

HFS, We Have a Problem:
Making Bush's Memex into a Reality & Leaving the HFS Behind

The internet, as it is today is a vast network of users, computers, bots, and websites all of which are conveniently accessible via a computer terminal anywhere in the world. Information is stored on servers around the world, and is arranged in the typical hierarchical file system (HFS). These files are linked together through folders and hyperlinks from other documents. Navigating through many folders and links can become cumbersome and repetitive, as well as confusing and difficult to remember any path to the single copy of a file which the user is seeking. The use of this HFS causes information and paths to information to be forgotten, which causes frustration for the user. In the 1990's, many internet browsers and internet-based companies began attempts to solve this problem with the use of bookmarks, meta-tags, and web crawlers, only to find that the problem still exists. Although these attempts have taken the appropriate steps to organize information in a more logical method, the main problem is not creating a system to retrace steps, but creating a system that does this organization and re-tracing automatically, and has the ability to grow with time, and with users.

In 1945, Vannevar Bush published his world-famous essay "As We May Think" in *Atlantic Monthly*; this paper will use the 1996 reprint of the essay. Bush described a tool 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" (Bush, p. #43). This tool, which he called the Memex, would bring an entire library into a piece of furniture that would be about the size of an ordinary desk. Users would create "trails" of information based on association,

much like the way the human mind links information, but unlike the human mind, these links would not fade with time.

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. At the bottom of each there are a number of blank code spaces and a pointer is set to indicate one of these on each item. The user taps a single key and the items are permanently joined (Bush, p. #44).

Using this type of system, not only is the information that the user is looking right there, but the trail which led that user to that information, as well as other articles related to the same bit of information; thus, creating a web of information based on the users own links.

Bush wrote this essay in response to the growing amount of scientific information being produced and the inaccessibility of information for research. “Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose” (Bush, p. #37). Bush was concerned with, not only the storage of information, but the ability to consult it once stored. Information should be retrieved as easily as it was produced and stored. Today, we can recall information, as stated above, by traversing through a redundant and confusing HFS in which it is easy to become lost, and lose information which was found in previous attempts. For example, if one lands on a website and is looking for information on guitars, but the website is about all instruments, the user must go through many clicks, different levels in the HFS, to find that page. When the user finally reaches the page, the entire process of clicking must be remembered, or re-traced in reverse order; this causes confusion, as well as loss of information. We are once again facing the dilemma that Bush was trying to fix, and what needs to be done is a complete restructure of how we store, and retrieve information in the digital world.

Attempts in making the Memex a reality

In years, recent and not so recent, attempts have been made to convert the known HFS to an associative filing system (AFS), more natural to the human mind, which would make information much simpler to find.

In the 1970s, an American philosopher and sociologist named Ted Nelson wrote of a software framework known as Xanadu. Xanadu was, in a sense, a writing system which stored multiple versions of a document, in what Nelson calls the *docuverse*, allowing users to view all changes in the documents, and the paths which the document went through to get to its current state. Although the project was never completed, and has been considered a failure by critics, Nelson took the first step and proposed of a problem with the technology at the time, the HFS problem. In 2000, a group of computer scientists proposed the idea of a Memex once again, this time in the form of a computer application which was also known as “Memex”.

The software memex “provides a platform for capturing hypertext and meta-data based on content as well as surfing events, in a structured manner” (Chakrabarti). The system used technologies such as bookmarks, history, meta-tagging and learning folder structures. Although this application relies on the current HFS models, it proves that the technology is available to make Bush’s Memex a reality.

A small group of researchers from Microsoft have also written another proposal to make the Memex into a reality. In 2002, the *MyLifeBits* project proposal was written; this proposal described in detail how the use of current technology, and the growing affordability of storage space, can allow users to store all information they come across.

It is a system for storing all of one’s digital media, including documents, images, sounds, and videos. It is built on four principles: (1) collections and search must replace hierarchy for organization (2) many visualizations should be supported (3) annotations are critical to non-text media and must be made easy, and (4) authoring should be via transclusion (Gemmell, p. #235).

In the fourth point above we are introduced to a new term, “transclusion”. What transclusion does is store data in one place, rather than copying it into multiple places, and allows it to be viewed in different contexts. The reference to the information also creates a link to each article. If one finds a website about a specific guitar, on a large guitar website, the user may link that piece of information with a picture of that guitar that is found on another website. Only that one guitar is viewed in the linking, but the documents are linked. Filing systems should support this idea. The *MyLifeBits* project is based solely on Bush’s Memex stating that, “It is Memex, extended beyond Bush’s vision to handle audio and video” (Gemmell, p. #235). The increasing inexpensiveness of data storage means will allow for unlimited storage, and allow users to store everything they view. The *MyLifeBits* project is currently in alpha testing at the Microsoft Research department, and many years from release.

Currently, the use of weblogs (blogs), bookmarks, Google, web-crawlers, and meta-tagging are all based on the Memex structure of information linking and gathering. “Google’s origins are strikingly similar to the memex. Google initially began as a tool for rating annotations, according to Larry Page (inventor of Google’s eponymous PageRank)” (Jerz). With Google, page annotations and rankings (amount of visits/links) are now possible.

Each example described above shows that there is a great need for change in the current system of data storage and retrieval. The problem was realized in 1945 by Bush, and has yet to be resolved, even with many attempts. Each example has used different methods in order to solve the HFS problem, but have only gone so far in their attempts; however, each example still relies on current HFS technology, this is the real problem. A

greater change needs to take place, not just a program that fits into current media and technology, but a new, completely different technology. What is really needed is a new system of storage, not just a revamp of an old one.

Where do we need to go next?

