



Journal of the American Society for
Information Science and Technology

Memex at 60: Internet or iPod?

Journal:	<i>Journal of the American Society for Information Science and Technology</i>
Manuscript ID:	JASIST-2005-05-0037.R1
Wiley - Manuscript type:	Opinion Paper (ASI)
Date Submitted by the Author:	27-Jul-2005
Keywords:	information technology < (hardware, software, and equipment)

powered by ScholarOne
Manuscript Central™

Memex at 60: Internet or iPod?

Richard H. Veith

R. Veith Consulting, 435 Hoffman Road, Port Murray, NJ 07865. E-mail:

rveith@gti.net

Abstract

It has been sixty years since the 1945 Memex article, and so much has changed since then that we might well wonder whether the article is still worth looking at. It certainly inspired some of the leading figures in information technology, but now it seems to be cited either for things it did not really say, or because everything it proposed has been pretty much accomplished, albeit with alternate technology. If we take another look at the Memex description, though, there are a few key ideas that can still be goals in terms of an easy-to-use personal collection that is a supplement to one's own memory. Perhaps in today's terms, the device would be a combination of the iPod design and a tablet computer. As such, it could function as a handy information pod, with certain Memex features, serving as an extended personal memory.

Introduction

Thomas Watson, Sr., chairman of IBM, is quoted as saying in 1943, “I think there is a world market for maybe five computers.” Vannevar Bush said something similar in his 1945 “As We May Think” article: “With machines for advanced analysis ... there was and is no extensive market.” But nobody pays much attention to that. Instead the article is remembered and cited for other reasons, and even (many times it seems) for ideas it did not espouse at all.

My purpose here is to take another look at what the 1945 article really said about the Memex, as opposed to extrapolations and extensions that have accumulated over the years, to see if the Memex article has anything to say to us today. The grand thing about the 1945 Memex description is that technology has so surpassed the dated description of levers and relays and microfilm that we all feel free to pick the parts to take literally, the parts to treat liberally, and the parts to ignore.

Background

Before addressing the 1945 article directly, it may be helpful to summarize the historical setting for the article in three respects: its publication history; the related (but different) machine called the Rapid Selector; and the broader context of the Memex concepts. Because there seems to be much misconception about the 1945 article, this

summary is intended to establish a basis for discussion even though some of it may be familiar to those who know the article and its history well.

Drafts and Delays

There are a number of texts that explain the origin and eventual publication of the Memex article, principally the compilation by Nyce and Kahn (1991b), and sections of the Zachary (1997) biography. As Nyce and Kahn point out, the article was really based on ideas and technology from the early 1930s, and Bush later estimated that he first formulated the Memex ideas around 1932. In 1937 he sought funding from the Rockefeller Foundation to build a Memex-like machine although it was not called that and it differed in some respects from the later Memex description (Nyce & Kahn, 1991a). By the end of 1939, Bush had a draft of an article called “Mechanization and the Record” which he sent to the publisher of Fortune magazine to get the publisher’s reaction. This draft used the name “Memex” for the first time and in fact contained all the material in the 1945 article except for the opening and concluding sections. For a number of reasons, particularly the pressure of national defense matters, Bush set the article aside and did little or nothing with it until 1944, submitting the draft this time (with a new opening section reflecting the coming end of the war) to the Atlantic Monthly. The Atlantic Monthly editor, in addition to publishing the article, decided to add an editor’s introduction to highlight the importance of the article.

The interest in the article was such that Life Magazine requested and received permission to print a condensed, illustrated version of the article in the September 1945 issue. The editors of the Life article, under the heading “What Dr. Bush Foresees,” highlighted five significant things: (1) the mini-camera and dry photography; (2) miniaturized microfilm; (3) speech to text (i.e., a machine that would type when spoken to); (4) advanced mathematical calculators that could generate conclusions from premises (called “thinking machines”); and (5) the Memex, a machine that files items by association, and can run through trails of associations at the touch of a button (Bush, 1945b).

During the several decades following that, Bush began to adjust and modify his Memex concept in light of newer technology and perhaps in response to critical comments about one aspect or another. By the end of 1958, he was writing drafts of a manuscript called “Memex II” (Kahn & Nyce, 1991). For a year or two, Bush tried to get it published in Atlantic Monthly and Life. The Atlantic Monthly was willing to publish it, but wanted Bush to “polish it, which it badly needs.” Instead, Bush dropped the idea of getting it published at that time. In 1965, he was apparently looking again at the Memex idea, and working on an essay that became “Memex Revisited,” published in 1967 in the final collection of Bush’s essays, *Science Is Not Enough*.

Despite the fact that Vannevar Bush modified his Memex concepts somewhat as time passed, it is the 1945 Atlantic Monthly article that is still cited as the basis for almost all of the subsequent claims concerning the Memex.

The MIT Rapid Selector

The Rapid Selector was a mechanical device for selecting microfilm frames, designed by Vannevar Bush, built at MIT beginning in 1938 and developed over a period of several years. The basic idea of the MIT Rapid Selector was that document abstracts would be contained on reels of 35 mm microfilm such that one half of each frame was the 16 mm image of an abstract, and the other half provided space for codes (represented as optical dots). When searching for frames matching a given set of codes, a card would be created (though later the card was replaced by a drum device) with the desired dot pattern, and the reeled microfilm would spin past the card. Photocell detectors would determine when a code pattern in the microfilm matched the pattern on the card, and that document image (the abstract) would be optically copied to 16 mm film. All of this was to happen so fast that 60,000 frames could be searched in a minute. In practice, though, the machine that was built had to be slowed down considerably due to mechanical and electronic problems, it was not too reliable in operation, and the effort associated with preparing and coding the abstracts appeared to be operationally and financially prohibitive (Burke, 1991). This selector, or more specifically a version of it, was really only used for a short period during the war “as something of a pattern locator” for analyzing codes in Japanese messages.

