



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

**“A Brief History of Ebooks”
INFO 290-5
Fall 2013**

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

3 September 2013



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

A Brief History of Ebooks

- The idea of the ebook is at least 50 years older than the word “ebook”
- How we understand the history of the ebook depends on how we define ebooks



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Important Events / Eras in Ebook History

- Visions of ebooks before there were any
- Visionary pragmatists; ebooks as digital texts that take advantage of their computing context before book-like ebooks are possible
- Formative design duels: card sharks vs holy scrollers; literati vs the engineers
- Dismissing and missing the web as an ebook platform
- 10 years of ebook readers
- Today: Books *on the web* vs books *of the web*

3




UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Defining the Ebook

- Any definition of ebook has to be consistent with its technology context
 - What could an ebook be before there were any computers?
 - What could an ebook be before there were book-size computers?
 - How did / does the web change the notion of ebooks?

4



 UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Definitions of Ebooks


- Electronic or digital texts:
 - viewable on a computer OR
 - downloadable to a device OR
 - printable on demand
- Electronic or digital books – a package of hardware and software that enables “book-like” interactions
 - dedicated functionality OR
 - one of many functions on the device OR
 - book reading apps
- Any app with interactive information content

More “book like”

Less “book like”

5

 UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Definitions of Ebooks

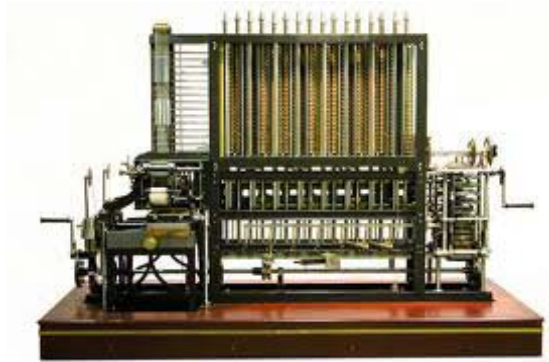
- How important is the book metaphor to your notion of an ebook?
- How is the definition of an ebook shaped by document type or genre?

6



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

When This is a Computer, What Can An "Ebook" Be?



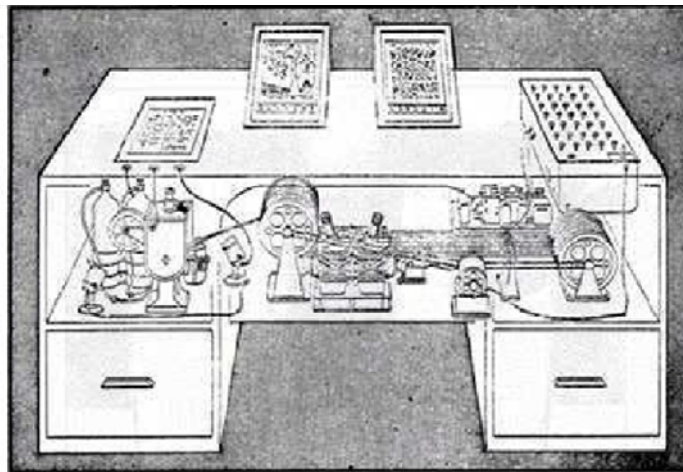
Babbage 19th Century "Difference Engine" recreated at Computer History Museum

7



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

The Memex (about 1940)



8



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Memex – The “Knowledge Workstation”

- Bush’s Memex was a device in which an individual stores all his books, records, and communications
- It is a mechanical device, but the information it contains (on microfilm) “may be consulted with exceeding speed and flexibility.”
- Bush's most influential idea was his proposal for organizing sets of related resources as “trails” connected by associative links, the ancestor of the hypertext links that define today's web
- Bush, Vannevar. “As We May Think”: The Atlantic. 1945
www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

9



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Memex – The “Knowledge Workstation”

- Bush imagined that Memex users could share these packages of trails and that a profession of trailbuilders would emerge.
- However, he did not envision that the Memexes themselves could be interconnected, nor did he imagine that their contents could be searched computationally.

