

2. INFO 202 - ISSUES AND CONTEXTS

INFO 202 - 3 September 2008

Bob Glushko

Plan for INFO 202 Lecture #2

Discussion of 8/27 readings: Memex, My Life Bits, Library of Babel

Discussion of today's readings

- Operation Clean Data
- Virulent Veggies
- Electronic Health Records
- Data Unification in PIM
- Enterprise Search
- Machine Translation

"202 in the News" Stories

As We May Think

Who was Vannevar Bush?

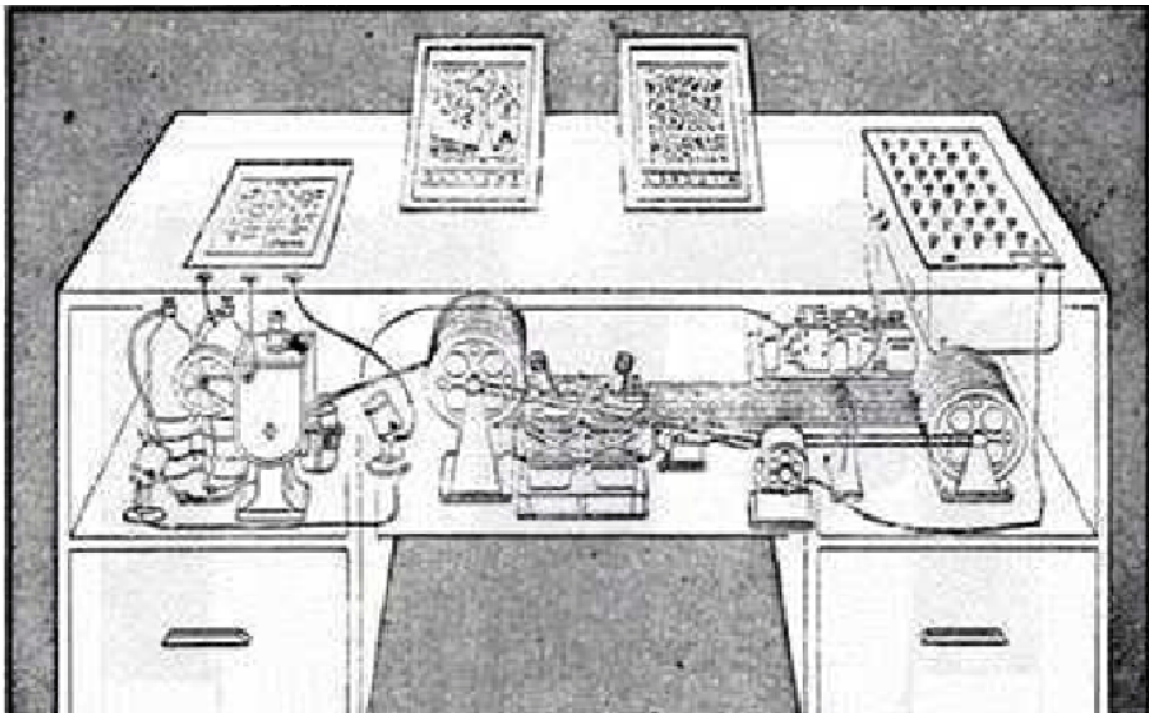
What problems is the Memex designed to solve?

What are the key TECHNOLOGY innovations in Memex? Are they plausible?

What are the key CONCEPTUAL innovations in Memex?

Is the Memex more like a precursor to the web, or the ipod?

The Memex



My Life Bits

What is the relationship of "My Life Bits" to Memex?

What types of content are being captured and stored?

What does the Sensecam sense, and what's the point?

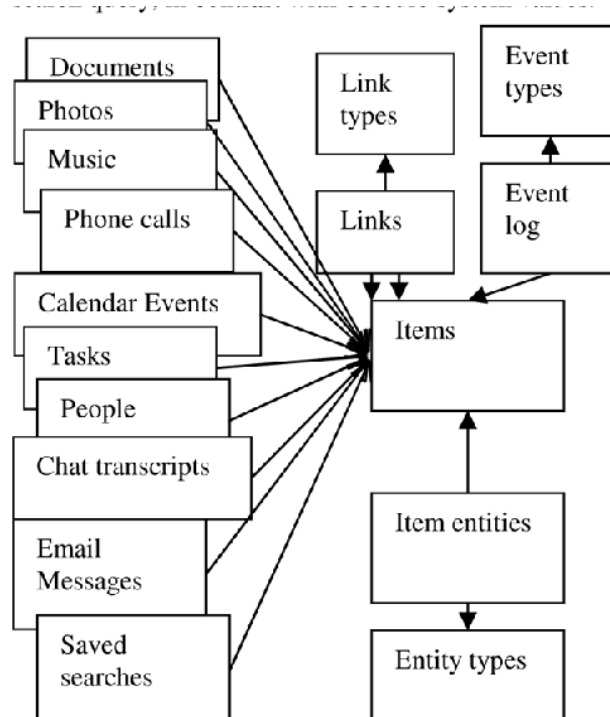
How much and what kind of information organization is going on?

