

Challenges in Using Lifetime Personal Information Stores

Gordon Bell

Microsoft Bay Area Research Center
455 Market St. Suite 1690
San Francisco, CA 94105
(01) 415 778 8225
gbell@Microsoft.com

Jim Gemmell

Microsoft Bay Area Research Center
455 Market St. Suite 1690
San Francisco, CA 94105
(01) 415 778 8222
jgimmell@Microsoft.com

Roger Lueder

Microsoft Bay Area Research Center
455 Market St. Suite 1690
San Francisco, CA 94105
(01) 415 778 8235
rlueder@Microsoft.com

Extended Abstract

Within five years, our personal computers with terabyte disk drives will be able to store everything we read, write, hear, and many of the images we see including video. Vannevar Bush outlined such a system in his famous 1945 Memex article [1].

For the last four years we have worked on MyLifeBits www.MyLifeBits.com, a system to digitally store everything from one's life, including books, articles, personal financial records, memorabilia, email, written correspondence, photos (time, location taken), telephone calls, video, television programs, and web pages visited. We recently added content from personal devices that automatically record photos and audio.

The project started with the capture of Bell's content [2], followed by an effort to explore the use of the SQL database for storage and retrieval. Work has continued along these lines to extend content capture from every useful source e.g. a meeting capture system. The second phase of the project includes the design of tools and links for annotation, collections, cluster analysis, facets for characterizing the content, creation of timelines and stories, and other inherent database related capabilities, e.g. the ability to pivot on an event or photo or person to retrieve linked information [3]. Ideally we would like to have a system that would read every document, extract meta-data (e.g. Dublin Core) and classify it using multiple ontologies, faceted classifications, or the relevant.

While such a system has implications for future computing devices and their users, these systems will only exist if we can effectively utilize the vast personal stores. Although our system is exploratory, the Stuff I've Seen system [4] demonstrates the utility and necessity of easy search and access to one's own data. Other research efforts with similar goals relating to personal information include Haystack [5], LifeStreams [6], and the UK "Memories for Life" Grand Challenge.

There are serious research issues beyond the problem of making the information useful through rapid and easy retrieval. The "Dear Appy¹" problem ("Dear Appy, My application, or platform, or media left me unreadable. Signed, Lost Data") is unsettling to archivists and computer professionals –and must be solved.

Just navigating the stored life of individual would at first glance appear to take almost a lifetime to sift through. While we are making progress in the capture of less traditionally archived content (e.g. meetings, phone calls & video), automatic interpretation and index of voice are illusive. MyLifeBits is currently focused on retrieval including the *hopefully automatic*, addition of meta-data e.g. document type identification, high level

knowledge. While such data is essential for the archivist, it is unclear how useful such meta-data is to a one's own information; without such higher level knowledge and concepts, the vast amount of raw bits may be completely unusable.

The most cited problem of personal archives is the control of the content including personal security, together with joint ownership of content by other individuals and organizations. In many corporations, periodic expunging of documents is the standard. Similarly, the aspects of a person's life not available in public documents is owned by the organization and all documents may have to be tagged in such a way that it can be expunged, if necessary, when an individual is no longer part of the organization. The HPPA law in the US and even more stringent privacy laws in other countries have major implications for personal stores.

Categories and Subject Descriptors

B. Hardware: General

H.3.3 Information Search and Retrieval; general, clustering, filtering, query formulation, search process

H.5.2 User Interfaces (H.1.2, I.3.6) Graphical user interface.

Keywords

Personal information storage and retrieval; recording devices for personal information, facets, hyper-text, ontology.

REFERENCES

- [1] Vannevar Bush. As We May Think, The Atlantic Monthly, 176(1), July 1945, pp.101-108.
- [2] Gordon Bell, A Personal Digital Store, Communications of the ACM, January 2001, 44 (1), pp. 86-92.
- [3] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker and Curtis Wong, MyLifeBits: Fulfilling the Memex Vision, ACM Multimedia '02, December 1-6, 2002, Juan-les-Pins, France, pp.235-238.
- [4] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, Daniel C. Robbins, Stuff I've Seen: A System for Personal Information Retrieval and Re-Use, ACM SIGIR 03', Toronto, Canada.
- [5] Eytan Adar, David Karger, and Lynn Andrea Stein. Haystack: Per-User Information Environments, In Proceedings of CIKM '99 (Kansas City, MO, November 2-6, 1999), ACM Press, 413-422.
- [6] Freeman, Eric, Gelernter, David. LifeStreams: A storage model for personal data, ACM SIGMOD Bulletin 25, 1, March 1996, pp. 80-86.
- [7] www.nesc.ac.uk/esi/events/Grand_Challenges/gcconf04/