10



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

When This is a Computer, What Can An “Ebook” Be?



11



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Building on Bush

- Doug Engelbart
 - Builder of “augmentation” technology, invented many ideas and devices fundamental to interactive computing, including the mouse, window displays, collaborative annotation
- Ted Nelson
 - Visionary who conceived of many hypertext publishing concepts, ideas of reuse by linking and transclusion
 - But his Xanadu system never implemented successfully

12



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Alan Kay, PARC, and the Dynabook

- Allan Kay's 1969 PhD thesis imagined a tablet computer called the Dynabook with text and animated graphics, multiple windows, and a user interface that enabled interactions with documents stored on the device and available via networks
- Kay came to the legendary Xerox PARC (Palo Alto Research Center) with the goal of designing and deploying the Dynabook, but it was ahead of its time
- However, Kay's work at PARC was fundamental in changing the computer from a computational machine into something that supported communication and collaboration, and even though the Dynabook didn't get built, the Alto did (which was also inspired by Engelbart's work)

13



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

The Alto – the “proto Dynabook”



Figure 8. Xerox Alto II Workstation. (Courtesy of Computer History Museum.)

The Alto was a desktop computer with a specially constructed monitor. It was self-contained, essentially a small box in which disk memory could be inserted. Each disk had 1,500-page equivalents of storage.

Unlike the character-based displays of the time, the Alto bit-mapped display could render documents that looked professionally typeset.

The Alto also had a mouse and the popular desktop environment of icons, folders, and documents

Barnes, Susan B. "Alan Kay: Transforming the computer into a communication medium." *Annals of the History of Computing*, IEEE 29, no. 2 (2007): 18-30.

14



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

The Pragmatists at Bell Labs, IBM, and elsewhere

- Computers store the text of computer programs - why not use them to store documents?
- Let's treat documents more like software in other ways – version control, configuration management, automated error checking, makefiles, updating
- SGML (“Standard Generalized Markup Language”) emerged from work at IBM to improve mainframe-based authoring and publishing of documents (in the legal department)
- PCs would be underpowered for ebooks until almost 1990; most ebook work was in mainframe, minicomputer, and workstation environments because of greater computing capability, storage capacity, networking

15



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Project Gutenberg

- Started in 1971 at the University of Illinois, first “ebook library,” when computing technology was becoming generally available for entering, storing, and displaying text
- ASCII text as least common denominator format to increase interoperability
- Thousands of volunteers have typed and validated tens of thousands of out-of-copyright books

<http://www.gutenberg.org/>

16



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Bell Labs

- For decades the world's leading research laboratory, lavishly funded by AT&T phone monopoly in US
- Numerous innovations in electronic documents (1975-1990)
 - Online docs (preformatted, downloadable)
 - SOLID - Single-source reuse, automation – treat documents like software
 - Superbook – software for automatically creating ebooks from the source files used for print formatting
 - Documenter's Workbench – tools for measuring and improving the quality of documents

17



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

System for Online Information Development (SOLID)

- Best way to develop documents that support complex software systems is with tools and methods analogous to those for software
- Best way to deliver software documents is online, using the computer that runs the software
- By 1984 SOLID was used by dozens of software development teams in Bell Labs and later spawned SOFT, first workgroup document management system, with collaborative workflow support, version control, configuration management
- Bianchi, M., Glushko, R., & Mashey, J. "[A software/documentation development environment built from the UNIX toolkit](#)," 1982
- Glushko, Robert J. [Text development and management in Unix-based projects](#), 1984

18

Copyright © 1982 American Telephone and Telegraph Company
 THE BELL SYSTEM TECHNICAL JOURNAL
 Vol. 61, No. 6, July-August 1982
 Printed in U.S.A.