Some commentators have suggested that the MIT Rapid Selector is a good example of how the Memex might have been built, but there are several major differences between that selector and the Memex described in the 1945 article.

The first major difference is the method of selection. Because the Memex concept required extremely fast access to hundreds of millions of microfilmed images, Bush discounted the sequential scanning method of the MIT Rapid Selector, and instead proposed a system more like a telephone exchange, where each digit rapidly narrows the selection range. Thus, if you know the number of the item you are seeking, the machine would arrive at the selected document as soon as the final digit was processed. Bush pointed out that the mechanical relays of the telephone-exchange method could be replaced by vacuum tube switches, enabling selections within a hundredth of a second. It is worth noting that this same idea has subsequently been used in digital storage systems, where the ID number of an item is directly related to the physical path to the data on disk. For example, the British Prestel system was developed in the mid-1970s using such a scheme in order to rapidly handle potentially thousands of simultaneous online requests for Prestel pages.

The second major difference is the method of associating one frame with another. In the MIT Rapid Selector, frames related to each other (i.e., in the same category) would have the same code or set of codes. But in the Memex, the dots in a microfilm frame would actually be the index number of the next related frame. Thus, one microfilm frame is linked to another with the link embedded in that frame, so that when viewing any given

frame, the information is there for the machine to rapidly select the next linked frame in a given trail.

More generally, the MIT Rapid Selector and the Memex had different intended users, and the timeframes for development were different. The Rapid Selector was intended for organizational use where the organization would expend the effort to create the abstracts, classify them, and add the classification codes. Further, it was designed to be built with available technology. The Memex, on the other hand, was intended for personal use (more about this later), and was proposed with near-future technology in mind, i.e., using primarily known methods and elements, extending those in terms of smaller size and faster speed, and assuming near-term resolution of the technical difficulties not yet solved.

The Broader Picture

The 1945 article pulled together a lot of ideas, and it is not a criticism of the article to acknowledge that many of the ideas had also been voiced by others. Robert Fairthorne is quoted as saying that “few of his [Bush’s] suggestions were original” in Buckland’s (1992) review of relevant technological developments that predated both the MIT Rapid Selector and the Memex article. Vannevar Bush was not the first to comment on the accumulating volume of recorded information nor the first to look for machine-based solutions. The following quick summary of the Memex article’s relationship to prior developments addresses three aspects: microfilm for storing masses of documents;

techniques for extremely fast selection of desired frames; and techniques for associating ideas.

With regard to microfilm itself, the Memex concept was on solid ground. In the 1920s and 1930s, microfilm was a popular storage technology, and a grainless microfilm able to contain the entire Bible fifty times within a square inch had been demonstrated in 1925. A workstation containing hundreds of thousands of microfilmed document pages had been developed in the 1930s by Sebille (Buckland, 1992). A microfilm selector using photoelectric cells had been developed in Germany by Emanuel Goldberg at Zeiss Ikon, and the machine was patented in both Germany and the United States around 1931 or 1932. And in late 1932, the same year that Vannevar Bush says he first began putting together the idea of a Memex, Bush was visited by Watson Davis. Davis was well acquainted with developments in both the U.S. and Europe regarding the use of microfilm for storing and viewing document collections, and pitched his own ideas for microfilm to Bush, possibly planting the seed in Bush's mind (Zachary, 1997, pp. 74-75).

At the time Bush was working on his Rapid Selector, in the United States companies like NCR and Eastman Kodak were also working on microfilm retrieval machines. In fact, both were financial sponsors of the MIT Rapid Selector, and Kodak even shared information about its own work with Bush (Burke, 1991). Bush and his team made several attempts to patent their own rapid selector, but the patent claims were denied due to a list of prior patents that "continued to grow" (Burke, 1991).

Regarding the techniques for selecting desired frames, the Memex description does depart from earlier microfilm selectors. Most of the schemes suggested involved attaching a code, or set of codes, directly to a microfilm frame, and then employing a sequential search through all the frames to find the matching codes. This was the method used in the MIT Rapid Selector. But, as mentioned above, it was not the method suggested by the Memex article. Instead, a system similar to telephone switching exchanges would be employed, with the switching done using vacuum tubes for high-speed selection.

Actually, there are two aspects of this selection technique that are worth reviewing. First is the telephone exchange model. Bush seems to have been aware of ongoing developments at Bell Telephone Laboratories, so he may or may not have been familiar with specific attempts to use telephone switch technology for other purposes. For example, beginning in 1937 George Stibitz was using telephone relays to create first a simple adder and later a complex number calculator, called the Relay Calculator. The telephone switch idea, as a selection mechanism, implies that the microforms within a Memex may need to be stored in a physical arrangement corresponding to each item's ID number (perhaps an accession number). For example, when the final number in the ID has been processed by the last switch, the machine should be pointing at the physical location of the microform frame, just as in a telephone switch it finally arrives at the location of the dialed line.

