The Encyclopedia of Applied Linguistics

Bush, Vanevar

Journal:	*The Encyclopedia of Applied Linguistics*
Manuscript ID:	Draft
Wiley - Manuscript type:	article
Date Submitted by the Author:	n/a
Complete List of Authors:	Cobb, Thomas
Keywords:	language learning technology, reading, literacy

Vannevar Bush: Hypertext past, present and future

Widely known as the father or even guru of hypertext and its principle medium the World Wide Web, Vannevar Bush (1890-1974) never employed either of these names. Rather he envisaged a "memex" system for organizing and retrieving multimedia information on microfilm, memex being a word like rolodex or index (card) that is more redolent of the actual period of his activities. Yet Bush's memex system went far beyond any index based information retrieval system and in many ways is the true precursor of the Internet, at least on the concept level.

Vannevar Bush in the 1940's
(from http://en.wikipedia.org/wiki/Vannevar_Bush)

An engineer, technocrat, and pioneer of analogue computing, Bush became part of the mass mobilization of US intellectual resources in WWII. Director of the Office of Scientific Research and Development 1941-1947, he was charged with focusing the minds of 6,000 leading US academics and scientists on matters of defense, via for example the Manhattan Project that developed the atomic bomb. After the War, Bush became presidential science advisor and hoped to carry the war time mobilization into useful peace time projects. The basic details of Bush's work and life are well presented in his Wikipedia entry.

What emerged as interesting for Bush in his war work was the issue of information management. *Information* for Bush was roughly equivalent to *civilization*, with retrieval its principle mechanism. In his words: "[Information retrieval] is a much larger matter than merely the extraction of data for the purposes of scientific research; it involves the entire process by which man profits by his inheritance of acquired knowledge" (1945, p. 106). And yet, at about mid-century effective information retrieval was starting to become a problem. Bush's efforts to mobilize a nation's intellectual resources had made him aware that there was already inexistence more printed and other information available on any given topic than was easily manageable, even as the creation of new information was accelerating. Shortly after the War, in an early and prescient encoding of the sense of being information ridden, Bush described the present state of info-civilization:

> Science has provided the swiftest communication between individuals; it has provided a record of ideas and has enabled man to manipulate and to make extracts from that record so that knowledge evolves and endures throughout the life of a race rather than that of an individual. … 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. (p. 101)

(Lengthy citations will help establish the prescient yet at the same time dated and even quaintly optimistic quality of Bush's thinking as viewed from the present. This citation is from the main document for which Bush is still known, an article written just after the War entitled "As We May Think" and published in The Atlantic Monthly in July, 1945.)

Bush's problem with information was not really that there was getting to be too much of it (he seemed never to question the engineer's maxim that more information is usually better), but that the means available for handling it had not kept pace the means of producing it:

> The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present day interests, but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships. (p. 102)

The square-rigged ship in question was basically the alphanumeric, name-based hierarchical index system, like the names-within-names of the folder system on the hard drive of your computer:

> The real heart of the matter of selection … 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. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path. (p. 107)

The solution to this problem is to organize information more as humans do, or "as we may think," i.e. in associative leaps and links:

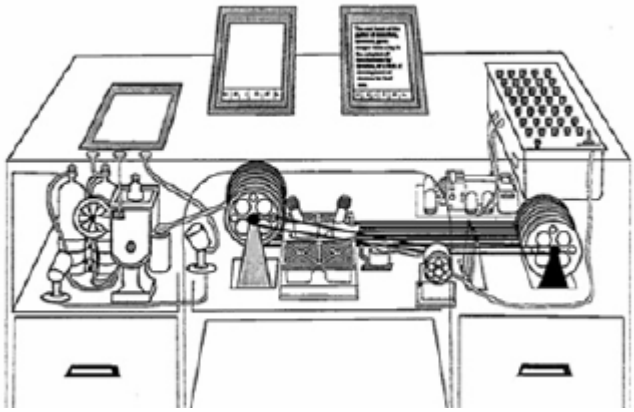
> The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. (p. 107)

