couplings to previous review article — Particular summaries for different audiences — Couplings from private files — Notes of interest to different audiences

Figure Two.

Most recent material — Separately Available — Private Distribution — Private Publishing — HYPER-MAGAZINE, commonly available under official auspices — Official — Volunteer (satisfying certain acceptability criteria) — Documents — Annotations — 'Documents' — Annotations (independently submitted) — Miscellaneous — News (meetings, items, obituaries) — Advertising — Previous materials no longer "current" but available — Complex couplings — Private Files