The second aspect of the selection technique described in the Memex article is that any given item contains the actual ID number of the next desired item. In other words, when a topic trail is being followed, one item will contain in the microfilm a dot pattern representing the “index number of the other item” (Bush, 1945a). Given that item A can point to item B, and B can point to C, as well as C pointing to B and B to A, and so on through a series of items, this is a good example of a doubly linked list. In fact, this might be one of the first articles, if not the first article, in computing literature to describe a doubly linked list, since linked lists are often attributed to developments in the 1950s. In any event, the linked list idea meant that selection is still by ID number only, rather than sequentially scanning for all items with a given topic code. In the Memex article, any item could belong to more than one linked list, and the maximum number of lists that any one item could belong to would be limited by the available space on the microfilm frame for dot patterns.

Finally, there is the matter of the association of ideas. Vannevar Bush wanted to break away from hierarchical classifications and he wanted to provide a mechanized way to support multiple associations among items. His argument was that the mind works by association, that one item suggests another because of a “web of trails” in the brain. With the Memex, he was not getting too carried away though; he simply wanted a mechanized way for a person to link any two items together based upon that person’s own reasons for the association. The Memex article was not concerned with long-standing philosophical views of how ideas in the mind are associated, but rather with the subset of logical associations that a researcher might make. For example, items would be associated using

the same principles as when gathering items “together to form a new book” (to use a phrase from the article slightly out of context). It was this restricted notion of association that Bush sought to mechanize.

As Rayward (1994, 1997) has pointed out, others were at work in the early 1900s to find mechanical means for associating ideas, and in particular Paul Otlet and the International Institute of Bibliography beginning around 1895. (We might speculate that Vannevar Bush heard about Otlet’s work indirectly, via the visit from Watson Davis in 1932, but there is no solid evidence one way or the other.) Essentially, Otlet’s scheme was to copy individual ideas or “chunks” of information from articles and books onto standardized cards, which allowed “all the manipulations of classification and continuous interfiling.” The cards were filed within a classification system (Universal Decimal Classification) such that associated ideas would be filed under the same classification number, perhaps in separate cabinets but still under the same classification. This allowed researchers to find related concepts that may have been expressed anywhere within any of the pamphlets, articles or books covered by the system. By 1930, one such collection of cards had nearly 16 million items.

Otlet also worked with microfilm, and was impressed by the potential of radio, television and records in achieving the same information and communication goals as books. In Otlet’s 1934 *Traité de Documentation*, he envisions someday books being replaced by the newer technologies. He also envisioned a researcher’s desk with multiple desktops, electrically controlled filing cabinets such that a desired cabinet would slide

quickly (on rails) into place at deskside, and even television screens for remote viewing of items. In fact, Otlet believed that eventually the researcher's desk would be just a telephone and viewing station, with all the actual materials housed in massive collections elsewhere.

A major difference between Otlet's suggestions and the Memex is that Vannevar Bush pretty much ignored the use of any classification scheme to govern the association of ideas. For the individual researcher this may be reasonable, and the researcher could implement his or her own schemes if desired. But for all the material that a Memex might acquire with pre-established linked lists, there is no statement in the Memex article indicating what would govern the logic of the connections. It is as if Bush, as a scientist/engineer, was principally interested in the mechanics of the device, and would let others decide how to exploit the capabilities of the machine.

Regardless of the degree of originality, there is no doubt at all that the 1945 article was popular in its own time, due in part to Bush's academic and political prominence. It is also clear that the article provided inspiration to later influential system designers. Douglas Engelbart (reprinted in Nyce & Kahn, 1991a) has said "... the article influenced me quite basically." Ted Nelson (reprinted in Nyce & Kahn, 1991a) considers himself a "counter-disciple" of Bush, believing that the article "runs counter to virtually all work being pursued under the name of information retrieval today [in 1972]" but that "Bush was right." J. C. R. Licklider dedicated his 1965 book, *Libraries of the Future*, to Dr.

Bush, saying that while he had not read the article until recently, he had “often heard about Memex and its ‘trails of reference’” (Kahn & Nyce, 1991).

Basic Concepts

The following section examines the 1945 Atlantic Monthly article itself for what it said and what it did not say, as a basis for distilling what I believe are the essential elements and comparing that to what we have today.

Article Components

The 1945 article, taken as a whole, makes a lot of suggestions about future technology not necessarily related to the Memex. The Atlantic Monthly article did not contain any subheadings, only numbered sections, and it is not particularly easy to see the overall structure. However, the article can basically be divided into two parts (aside from the introduction and conclusion) that discuss the creation of data collections and the selection of items from those collections. (See Table 1.) These two topics account for 75 percent of the article, with slightly more space given to the topic of creating the record (43 percent) than to the topic of selecting from the record (34 percent). The specific discussion of the Memex, its trails, and new types of publications occupies just 18 percent of the total article.

One of the largest single topic areas is that of photography and the possibilities associated with electronic photography. Unlike Otlet, who saw a future with television-like terminals for viewing remote pages, Bush saw television as a way to electronically capture a picture so that it might be printed on microfilm.

There is also a section at the end of the article (labeled “Really Futuristic” in Table 1) in which Vannevar Bush engages in a little speculation about distant future prospects for coupling electronic information outside the body directly with electronic information inside the body, thus bypassing the eye, the ear and the hand. Developments in these areas were in their infancy when the article was written and have progressed considerably since then. However, this section is rarely cited, even though it may presage the ultimate user interface for information systems.