On-line Documentation: Mechanizing Development, Delivery, and Use

By R. J. GLUSHKO and M. H. BIANCHI

We describe the design and development of on-line documentation for a minicomputer-based management information system. We outline the design choices, compare on-line documents with paper ones, and review human engineering and "software psychology" issues. On-line documents are accessed from any dial-up terminal. Document retrieval shares a common user interface with other information activities like report generation, trouble reporting, and interuser communication. Documents are "modular" with properties that make them easier to create, use, and maintain.



UNIVERSITY OF CALIFORNIA, BERKELEY
 SCHOOL OF INFORMATION

SuperBook

- SuperBook is a "structured browsing system," designed to improve the usability of existing documents through computer-based enhancements.
- SuperBook takes as input ordinary text formatting language and converts it into a multi-windowed browser with rich search, navigation, and annotation, and display enhancements.
- Egan, Dennis E., Joel R. Remde, Louis M. Gomez, Thomas K. Landauer, Jennifer Eberhardt, and Carol C. Lochbaum. "Formative design evaluation of superbook." *ACM Transactions on Information Systems (TOIS)* 7, no. 1 (1989): 30-57.

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

SuperBook

S - An Interactive Environment for Data Analysis

HELP NEW TEXT PAGE

Table of Contents

- Preface
- 1 •Chapter 1 -- How to Beat the Lottery
- 7 •Chapter 2 -- Tutorial Introduction to S
- 5 •Chapter 3 -- Basic Use of S
- 15 •Chapter 4 -- Graphical Methods in S
- 1 •Chapter 5 -- Advanced Use of S
- 13 •Chapter 6 -- The S Macro Processor
- 1 •Chapter 7 -- Data Analysis Using S
- Chapter 8 -- Presentation: Graphics and Reports
- 34 •Appendix 1: Function and Macro Documentation

Expand Collapse Find word Move focus

Distribution of word(s) data, structure

Find the section discussing the basic concept that the value of any expression, however complicated, is a data structure.

Word lookup

data, structure

Enter word

Found 76 occurrences of: data, structure

Find word Find section Page Remove

Page of text

21

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Xerox PARC NoteCards

Capabilities of New Missiles

Even though the weapons in question replace older weapons (the Pershing IA and the Vulcan bomber), both are capable of more destruction faster than their predecessors. This is the result of new radar guidance systems, with new levels of accuracy. Also have sufficient range to make vulnerable installations and cities in the Western USSR, in the case of the P 2, within a matter of minutes.

(p. 371) See [Guidance of Pershing II](#)

Map: Missile Ranges

Guidance of Pershing II

"The new American Pershing II missile, fitted with a radar-homing warhead, is designed to be even more accurate. As it falls back to earth this compares a radar image of the target with an image stored in its computer memory. It should then be able to adjust its flight path so as to hit its target with pin-point accuracy after a journey of 1,800 kilometers." (p. 13)

See [Unspecified Tomahawk Characteristics](#)

Tomahawk Characteristics

Tomahawk cruise missile: jet engine produces speeds of 800km/h over distances of 2,600 km. Missile carries a computer which is programmed with maps of the area missile is to fly over, so can compare actual position with programmed course and correct course. Computer is designed to allow missile to follow a zig-zag

Halasz, Frank. "Reflections on NoteCards: seven issues for the next generation of hypermedia systems." *Communications of the ACM* 31, no. 7 (1988): 836-852.