"The more that is captured, the more correlation is possible to help find things"-- why? how?

My Life Bits -- the SenseCam



My Life Bits -- Platform and Tools



The Library of Babel

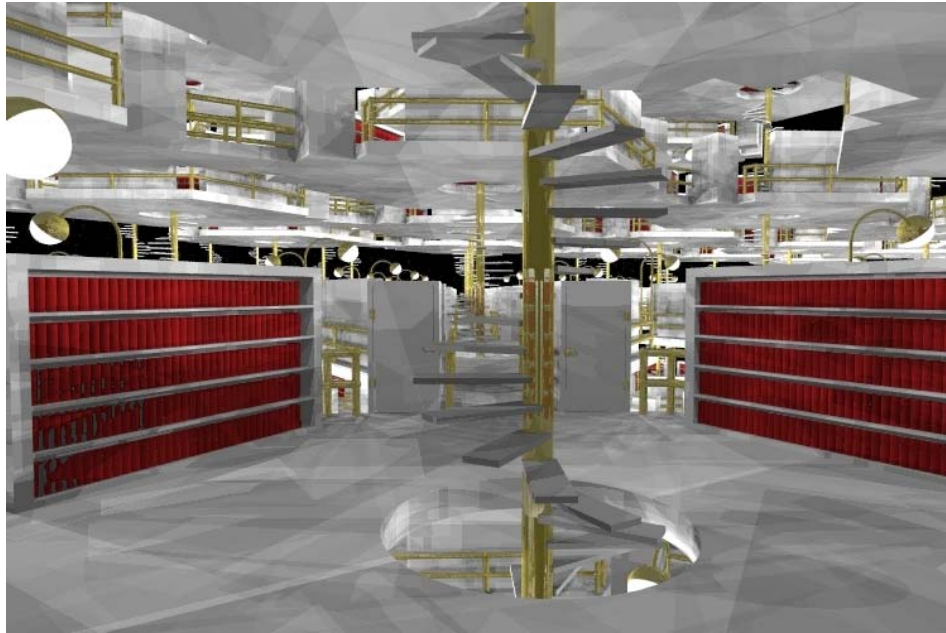
"When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness"

"... this inordinate hope was followed by an excessive depression"

Is a library containing all the world's knowledge feasible?

How can a Man of the Book be sure he has found the true Book, and not an imperfect (or even deceptive) copy?

A Depiction of Borges' Library



"Operation Clean Data" -- British Military Case

What were the symptoms or implications of "dirty" data in the British army's supply chains?

What were the primary causes of this "dirty" data?

Which data items were the focus of the data cleanup effort? Why?

What technologies or tools were used in the data cleanup effort?

The US Didn't Learn from the Brits

The New York Times
nytimes.com

March 26, 2008

U.S. Sent Missile Parts to Taiwan in Error

By [THOM SHANKER](#)

WASHINGTON — Four high-tech electrical fuses for Minuteman nuclear warheads were sent to [Taiwan](#) in place of helicopter batteries, a mistake discovered only last week — a year and a half after the erroneous shipment, Pentagon officials disclosed Tuesday.

Officials said the nose-cone fuses, designed to send an electrical signal to the warhead's trigger as it plummets toward its target, resembled those used for conventional munitions and contained no nuclear material.

"Operation Clean Data" **-- Carlson Wagonlit Case**

What were the symptoms or implications of "dirty" data for the Carlson Wagonlit travel agency?

What were the primary causes of this "dirty" data?

How is Carlson Wagonlit improving its data quality?

"Operation Clean Data" -- Cendant Case

What were the symptoms or implications of "dirty" data for Cendant?

What were the primary causes of this "dirty" data?

How is Cendant improving its data quality?

"Tailing Virulent Veggies" (WSJ, 3/13/07)



March 13, 2007

Tailing Virulent Veggies

**Produce Industry Develops
Means to Pinpoint Origin
Of Contaminated Products**

By JANE ZHANG



Scanning RFID Tags on Vegetable Boxes



Electronic Health Records: Just Around the Corner? Or over the Cliff?

A case study of the adoption of an electronic health record system by a small (4 physicians) medical office