Today, we have the internet: a network of insecurity, old technology, and the inability to personally add and annotate information as Bush would want us to. From the examples given above, one can see that the technology is available to make the Memex a reality, and these examples all share a common theme: they all attempt to change how we retrieve information, because our current system of information retrieval/storage is lacking many features, and seems to be incompatible with modern standards. Users do not have the ability to annotate, add, edit, delete, link, and copy information as they please, and the option of re-tracing their steps to figure out how they got to where they are is almost impossible. A user can copy a file to a more suitable place, but that is not the same bit of information, it is a new copy. If a user wants to annotate a file, the file must itself be edited, which may cause problems for the user.

Using many current technologies, this difficult task to create a fully user-based, annotatable, and associative network can become a reality. Using audio/visual technology, video game technology (or the ability to create what is also known as a simulacra), and engineering, a new media can be created that allows users to do all of these various features, and even more. Current technology allows for greater storage of information, links, annotations, and alterations. As stated in the proposal for the *MyLifeBits* project, “such abundant storage is finally on the horizon. Within five years [from 2002], terabyte hard drives will be common and inexpensive... Thus, purchasing

an additional terabyte of personal storage every year will be feasible for the average computer user” (Gemmell, p. #235). This vast amount of storage is only the first step taken to allow the Memex idea to become a reality.

The system should work on more than one plane of space, because information is not flat. Information can be found anywhere, and in any medium, current technology usually limits users to a single medium. Since information is stored in more of a three-dimensional world, information should be viewed, annotated, edited, and stored in a three-dimensional world. This concept may be difficult for many to envision; however, our current flat filing systems can only grow so far, and do not allow users to link multiple files easily. A three-dimensional system will enable many files to be linked to many other files, and allow all trails to be viewed simultaneously; thus, speeding up information retrieval and production. For example: if a user links 17 files to one file on guitars, that user can view a three-dimensional map of those links, and the links from other files to the linked files. If a user finds a link to be unnecessary, or finds another more suitable path, the user can chose to make the new link, bypassing all links that are not used in the process, and avoiding the problem of HFS redundancy. These links will be like the links in Bush’s Memex. The links are set by the user, and permanently linked unless the user decides to remove the link, the user should also have the ability to view multiple linked files on a single screen, or even multiple screens like that of the Memex.

Users should have the ability, just as in the *MyLifeBits* project, to store links to pictures, documents, videos, audio, as well as any other digital mediums. Information is not limited to print, and microfilm, thus the technology should allow for additions of new

mediums, as well as expansion of current mediums. Users should be allowed to interact on a much more direct level than how they interact today.

Vannevar Bush realized the problem, in 1945, with information retrieval and collection:

There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial.

Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose. (Bush, p. #37)

Still, today we have this same issue. There is an unlimited amount of information to be found and linked together; the time it would take to remember and link all of this information would be staggering for a single individual. The Memex would allow all users to link information together, thus allowing much less time to be spent on drawing these links from scratch, and much more time to be spent discovering new information. The Memex was not only a tool to save space, but a tool to save incredible amounts of time. The current HFS that is used in computer technology today creates unnecessary redundancy and confusion, and wastes valuable time.

Although Bush's machine was an analog computer system from the 1930s [his first draft of "As We May Think" was written in 1937 (Nyce, p. #42)] the ideas in which he built the Memex on are still prevalent in today's technology. Although attempts have been made, in 60 years we have not solved the problems of the HFS. The adoption of an AFS would allow information growth to flourish, and would allow other technologies to be used to replace the current, aging technology we use today. The use of associative links to find and store information is possibly the fastest method of information storage;

the human brain stores all information using association, this was Bush's inspiration for the Memex. "Memex got its name because it was to support and extend the powers of human memory and association" (Nyce, p. #57). Links can be seen as neurons in the human mind, linking various functions and information together which can be retrieved in an instant. Let us see a modern HFS retrieve information that quickly, and link it to other objects just as quickly.

Works Cited

Bush, Vannevar. (Mar. 1996). *As We May Think. Interactions. Vol. 3, Issue 2, 35-46.*

This is a re-print of Vannevar Bush's original article which was first printed in 1945 in *The Atlantic Monthly*. This is the introduction of Bush's Memex idea, and the basis for my entire presentation.

Chakrabarti, Soumen; Srivastava, Sandeep; Subramanyam, Mallela; Tiwari, Mitul. (Jun. 2000). *Using Memex to archive and mine community Web browsing experience. Computer Networks. Vol 33, Issues 1-6, 669-684.*

This is an article found on the Science Direct website. It's a plan for a computer web browsing assistant that is called *The Memex* and tracks users browsing. It also discusses more primitive forms of a "memex" such as bookmarks, history, and search engines. Many aspects are written about: architecture, server design, and other data models.

Gemmell, Jim; Bell, Gordon; Lueder, Roger; Drucker, Steven; Wong, Curtis. (2002). *MyLifeBits: fulfilling the Memex vision. Proceedings of the tenth ACM international conference on Multimedia. 235-238.*

The *MyLifeBits* project is a project by Microsoft researchers to make Bush's Memex idea a reality. To create a system that is capable of storing ones complete digital life.

Jerz, Dennis G. (2003). *On the Trail of the Memex: Vannevar Bush, Weblogs and the Google Galaxy.*

This is an article which discusses how Bush's memex and the idea of hypermedia have aided in the creation of weblogs and Google.

Nyce, James M.; Kahn, Paul. (1991). *A Machine for the Mind: Vannevar Bush's Memex.*

Paul Kahn is a "Memex Historian"; I found this book by searching for his name after I found info on a symposium he spoke at. The book is a reprint of Bush's writings and related studies and sources, as well as comments on Bush's writing.