What Was Not Said

There are a number of things described in the 1945 article that caught the imagination of readers and that continue to invoke admiration, but the article has also been cited many times for things it did not say. In her surveys of citations in 1981 and 1991, Smith (1991) noted that occasionally authors citing the Memex article do not seem to have actually read it. More recently, the Web appears to have increased the quantity of misleading or plainly inaccurate references to what Vannevar Bush actually said in the 1945 article. (Table 2 provides a few such examples from Web searches.)

Contrary to suggestions made in some articles and on some Web sites, the article was not about a world brain, it was not about libraries (except for personal ones), and it did not foresee networks of Memex devices, let alone the Internet. The Memex was not the Rapid Selector, the Memex section did not describe a computer, and the article did not foresee the digital revolution. It also did not envision handwriting recognition, it did not envision machine searching of the information itself, and it was not a multi-media machine (everything was on microfilm only). It described only a limited notion of what we call hypertext (i.e., just linked lists), and it was only marginally related to collaborative research (i.e., sharing a trail with a colleague is along the same lines as sharing a digital folder).

The article could have gone into some of the above areas, but did not. For example, a good deal of space (18 percent of the article) is devoted to analog computers used for repetitive thought processes ranging from calculating machines to formal logic processors (Bush had, after all, gained fame as the creator of the Differential Analyzer at MIT). But these analog machines were described only as a means of creating information. There was no connection in the article between what these machines might be capable of and the task of selecting desired information from the stored collection.

The article also explicitly stated that it was not about libraries. Bush acknowledges in the article that some of the same ideas could be applied to libraries (and by extension to organizational collections of knowledge), but he says, “that is another story.” And he meant it. Perhaps he did not blame libraries, but he definitely was

disappointed by their slowness to adopt technology, even if it was due to lack of funds. The MIT Rapid Selector was initially intended for libraries but no library was willing to pay for one. Bush's real feelings about libraries may be in his comments in "The Inscrutable Thirties," published originally in 1933 and again in a slightly revised form in 1946 (reprinted in Nyce & Kahn, 1991b). In that essay, Bush describes a researcher in a library who has to "paw over cards, thumb pages, and delve by the hour. It was time-wasting and exasperating indeed," especially so given the idea that with improved microfilm thousands of volumes could be housed in a couple of cubic feet, with any page almost instantly selectable. He then went on to grumble about some people being slow to accept a basically sound idea.

In the case of electromagnetic recording, Bush was well aware of Valdemar Poulsen's device in 1898 for electromagnetically storing speech onto steel wires, and he mentioned in the Memex article the possibility of similarly storing data as "magnetic dots on a steel sheet." Nevertheless, he believed microfilm was better because it could be projected and it could be transmitted (between the researcher and the desk) using a television-like process. He apparently did not see any way to use that same television-like process to display and transmit information stored as magnetic dots.

The article also describes using radio so that a researcher in the field could communicate with the recording device in the office. However, as in the case of the brief mention of television transmission, the article does not see any use for a wireless connection beyond a point-to-point connection so that the researcher is not necessarily

right at the desk. As the 1945 article explains, the Memex “can presumably be operated from a distance, [but] it is primarily the piece of furniture at which he works.”

Enduring and Endearing Concepts

Without resorting to attributing unwarranted insights to the Memex article, there are still concepts from the article that fire the imagination and can still be held up as goals.

One of the most important aspects, clearly intended by Bush, was that the Memex was a personal machine storing an individual’s personal collection. If there are any doubts about the fact that it was a personal machine, Bush (1967) explicitly stated later that the 1945 article “proposed a machine for personal use.”

Following from the personal nature of the device is the fact that it was intended to be an extension of one’s own memory. One of the major benefits of the Memex was that a person would no longer have to rely solely on memory to find things previously seen or read. This means that the Memex was to be a personal archive, housing a collection of all of a person’s books, records, communications (i.e., written communications), newspapers, periodicals subscribed to, and so on, accumulated over the years.

As a personal, archival collection, the Memex acquired most things, and stored everything, in a single format (microfilm or microfiche) and it was intended to be a

format that would be just as useful at the end of a lifetime of collecting information as at the beginning. In our current world of software, computers and storage media that are continually being ‘improved,’ we may indeed wonder if we will ever have a personal digital device that will be just as useful thirty or forty years from now as it is today.

Another major aspect of the Memex is that it was supposed to be easy to find associated information as well as to create associations. This is still a good goal, but, like the use of microfilm, the techniques suggested in 1945 can be bettered by newer technologies. After all, the 1945 article had no idea that it would be possible to rapidly search for any word, or any string of characters, occurring anywhere within any text in the collection, and to use occurrence statistics to establish associations. Lesk (2005, p. 225) points out that even though the hyperlink was Bush’s primary information seeking method, “chasing hyperlinks is not effective now” as a way of generally finding information in large collections. The reference is to searching the Web, but the same applies for any really large collection.

A related concept in the 1945 article was that information in one’s own collection, including the linked associations, would be easily shareable with others by virtue of quickly copying microfilm frames. Because the links from one item to another were embedded in the microfilm frames containing the items, sharing links that have been added to formal materials like books and articles would involve copying the materials themselves. Bush, though, did not address the question of copyright or other ownership issues.

Regarding the mechanics of the Memex, there are some suggestions that are still good design goals. One, of course, was that the machine be easy to use. Secondly, the process of reading, and skimming, should be just as easy as handling a real book, including the ability to flip through pages with “a recognizing glance.” For adding material to the Memex that was not acquired in microfilm format, there were four possibilities: typing on the keyboard; placing materials face down on a glass for photographing; writing with a stylus that would cause the lines to be copied to microfilm; and perhaps speaking to the machine so that text would be generated (possibly directly onto the microfilm).