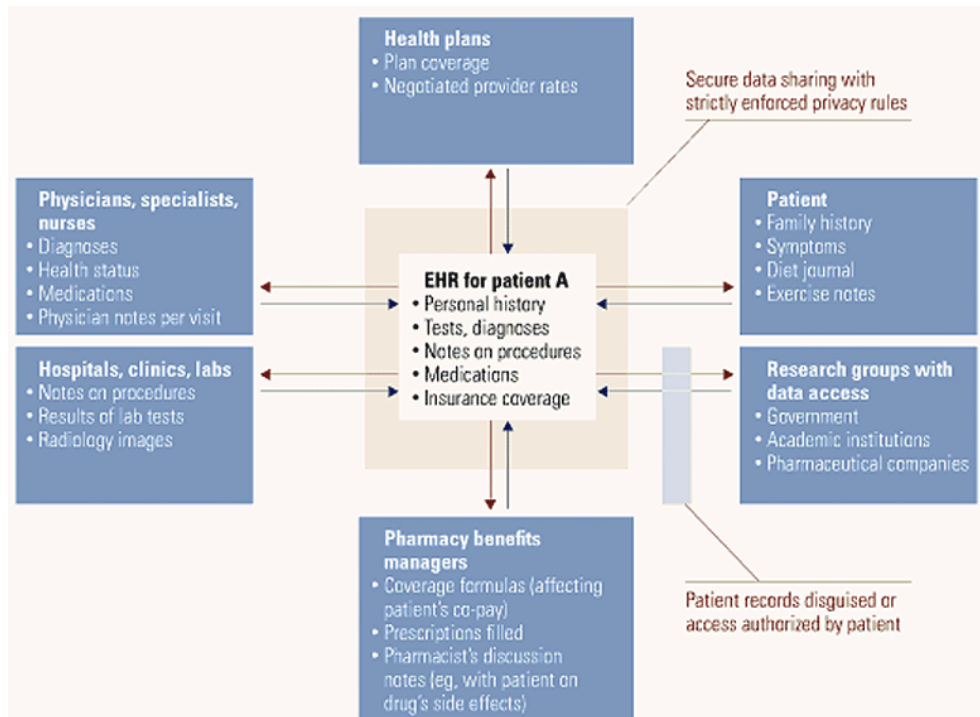
What were the primary motivations for installing the system?

Did they do a good job selecting the system and ensuring that it was appropriate to their requirements?

Were their expectations about installation, training, and operation reasonable?

Of the problems they encountered, which were preventable, and which ones weren't?

EHR Constituents and Perspectives



EHR Models and Formats

Some EHRs have richly-structured multimedia information models

Others are much simpler - word processing documents or text "notepad" formats

What are the rationales and implications of each of these models and formats?

EHR Architectures

Microsoft and Google have proposed a patient-centered or Personal Health Record that collects information from healthcare providers

However, most EHRs are designed not as a "single physical entity" but as a "functional view assembled when needed" from data stored in multiple repositories

What are the rationales and implications of each of these architectures?

Challenges on the way to EHRs

The healthcare ecosystem is extremely complex and fragmented

Large short-term investment required, and cost reductions are uncertain and longer-term

Mismatch of benefits and incentives - who pays, and who benefits?

Insufficient standardization, interoperability problems

Data Unification in PIM

What problems will be solved by data unification?

What aspects of these problems are related to technology?

What aspects of these problems are related to individual choices and decisions?

What aspects of these problems are related to business models?

Are common data formats the solution? Should they be "least common denominator" formats? Who should specify the format?

Data Unification in PIM -- Techniques

<i>APPROACH</i>	<i>EXAMPLE</i>	<i>OPERATIONS</i>	<i>ENABLES</i>
Standard datatype	Text	Cut/copy/paste	Unified search
Unified presentation	Window manager	Layout, tile, show, hide	Simultaneous view
Unified namespace		Reference, de-reference	
Grouping	Directories	Group, ungroup	Organize, browse
Metadata	ID3 tags, XML	Annotate, query	Search, organize, browse
Cross-ref	Web links, OLE	Link, follow, embed	Simultaneous view, orient
Relations	RDF, Haystack	Record named relations	Unified search, browse, orient

Enterprise Search: Tough Stuff

How is enterprise content differ from web content?

How do these differences in content types influence the mechanisms and methods of enterprise search?

How does "the way an enterprise works" influence the goals, mechanisms and methods of enterprise search?

The Elusive Goal of Machine Translation

Have you ever used an application of machine translation?

How can "brute force" techniques that use statistical properties of text ("machine learning") sometimes outperform those that use grammatical and lexical knowledge?

How has the web changed the importance of machine translation?

How has the web changed the cost and feasibility of machine translation?

Will machine translation ever be as good as translation by people? Is this an important requirement?

What Are the Common Themes in Stories?

Many existing information organization, management, and transformation practices would benefit from redesign, automation, integration, interoperability

Standards / patterns for information components / documents and the associated processes are essential

Practices evolve and co-evolve with new technologies

But projects can be challenging, and their success depends on many factors besides technology

"202 in the News" Stories

Let's talk about some of them...

Are there any connections between the stories we read for today and ones that you posted?

Readings for Monday 8 September

Elaine Svenonius, *The Intellectual Foundation of Information Organization*, Preface & Chapters 1-2

David Weinberger, *Everything is Miscellaneous*, Preface and Chapter 1
<http://www.everythingismiscellaneous.com/samples/>

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Sections 8.1-8.3

Class of 2010 202 Party

Tuesday, September 9, 6-9pm

At my house in the Berkeley hills, 34 Stephens Way

No public transportation, so this will encourage you to talk to each other