22



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Cards vs Scrolls

- “Block mode” interaction with mainframe computers implies a “card” or “frame” unit
- In the early 1970s, ZOG at CMU pioneered the idea that long documents could be implemented as sets of interconnected frames
- Scrolling is possible when local storage enables a unit of text to contain more than one screen-size of text
- But the debate involves questions about navigation and organization, not just implementation, and search capability makes much of the argument moot

23



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Literati vs Engineers

- Literati: Hypertext is a new form of writing
- Engineers: Hypertext is a new information management and delivery model for existing text
 - “ebooks” are a means, not an end, and creating them requires a systematic and disciplined approach to task and document analysis
 - Card model programs impose explicit or implicit constraints on the size of the hypertext units, the nature of links, and the complexity of access methods that fail to capture the structural complexity of existing printed materials

24



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Advice on Authoring for Hypertext

- Instead of a single string of paragraphs, the author lays out a textual space in which the fiction operates.
- The reader joins in actively constructing the text by selecting a particular order of episodes
- The extent of the reader's choices and therefore his freedom in examining the literary space depends on the links that the author creates between episodes
- The links come with a condition statement... and may require that the reader answer a question, or may require that the reader has already visited a particular episode
- Revisiting a place that one has seen before – was once seen as a sign of disorientation, inefficiency, or artistic affectation. As hypertext readers gained experience, however, they recognize that recurrence was the way to perceive structure

Bolter, Jay David, and Michael Joyce. "[Hypertext and creative writing](#)." In *Proceedings of the ACM conference on Hypertext*, pp. 41-50. ACM, 1987.

Bernstein, Mark, Michael Joyce, and David Levine. "[Contours of constructive hypertexts](#)." In *Proceedings of the ACM conference on Hypertext*, pp. 161-170. ACM, 1982.

25



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Advice on Converting Books to Ebooks

- How closely will the computer version preserve the physical structure of the document?
- How closely does the physical structure preserve the logical structure of the document?
- Before preserving all aspects of the logical structure of the document, consider the implementation tradeoffs involved
- It is always a challenge to separate the aspects of physical structure that reflect the logical structure from those that are shaped by stylistic, aesthetic, or entirely arbitrary factors

Glushko, Robert J., Mark D. Weaver, Thomas A. Coonan, and Janet E. Lincoln. "[Hypertext engineering: practical methods for creating a compact disk encyclopedia](#)." In *Proceedings of the ACM conference on Document processing systems*, pp. 11-19. ACM, 1988.

26

SANTA FE, NEW MEXICO, DECEMBER 5-9, 1988

"Hypertext Engineering": Practical Methods for Creating A Compact Disc Encyclopedia

Robert J. Glushko
Mark D. Weaver
Thomas A. Coonan
Search Technology, Inc.
Norcross, GA

Janet E. Lincoln
Stuyvesant, NY

Since no established methodology existed for creating an electronic encyclopedia from a printed one, we set out to develop our own approach to "hypertext engineering."

27



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Hypercard, Guide, Hyperties, Dynatext, etc.

- Around 1990 several tools for building hypertext documents and related applications on personal computers emerged
- Many used the new CDROM format to store large document libraries or multimedia content managed by the application
- Most used the card metaphor and were designed to create new applications

28



HyperCard

- http://archive.org/details/CC501_hypercard
- Software came with every Apple Macintosh computer and quickly became the default development platform for many types of applications
- Information is stored in a series of "cards" arranged into "stacks."
- Cards can be linked to each other
- A built-in, plain-English programming language, HyperTalk, executes commands.
- HyperCard supports pictures, audio and video, and can be used to make almost anything -- from games to accounting systems, scientific data tools and teaching aides.



A Hypertext Engineer Compares the Programs

Table 1—Comparison of the Three Programs

PROGRAM DESIGN ISSUE	Guide	HyperCard	HyperTIES
Granularity	good	OK	fair
Non-text components	good	excellent	poor
Links	excellent	good	OK
Access methods	good	fair	good
Navigation / sessions	good	good	good
Extensibility	OK	excellent	poor

Glushko, Robert J. "Using off-the-shelf software to create a hypertext electronic encyclopedia." *Technical Communication, First Quarter 1990 (as listed in the ACM HyperText Compendium)* (1990).