Finally, the article was based on the idea of shrinking massive quantities of recorded information into an incredibly small space. Some writers believe we have already passed Bush’s goal, but it depends on how you look at it. The Memex described in 1945 could accumulate 5000 pages a day (or a little under 2 million pages per year) for “hundreds of years.” After 200 to 300 years, we would have 400 to 600 million pages. With digital technology, the next step is to assume how many bytes per page, and that then depends upon the content and format of the document and any compression used. Just to take one example, a certain book available in PDF format via Amazon.com is 455 pages and about 6 Megabytes, or a little over 13 kilobytes per page. That may not be a representative figure, but at that rate, the personal Memex needs a minimum 5,200 gigabytes of storage to be good for 200 years. This is doable today, but it is not exactly a configuration marketed for personal or desktop systems.

Today's Manifestations

Taking some liberties with the Memex description, there are claims that the Memex's goals have been realized in a combination of the personal computer, the Internet and the Web, and hyperlinks. These claims are primarily looking at the ability to quickly access a huge collection of information, having the means to associate items non-hierarchically, and being able to publish or share materials with these embedded links.

Not the Internet

Any claim that the Internet/Web is an example of the Memex concept is, I believe, misguided. The problem is that the Memex was so clearly a personal collection, and intended to be a supplement to one's own memory. The Internet and the Web provide access to enormous collections, but that is not the same thing as having all material you have seen before on hand under your own control.

For one thing, material on the Web is by no means permanent. A Web page that I accessed last year may no longer be there. Secondly, even if an item still is there, I may no longer have access rights. I do not mean material that belongs to a former employer, for example, but material that I have a right to read and copy as a member of an organization, for instance. When I am no longer a member of that organization, I may be

required to pay again to access something I previously paid for to access and read, or I might not even be able to pay for access without being a member.

There are other differences too. Searching things I have collected myself is different than searching the vast Web, and both have their place. One author (Johnson, 2005) who has collected on his PC an archive of all his writings and notes plus a few thousand selections from books he has read over the past ten years is convinced that there are meaningful differences. He argues that there is “a fundamental difference between searching a universe of documents created by strangers and searching your own personal library.” Influenced by the Memex article, he adds that when you are “freewheeling through ideas that you yourself have collated” it is “uncannily like freewheeling through the corridors of your own memory.”

And of course there is the matter of the physical connection. Despite all the advances in networking, wireless and otherwise, there will always be times and places in the foreseeable future where a person will simply not be able to establish a high-speed connection to the Internet, rendering the Memex as part of one’s own memory temporarily amnesiac. Further, for some applications, even when a high-speed connection at today’s rates is indeed available, it is not fast enough.

Desktop Collections

The desktop, or personal computer, is closer to the Memex in that we have collections of digital and digitized material in a machine (or set of machines) that fit on a desktop. But it is not yet the unified type of environment portrayed in the Memex article. Most of the material is not coming into the system in a single format (although a lot of it might be acquired via the Web), and finding items on a personal computer with a lot of storage capacity has until recently often been slower than finding things on the Internet.

Some researchers in this area are concentrating on the Memex's inclusion of all of one's books, records and communications, and have been working on ways not only to accumulate everything but also to be able to find items amidst all that.

One of the most ambitious projects, for example, is Gordon Bell's MyLifeBits, where he and a small team are attempting to store everything from Gordon Bell's life, including books, articles, Web pages, photos, email, written correspondence, telephone conversations, radio and television, and just about anything else he looks at or hears (Bell, et al., 2004; MyLifeBits Project, 2005). The procedures for accumulating the information even include an automatic camera worn like a pendant on a necklace (or pinned like a brooch to one's shirt or coat). Aside from the challenges in accumulating information from these varied inputs, the project is facing challenges in organizing, browsing, optionally annotating, and retrieving items from such a volume of disparate types, and especially for the non-traditional formats such as audio and video. There are also issues with extracting or creating metadata, and with ownership rights, particularly when information is accumulated as part of one's work for an employer. In terms of

creating links among items, the project demonstrates anew that there are considerably better methods than the 1945 Memex proposal for linked lists. For example, it is much easier to use saved queries to find related items (when all items are maintained using a database structure) than to manually establish the links.

Another project, perhaps a little less all-inclusive, is the Stuff I've Seen system (Dumais, et al., 2003; Stuff I've Seen, 2005). This system for personal computers provides a unified index covering a wide variety of the information that might be stored there, including document files, emails, calendar appointments, Web pages and media files other than documents. One of the more interesting findings from the research is that chronology (the date or time something was initially seen) provides a strong associative link when searching for something seen before. The research also showed that further development is needed in the area of automatic background full-text indexing, so that the user is completely unaware of the activity. Although retrieval is based on the power of full-text indexing, the system also contains the foundation for users to tag items with metadata, thus also providing something akin to the Memex's idea of the user specifically linking items into groups.

There are, certainly, other projects along the same lines, and in the wider area of personal information management. For a fairly extensive list of such projects, with a concentration on understanding and improving the human-computer interface, see Boardman (2003, 2004). In addition, there are a number of products now available in the category of desktop search engines that vary in their capabilities and in the file formats

that they can process for full-text indexing. These projects and products as a group clearly demonstrate several points: (a) desktop collections of information have gotten a lot more complex than what was described in 1945; and (b) the tools needed to deal with finding information within this extended memory are similarly a lot more complex. Furthermore, these projects and products show that, in a unified environment, there can be better ways to find things than looking through digital folders (based on the associative technique of placing related files or file pointers in folders) or using separate application-specific search tools.

