A new paradigm for domain-specific search to fight human trafficking

February 10 2014
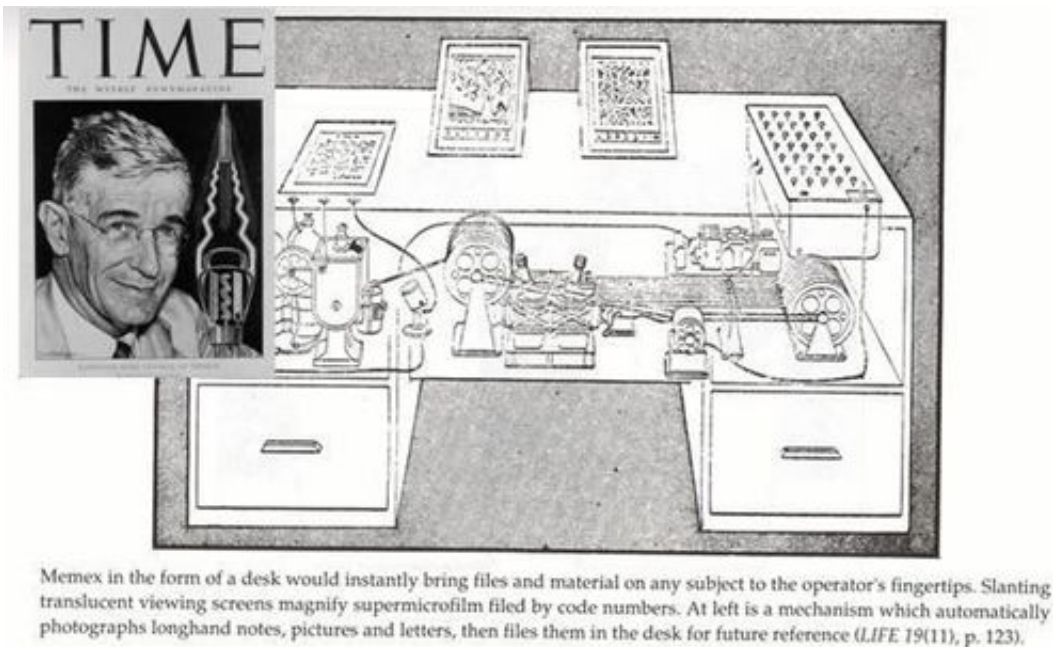
Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

Memex seeks to develop the next generation of search technologies and revolutionize the discovery, organization and presentation of search results. The Memex program gets its name and inspiration from a hypothetical device described in "As We May Think," a 1945 article for The Atlantic Monthly written by Vannevar Bush, director of the U.S. Office of Scientific Research and Development (OSRD) during World War II.

Today's web searches use a centralized, one-size-fits-all approach that searches the Internet with the same set of tools for all queries. While that model has been wildly successful commercially, it does not work well

for many government use cases. For example, it still remains a largely manual process that does not save sessions, requires nearly exact input with one-at-a-time entry, and doesn't organize or aggregate results beyond a list of links. Moreover, common search practices miss information in the deep web—the parts of the web not indexed by standard commercial search engines—and ignore shared content across pages.

To help overcome these challenges, DARPA has launched the Memex program. Memex seeks to develop the next generation of search technologies and revolutionize the discovery, organization and presentation of search results. The goal is for users to be able to extend the reach of current search capabilities and quickly and thoroughly organize subsets of information based on individual interests. Memex also aims to produce search results that are more immediately useful to specific domains and tasks, and to improve the ability of military, government and commercial enterprises to find and organize mission-critical publically available information on the Internet.

"We're envisioning a new paradigm for search that would tailor indexed content, search results and interface tools to individual users and specific subject areas, and not the other way around," said Chris White, DARPA program manager. "By inventing better methods for interacting with and sharing information, we want to improve search for everybody and individualize access to information. Ease of use for non-programmers is essential."

Memex would ultimately apply to any public domain content; initially, DARPA intends to develop Memex to address a key Defense Department mission: fighting human trafficking. Human trafficking is a factor in many types of military, law enforcement and intelligence investigations and has a significant web presence to attract customers. The use of forums, chats, advertisements, job postings, hidden services,

etc., continues to enable a growing industry of modern slavery. An index curated for the counter-trafficking domain, along with configurable interfaces for search and analysis, would enable new opportunities to uncover and defeat trafficking enterprises.

Memex plans to explore three technical areas of interest: domain-specific indexing, domain-specific search, and DoD-specified applications. The program is specifically not interested in proposals for the following: attributing anonymous services, deanonymizing or attributing identity to servers or IP addresses, or accessing information not intended to be publicly available. The program plans to use commodity hardware and emphasize creating and leveraging open source technology and architecture.

The Memex program gets its name and inspiration from a hypothetical device described in "As We May Think," a 1945 article for *The Atlantic Monthly* written by Vannevar Bush, director of the U.S. Office of Scientific Research and Development (OSRD) during World War II. Envisioned as an analog computer to supplement human memory, the memex (a combination of "memory" and "index") would store and automatically cross-reference all of the user's books, records and other information.

This cross-referencing, which Bush called associative indexing, would enable users to quickly and flexibly search huge amounts of information and more efficiently gain insights from it. The memex presaged and encouraged scientists and engineers to create hypertext, the Internet, personal computers, online encyclopedias and other major IT advances of the last seven decades.

Provided by DARPA