This scheme however is as we "may" think since we do not yet have all the details:

> Man cannot hope fully to duplicate this mental process artificially, but he certainly ought to be able to learn from it. In minor ways he may even improve, for his records have relative permanency. The first idea, however, to be drawn from the analogy concerns selection. Selection by association, rather than indexing, may yet be mechanized. One cannot hope thus to equal the speed and flexibility with which the mind follows an associative trail, but it

> should be possible to beat the mind decisively in regard to the permanence and clarity of the items resurrected from storage. (p. 107)

The plan probably seemed idle futurism at the time, but Bush envisaged a real device, which he named a memex (short for memory extender), 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" (p. 107). Physically, the memex would resemble a desk, the top of which would bear "slanting translucent screens on which material can be projected for convenient reading," as queried from a "keyboard, and sets of buttons and levers." Space for all this information is no problem: "The matter of bulk is well taken care of by improved microfilm. […] if the user inserted 5000 pages of material a day it would take him hundreds of years to fill the repository, so he can be profligate and enter material freely" (p. 107). The information to be thus collected would not be limited to a text format:



As Bush's Memex may have looked
(From Thomas Hofmann, Director of Engineering at Google, http://www.dbis.ethz.ch/education/ws0708/inf_retrieval/IR2007-Lecture01.pdf and on Google Images)

> Most of the memex contents are purchased on microfilm ready for insertion. Books of all sorts, pictures, current periodicals, newspapers, are thus obtained and dropped into place. Business correspondence takes the same path. And there is provision for direct entry. On the top of the memex is a transparent platen. On this are placed longhand notes, photographs, memoranda, all sorts of things. When one is in place, the depression of a lever causes it to be photographed onto the next blank space in a section of the memex film, dry photography being employed (p. 7).

As enthusiastic as Bush was for the simple collecting of information that a memex would make possible, it was more the accessibility of particular pieces of information that interested him, and following that their linking into useful associations. For this he envisaged making full use of traditional indexing: "There is, of course, provision for consultation of the record by the usual scheme of indexing. 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." (Like hyperlinks?) But more than index links, it is the associative links that Bush is really interested in:

> All this is conventional, except for the projection forward of present-day mechanisms and gadgetry. It affords an immediate step, however, to 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. (pp. 107-108)

More specifically:

> 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. … Thereafter, at any time, when one of these items is in view, the other can be instantly recalled merely by tapping a button below the corresponding code space. Moreover, when numerous items have been thus joined together to form a trail, they can be reviewed in turn, rapidly or slowly, by deflecting a lever like that used for turning the pages of a book. It is exactly as though the physical items had been gathered together to form a new book. (p. 108)

Minus the buttons, levers, microfilm, and dry photography, and with slanting screens replaced by virtual windows on a single but still slanting screen, functionally Bush's memex pretty much describes the laptop your writer is working on right now. But one big difference is that the latter can access not only all the information on a single hard disk but also that on the many other hard disks and stored in networks that "over a billion people have access to [in] the most extensive expansion of information ever compiled" (Woolf, 2007, p. 221)

Hits and misses

Most of the quarter million hits for "Vannevar Bush" currently itemized on Google, at least on the first few pages, emphasize the amazing accuracy of Bush's predictions, but in fact his track record is mixed. Take the notion of a named information trail, Bush's key idea about how we may think. Technically, this has turned out even better than Bush predicted: the trail of my research for this very article, i.e. the pages I actually went to from my search engine's offering yesterday, is automatically stored in chronological order on my browser and accessible in my Internet Explorer's Favorites under "History: Yesterday" (and similarly in other browsers). Here, I can edit out the false starts, dead links, and ads that a typical search now entails, something Bush almost certainly did not envisage. But more importantly, somehow the idea of a trail of stored associations has not been the hit that Bush predicted. Personally I rarely use my browser's History feature in Web searches, preferring to save good finds as a list of either as links or documents, either way losing the sequential character of a trail or other information about what I may have been thinking. I seem not to be alone: a Google search on "IE favorites history" leads mainly to articles from Microsoft and others about how to use the feature, and there are no studies of how or whether it gets used by information workers (in e.g. the open access database of *Educational Technology Research & Development* at http://www.springerlink.com/content/1042-1629/).