The iPod Model

Stepping aside from the desktop storage and retrieval environment, there is one product category today that has a lot in common with the basic elements of the Memex, if we restrict ourselves to a subset of current file types. This product category is the media player, or specifically the Apple iPod with iTunes.

If we ignore for the moment the fact that we are not dealing with documents here, we can examine the similarities:

- (a) the iPod is extremely easy to use;
- (b) the iPod can accommodate a very large amount of data in a very small space (currently, iPod versions are available with up to 60 gigabytes);

- (c) the iPod allows the owner to build very personal collections (see, for example, Schwartz [2004] for a description of what an iPod can reveal about the owner);
- (d) although everything is not in a single format, the 60 gigabyte iPod seamlessly accommodates a variety of formats (for music, photos, audiobooks, voice mail messages);
- (e) the owner can create playlists that are similar to the Memex's linked lists of associated items, with any item able to appear in multiple lists;
- (f) the owner can share playlists with others.

One of the great things about the iPod is that it is so simple to use, and it would be wonderful to have a small, handy document-oriented information pod that would house all the materials one would like to collect as part of a personal memory.

Conclusion: Memex as iPod/Tablet Combination

Looking back at the 1945 Memex article, I believe that one of the most important parts of the description is that it was a personal device. It was a compact device (relative to then-current technology) that stored, and provided access to, the things a person would like to keep as part of a personal or professional memory, with a means for associating certain items together when desired.

And just as we now ignore the fact that the 1945 article talked about a desk-sized contraption that was not a digital machine, I think we can also ignore the fact that it did not envision databases, full-text searching, or other desirable retrieval techniques. In addition, we are beginning to ignore the reliance on manually created associations among items as the primary retrieval mechanism (e.g., digital file folders). With a unified massive memory device, we could let the machine produce the bulk of the associations. Rather than manually putting files or file shortcuts into digital folders, an automated process could add grouping information based on file types, acquisition method, conceptual associations and the like at acquisition time, or the machine's retrieval methods could display grouped items based upon various associative techniques. In addition, we could still have the option of manually adding drag-and-drop labels or classification tags, or creating "playlists," when desired. Internally, instead of traditional file structures, the basic data structures could be optimized for this type of storage and retrieval. (The development of newer internal storage structures for enhanced retrieval of item contents and concepts is certainly underway, but beyond the scope of this paper.)

The core concept of a handy personal memory extension, with easy procedures for putting material into it, is still a fascinating one. The problem today is not so much that the individual capabilities of a Memex-like device are not available, because they all are in one form or another, but that they are not available in a single device in an environment that easily supports such devices. For example, paper documents that I wish to save digitally can be placed on a scanner and thus added to a digital collection, but I

am not aware of a unified compact device that incorporates all of the functions of scanner, processor, storage and display.

The ideal device might be a combination of the iPod design with the tablet computer format. Like the iPod, its primary purpose would be to house the collection and provide access to it. Just as the 1945 article suggested, most material should be available for acquisition in a suitable format. Today, that would mean either a direct or indirect connection to the Internet for downloading a common format such as HTML, XML, PDF (or PDF Archival), word processor formats, and so on. For material on paper, just as in the 1945 article, it should be possible to simply slap a document face down on the tablet and have it scanned in. And just as the 1945 article described, it should be possible to use a stylus to write and draw directly on the tablet screen for annotation purposes. In addition to all of that, though, we would want all text to be automatically recognized and fully indexed, and even non-text recognized and indexed to the extent possible. Finally, retrieving information stored on the device should be iPod-easy.

I do not propose that this device be a substitute for desktop computers, and it need not support all the software for creating documents and other forms of materials, nor does it need to do all the other things personal computers can do. On the other hand, it is not just a storage unit, because it is a self-contained device with everything necessary to find, view and exchange materials. It is essentially just what Bush described: a mechanized private library for the materials that I deliberately want to keep over the years in case I

want to refer back to something at a much later time. It is a replacement for my bookshelves, not for my personal computer or laptop.

Regarding the issue of storage capacity, we might be chasing a moving target, in that our ability to expand recorded information seems to be always ahead of our capabilities for compact storage and rapid retrieval. For example, what about storing all the movies one has seen? According to some estimates, a 60-gigabyte video iPod, using the H.264 video format, would hold just 30 two-hour movies of high definition. By the time an iPod-like device can handle all the television, movies and music videos one has seen and wants to save, along with all the audio, digital photographs, Web pages, emails and print-based materials, no doubt there will be even more things to push the storage limits. At the same time though, there does not seem to be any real worry that the technologies to store ever more massive amounts of information in very small containers will not similarly progress.

The non-technical hurdles to this iPod/Tablet type of Memex, as a personal memory of things read and seen, may be greater than the technical ones, particularly in the areas of formats and intellectual property rights. For example, I can buy books in electronic format (e.g., PDF) and store them on my own machine, but I may not be able to print them or to share them. If I borrow a book from a library and read it, my ability to save a copy is legally limited. There are many things I read, and would like to save, that are simply not available today for purchase in a digital format. Even when a book or article is available for purchase in digital format, I may be required to pay for it twice if I

want to read it in paper form and yet save it in digital form for later searching. (Reading lengthy texts on screens is still not as comfortable for many people as reading paper, while searching is of course much easier on the computer.)