The ACM Hypertext Conferences

- [Starting in 1987](#), for many years the ACM conference was the premier venue for research reports on hypertext and ebooks...
- TBLs paper on the web was rejected for 1991 conference, and the ACM community failed to embrace the web.. Which led to splintering and marginalization.
- <http://ibiblio.org/pjones/blog/the-story-behind-the-hypertext-91-demo-page-and-unc-and-me/>
- But still highly relevant to ebook design – now in 25th year



“Ebooks” in 1987

TABLE I. Architectural Features of an Average Current Generation Hypermedia System

Feature	Description
Nodes:	Typed (text, graphics, . . .), implemented using a type hierarchy Nodes cannot contain other nodes
Links:	Binary, bidirectional Labeled but not typed Anchors can be whole nodes or points/regions within the node
Overviews:	Browsers containing node/link diagrams of the network Can edit network via browser
Hierarchies:	Special support for hierarchical networks
User interface:	Multiple windows; mouse/menu driven
Extensibility:	Programmer's interface
Search/query:	Slow, full-text string match
Distribution:	Single-user or multi-user central server with limited concurrency control
Versioning:	None
Storage:	Standard files or relational DBMS

Halasz, Frank. "Reflections on NoteCards: seven issues for the next generation of hypermedia systems." *Communications of the ACM* 31, no. 7 (1988): 836-852.



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Seven Issues for the Next Generation

- Search across a network
- Hierarchical composites, not just simple nodes and links
- Virtual / dynamic structures
- Computation across a network
- Versioning
- Support for collaborative work
- Extensibility and tailorability

33



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Dismissing and Missing the Web

- The web's simple hypertext model (one way, binary, untyped links) and primitive UI was seen as a simplistic step backward and dismissed by the hypertext and document research community
 - TBL's 1991 paper rejected by ACM Hypertext conference
- But the web traded expressive power for scale, and Netscape's browser made it much simpler to create web pages and use the web
- Within a few years, the web was effectively the universal interface to electronic text, so we need to use it as the baseline against which we compare subsequent developments

34



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

The Web and “Ebook Architecture”

- Web returns to the card metaphor, but now a card can link anywhere in the web
- Linking remains highly simplistic
- Much of the “card” content is virtual or dynamically generated from a database, ERP, or other big software application, and often based on the user/reader’s behavior
- Perceived limitations of the “plain web” lead to proprietary add-ons (e.g. Flash), non-interoperable coding practices, and walled gardens that undermine the vision of an open hypertext web

35



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

The 13 Year Emergence of Book Readers

- [Two Decades of E-reader Evolution](#) ...
- begins with Sony Data Discman / Bookman – 1990
<http://www.youtube.com/watch?v=opDD0rS5ats>
- Amazon’s Kindle family, launched in 2007, is the most popular dedicated reader
- But still new entrants: New Kobo:
<http://www.engadget.com/2013/08/27/kobo-aura-hands-on/>
- Reading apps on tablets support most of the functionality in a more general-purpose device; Apple iBooks is most popular example
- Numerous specialized readers: Vook, Inkling

36



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Engineers vs Literati, Revisited

- Most ebooks today are essentially “engineering” ebooks – converted from the backlist of print books as efficiently as possible
- This typically means they offer little capability beyond simple page turning and search
- But highly valuable books (like widely used textbooks) are redesigned and redeveloped by “literati” because the cost of hand-crafting can be amortized over high sales volume
- Challenge as the backlist runs out - can we create enhanced ebooks using scalable engineering methods?

37



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Web vs Ebook vs App

- Emergence of tablet computers, especially the ipad, creates a third category of ebooks – those that run in proprietary software applications
- These are typically distributed on the web, but they don’t use native web standards and protocols and use proprietary formats in walled gardens
- If the same content is on the web, in an ebook reader, and an App, is it one ebook or 3?

38



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

A Brief History of Ebooks

- The idea of the ebook is at least 50 years older than the word “ebook”
- How we understand the history of the ebook depends on how we define ebooks
- History is far from being written. Lots of evolution and innovation to come