Other key ideas have fared better. The notion of a mechanized associative search began as pure fantasy but step by step became clear and operable. The notion of capturing an association once provided by a search engine is clear enough, but how

the search engine provides the raw material for associations was less so without the replacement of microfilm by digital media and then the many subsequent upgrades of digital processing power. This power was not long coming, and Bush's basic idea of an associative, non-hierarchical search was soon realizable in prototype. An early realization from IBM was Luhn's (1957) proposal to leave the to-be-searched archive itself largely unstructured (like the Web of today) and then retrieve information from it by "using words as indexing units for documents and measuring word overlap as a criterion for retrieval" (p. 316). This is a type of full text search, rather than a folder name and filename (index) search, such as we have enjoyed on our PCs since roughly 2005. The basic idea was developed and tested over the 1970s and 1980s, and particularly since 1992 through the ongoing Text Retrieval Conference series (TREC, sponsored by National Institute of Standards and Technology, which regularly cites Bush's ideas in its proceedings).

The search algorithms now deployed by Google et al. are too complex to discuss in the present forum except to say that there is continuity between the associative principle of Bush and the modern search engine of Google *et al*. Bush is regularly cited as an inspiration by Googleplex fellows including Singhal (2001). Some recent search refinements Bush did not anticipate but that will resonate for applied linguists are that search words are stemmed (fleshed out as families), and multi-word searches are recast as phrase-first (thus *White House* is searched as a whole phrase first, then conjunctively, and only finally disjunctively). But these refinements are all continuous with Bush's basic concept, and indeed Bush is in some ways still ahead of Google in that a Google search remains a word-form dominated keyword search, with all the limitations pertaining thereto, where for instance money *bank* may be sought but river *bank* found. The final realization of Bush's associative idea awaits the fuller implementation of the "semantic Web," apparently just around the corner, where searches will proceed on a conceptual level through XML-encoded documents (Antoniou & Van Harmelen, 2004).

To summarize, Vannevar Bush came up with a remarkably specific blueprint for not only where we have got to in the universe of digital media but also where we may yet be going. It is thus ironic that at the moment of realization the question is increasingly posed, Is this where we want to be going?

What Bush did not predict

We are presently seeing the simultaneous triumph of Bush's ideas along with the emergence of worries that a fully linked docuverse may just be another type of information overload that this time there really is no solution for. A whole semi-popular subgenre of worrisome assessments has recently sprung up to argue points like "The Internet is rotting your brain" (Miller, 2010) and "Google is making us stupid" (Carr, 2008). Nicholas Carr's *The Shallows* (2010) is a non-academic but still serious work expressing a growing unease about the kinds of mental experience that a fully linked docuverse leaves us with. The basic problem is that fully linked documents provide mainly distraction and overload, and far more of it than people can actually cope with, particularly in the case of those who have had little other experience with texts and information such as children.

This piece from *The Shallows* looks like a realization of Vannevar Bush's happy daydream of buttons and levers:

> For more than a decade now, I've been spending a lot of time online, searching and surfing and sometimes adding to the great databases of the Internet. The Web has been a godsend to me as a writer. Research that once required days in the stacks or periodical rooms of libraries can now be done in minutes. A few Google searches, some quick clicks on hyperlinks, and I've got the telltale fact or pithy quote I was after. Even when I'm not working, I'm as likely as not to be foraging in the Web's info-thickets, reading and writing e-mails, scanning headlines and blog posts, watching videos and listening to podcasts, or just tripping from link to link to link.

The difference however is that Carr here is building up to an ultimately negative assessment of this experience, where all this quick, targeted, or scanning type of linked, associative, and highly distractible "reading" is leaving him and presumably others less and less inclined, or possibly able, to get to the end of a longer texts with complex arguments.