On the other hand, the iPod is also an example of one way of resolving the issues of intellectual property rights, payment mechanisms, and format accommodation (or control). Given the apparent success of the iPod/iTunes arrangement, and similar online music distributors, some commentators have suggested that such arrangements will become the predominant way recorded music is sold to consumers. Would that kind of arrangement be appropriate for the majority of printed material?

Vannevar Bush acknowledged in his 1945 article that “means as yet unknown ... may come any day” that could change things just as dramatically as the vacuum tube did. On that point, he was absolutely right. He also still seems to be right in his assumption of the desirability of a supplement to one’s own memory. We do not have to make exaggerated claims for visions that were not in the article to find something that is still a goal today. It would be nice to have a personal device – an iPod-like information pod – that we can use to easily collect all the things we have read and seen that we would like to save, with an easy way to quickly find and view things within that collection.

Acknowledgments

I would like to express my appreciation to the anonymous referees for taking the time to offer helpful suggestions and references to additional material.

References

Bell, G., Gemmell, J., and Lueder, R. (2004). Challenges In Using Lifetime Personal Information Stores. In SIGIR '04, Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval (Sheffield, UK). New York: ACM Press.

Boardman, Richard. (2003). Who's Working on Personal Information Management? Retrieved July 20, 2005 from <http://www.iis.ee.ic.ac.uk/~rick/pim.htm>

Boardman, Richard. (2004). Improving Tool Support for Personal Information Management. Retrieved July 20, 2005 from <http://www.iis.ee.ic.ac.uk/~rick/thesis/boardman04-thesis.pdf>

Buckland, Michael K. (1992). Emanuel Goldberg, Electronic Document Retrieval, and Vannevar Bush's Memex. *Journal of the American Society for Information Science* 43 (4), 284-294. Retrieved May 3, 2005 from <http://www.sims.berkeley.edu/~buckland/goldbush.html>

- Burke, Colin. (1991). A Practical View of the Memex: The Career of the Rapid Selector. In J. M. Nyce and P. Kahn (Eds.), *From Memex to Hypertext: Vannevar Bush and the Mind's Machine* (pp. 145-164). San Diego: Academic Press.
- Bush, Vannevar. (1945a). As We May Think. *Atlantic Monthly* 176(1), (July 1945), 101-108.
- Bush, Vannevar. (1945b). As We May Think. Reprinted in J. M. Nyce and P. Kahn (Eds.), *From Memex to Hypertext: Vannevar Bush and the Mind's Machine* (pp. 85-110). San Diego: Academic Press.
- Bush, Vannevar. (1967). Memex Revisted. Reprinted in J. M. Nyce and P. Kahn (Eds.), *From Memex to Hypertext: Vannevar Bush and the Mind's Machine* (pp. 197-216). San Diego: Academic Press.
- Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Toronto, Canada). New York: ACM Press.

Johnson, Steven. (2005). Tool For Thought. New York Times Book Review (January 30, 2005), 27.

Kahn, Paul, & Nyce, James M. (1991). The Idea of a Machine: The Later Memex Essays. In J. M. Nyce and P. Kahn (Eds.), From Memex to Hypertext: Vannevar Bush and the Mind's Machine (pp. 113-144). San Diego: Academic Press.

Lesk, Michael. (2005). Understanding Digital Libraries. San Francisco: Morgan Kaufmann Publishers.

MyLifeBits Project. (2005). Retrieved May 16, 2005 from <http://research.microsoft.com/barc/mediapresence/MyLifeBits.aspx>

Nyce, James M., & Kahn, Paul. (1991a). A Machine for the Mind: Vannevar Bush's Memex. In J. M. Nyce and P. Kahn (Eds.), From Memex to Hypertext: Vannevar Bush and the Mind's Machine (pp. 39-66). San Diego: Academic Press.

Nyce, James M., & Kahn, Paul. (Eds.) (1991b). From Memex to Hypertext: Vannevar Bush and the Mind's Machine. San Diego: Academic Press.

Rayward, W. Boyd. (1994). Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext. *Journal of the American Society for Information Science* 45, 235-250. Retrieved May 3, 2005 from <http://www.lis.uiuc.edu/~wrayward/otlet/xanadu.htm>

Rayward, W. Boyd. (1997). The Origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). *Journal of the American Society for Information Science* 48, 289-300. Retrieved May 12, 2005 from <http://www.lis.uiuc.edu/~wrayward/otlet/OriginsofInfoSci.htm>

Schwartz, John. (2004). To Know Me, Know My iPod. *New York Times, Week In Review* (Section 4, November 28, 2004), WK 6.

Smith, Linda C. Memex as an Image of Potentiality. (1991). In J. M. Nyce and P. Kahn (Eds.), *From Memex to Hypertext: Vannevar Bush and the Mind's Machine* (pp. 261-286). San Diego: Academic Press.

Stuff I've Seen – Home Page. (2005). Retrieved May 17, 2005 from <http://research.microsoft.com/adapt/sis/index.htm>

Zachary, G. Pascal. (1997). *Endless Frontier: Vannevar Bush, Engineer of the American Century*. New York: Free Press.

For Peer Review

Table 1. Basic Concepts in the 1945 Memex Article

Section Subsection	Concept/Statement		
Introduction		14%	
Post-war Job for Scientists			< 3%
	No problem for biologists, medical scientists		
	Physicists need new work		
Role of Science			< 2%
	Better control of material environment		
	Increased knowledge of biological processes		
	Illuminating physiological-psychological relationships		
	Speedier communication among people		
	Greater ability to record ideas and select from that		
Mountain of Research			9%
	Amount of information staggering		
	Problem of specialization		
	Methods for sharing, reviewing are old and inadequate		
	Inability to remember everything read		
	Significant results can be lost in the mass		
	New technologies may help (photocells, advanced photography, vacuum tubes, relays, CRTs)		
	Now possible to manufacture cheap, reliable devices		
Creating the Record		43%	
Photography			14%
	Minicamera (walnut-size) w/ 100 frames per roll		
	Stereoscopic camera		
	Wearing camera on head, with shutter wire in sleeve		
	Dry photography		
	Instant photography		
	Fax as type of photography		
	Photocell snapshot of scenes, even distant scenes		
	Television camera to produce photographs		
Smaller microfilm			5%
	Microfilm, 100 times smaller than original page (a million books reduced to one end of a desk, or all recorded knowledge – a billion books – in a van)		
Voice input, wireless			6%
	Dictating (speech to typescript)		
	Wireless link to office when dictating, photographing		
Machines for Repetitive Thought			18%
	Machines for repetitive “thought” (adding, etc.)		
	Advanced arithmetical machines (100 times faster)		
	Machines for advanced analysis (but no big market)		
	Logic machines to crank out conclusions from a set of premises (new symbolic language needed)		
Selecting From the Record		34%	
Techniques, Fast Selection			12%
	With photocells and microfilm, search 1000s per sec.		
	Instead of sequential searching, use the same idea as in banks of telephone relays, with vacuum tube switches		
	A microfilm “card” could be selected in a second		
	The “card” could be magnetic dots on steel sheet, but		

	not as many advantages as microfilm		
	Voice commands for selecting		
Indexes and Memory			4%
	Current index methods artificial, cumbersome		
	Human mind operates by association		
	But human memory is transitory		
	Selection by association can be mechanized		
	Mechanized memory permanent and clear		
Memex			7%
	Individual's storage of books, records, communications		
	Desk with multiple screens, keyboard, buttons, levers		
	Lever used to scan through pages, either direction		
	Could be operated from a distance (wirelessly?)		
	Improved microfilm (could add 5000 pages per day for hundreds of years)		
	Most content purchased already on microfilm		
	Glass plate for photocopying other docs onto microfilm		
	Stylus for handwritten notes added to microfilm		
Trails			8%
	Normal index, with the index also on the microfilm		
	Associative index, any two items can be linked, with the index number of each embedded in the other		
	Numerous items in a chain form a trail, and any item can be in numerous trails, and a chain of items can be viewed just like the pages of a book		
	Trails can be reproduced and shared with friends		
New Publication Types			3%
	Encyclopedias with ready-made associative trails		
	Professions with their own "associated" information		
	Trail blazers who create the trails		
Really Futuristic		7%	7%
	Electrically recording what the eye is looking at		
	Doing manual tasks by thought control		
	Hearing without using the ear		
	Generally, electrical waves outside the body interacting directly with electrical impulses inside the body		
Conclusion		<3%	<3%
	Machines to find what we forget		
	Recorded knowledge aids humanity's growth		

Table 2. Sample Misleading or Inaccurate Statements about the 1945 Memex Article.

<p>A precursor concept to the modern personal computer ... also called the Rapid Selector</p> <p>(http://www.homoexcelsior.com/omega.db/datum/computer_science/memex/8683)</p>
<p>Memex was, basically, a complicated analogue computer</p> <p>(http://www.aedo-to.com/eng/inspiration/future/realizzati/01/art01.html)</p>
<p>[the Memex article] hinted at the potential of using television to provide network links</p> <p>(http://www.labyrinth.net.au/~saul/essays/02computer.html)</p>
<p>Unmistakably in this [the Memex article] is the technology now known as search by millions and known as <i>information retrieval</i> (IR) by tens of thousands.</p> <p>(http://www.acmqueue.com/modules.php?name=Content&pa=showpage&pid=148&page=1)</p>
<p><i>memex</i>: Vannevar Bush's conception of the computer as universally accessible electronic library</p> <p>(http://academ.hvcc.edu/~kantopet/misc/index.php?page=net+beginnings&parent=net+history&printme=true)</p>
<p>Dr. Bush foresaw a computer designed for the purposes of applications other than number manipulation. He called it the Memex.</p> <p>(http://www.thocp.net/hardware/xerox_star.htm)</p>
<p>[the Memex article] proposed linking all the world's libraries with all the world's knowledge via a system he called Memex</p> <p>(http://www.flora.org/oldfolks/ch2.html)</p>
<p>[the Memex article is] the first visionary articulation of the digital age ... Central to his vision of the memex was the notion that sound would circulate through the system, available for easy retrieval and manipulation</p> <p>(http://www.nothing.org/osc/FreezeFrame.htm)</p>
<p>Memex machine, a hybrid microfilm computer developed in the forties</p> <p>(http://www-rohan.sdsu.edu/dept/drwswebb/lore/1_3/tropeano.htm)</p>
<p>[the Memex article describes] a desk filled with microfilm, shrunken photographs, tiny movies, and miniature audio tapes</p> <p>(http://www.kenjordan.tv/Whats%20That%20Sound.doc)</p>
<p>[the Memex article describes] his vision of a distributed MEMEX in which he outlined the possibility of building digital network computers that incorporate multimedia functions and use of Compact Disc (CD)</p> <p>(http://iet.open.ac.uk/pp/d.mwanza/phd-thesis/Chapt2.pdf)</p>