Carr's evidence is mainly anecdotal, but he may have a point. Research studies both broad and narrow are starting to look at the costs and benefits of hypertextual, lateral, or search-based reading (to name just a few of the terms competing to name the phenomenon). An broad look is the five-year study (by Rowlands & Nicholas, 2008) of the computer logs from two heavily used research UK Web sites, showing that link-enabled researchers both young and old tend to skim rather than read, hop from one source to another, rarely return to any source they have already read, and rarely read more than two pages of any document, thus not so much reading as "power browsing horizontally" through abstracts and content pages in search of quick references for school assignments and academic tasks.

From the perspective of process and detail, reading researchers deStefano and Lefebvre (2007) review 38 experimental studies on the cognitive consequences of hypertext reading "in order to test the hypothesis that activities specific to hypertext can increase cognitive load and impair learning." (p. 1617). The researchers wonder whether, "When people follow links, they may lose track of where they are in the text, of their reading goals, of the larger context for the node, or of material activated in working memory" (p. 1627). Their review concludes that generally in the experiments reviewed, "many features of hypertext resulted in increased cognitive load and thus may have required working memory capacity that exceeded readers' capabilities" (p. 1636), or at least some readers' capabilities, since "the various manipulations [entailed in linked or hypertext reading] were often most detrimental to the reading processes of less-knowledgeable readers and for readers with low working memory capacity" compared to traditional linear presentations of text.

Few armchair ideas ever get realized, fewer still without surprises. Vannevar Bush formulated his vision of creating and reading linked texts from the perspective of a longstanding offline or linear reader. He almost certainly never thought about what kind of reading would take place once readers were "digital natives," virtually raised on associatively linked or hypertextual reading. The idea is now emerging that coping with multiple interlinked texts constitutes something of a new literacy, which like the

old linear model will have to be the subject of deliberate training and extensive practice, possibly involving either new cognitive skills, or a type of bi-modality between two literacies (see Ong, 1982, p. 175, for speech and text; and Wolf, 2007, Ch. 9, for text and hypertext), or both. It is often said that learning to talk is free while learning to read is expensive creating winners and losers (Gough & Hillinger, 1980; deHaene, 2009); it will probably turn out that learning to read hypertexts while beneficial to the able will be even more expensive, like reading but more so. At the moment, children appear to be floundering in the new literacy while their teachers persist with training and testing in the old---yet another digital divide.

See also

Critical Media Literacy, Language and the Digital Divide, New Literacies of Online Reading Comprehension, Searchlinguistics, Teaching Reading, Technology and Literacy

References

Antoniou, G. & Van Harmelen, F. (2004) *A semantic Web primer.* Cambridge MA: MIT Press.

Bush, V. (1945). As we may think. *The Atlantic Monthly*, July, 101-108. [Online: http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/3881/ .

Carr, N. (2008). Is Google making us stupid? *Atlantic Monthly,* New York, July.

Carr, N. (2010) *The Shallows: How the Internet is Changing the Way We Think, Read and Remember.* New York: Norton.

deHaene, S. (2009). *Reading in the Brain.* New York: Viking.

deStefano, L. & Lefevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior 23,* 1616–1641.

Gough, P., & Hillinger, M. (1980). Learning to read: An un-natural act. *Bulletin of the Orton Society 30,* 179-. 196.

Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development 1* (4), 309-317.

Rowlands, I., & Nicholas, D. (2008). Information behaviour of the researcher of the future. Centre for Information Behaviour & the Evaluation of Research (CIBER), University College London. [Online at http://www.bl.uk/news/pdf/googlegen.pdf.]

Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Data Engineering Bulletin 24(4), 35-43. [Online http://singhal.info/ieee2001.pdf]

Wolf, M. (2007). *Proust and the squid: The story and science of the reading brain.* New York: HarperCollins.

Further reading

Nyce, J. & Kahn, P. (1991). *From Memex to Hypertext: Vannevar Bush and the Mind's Machine.* San Diego: Academic Press.

Wardrip-Fruin, N. & Montfort, N. (Eds.) (2003). *The New Media Reader.* Cambridge MA: MIT Press.