Digital Libraries

Orientation

Week 0

Min-Yen KAN

What is a library?

1. A place set apart to contain books for reading, study, or reference.
   - (Not applied, e.g. to the shop or warehouse of a bookseller.)
2. A building … containing a collection of books for the use of the public or of some particular portion of it, or of the members of some society or the like;
3. a public institution or establishment, charged with the care of a collection of books, and the duty of rendering the books accessible to those who require to use them.

What is a library?

1. A private commercial establishment for the lending of books, the borrower paying either a fixed sum for each book lent or a periodical subscription.
2. a great mass of learning or knowledge;
3. the objects of a person's study, the sources on which he depends for instruction.
4. *Computers*. An organized collection of routines, esp. of tested routines suitable for a particular model of computer
5. *Biology*. a collection of sequences of DNA … that represent the genetic material of a particular organism or tissue

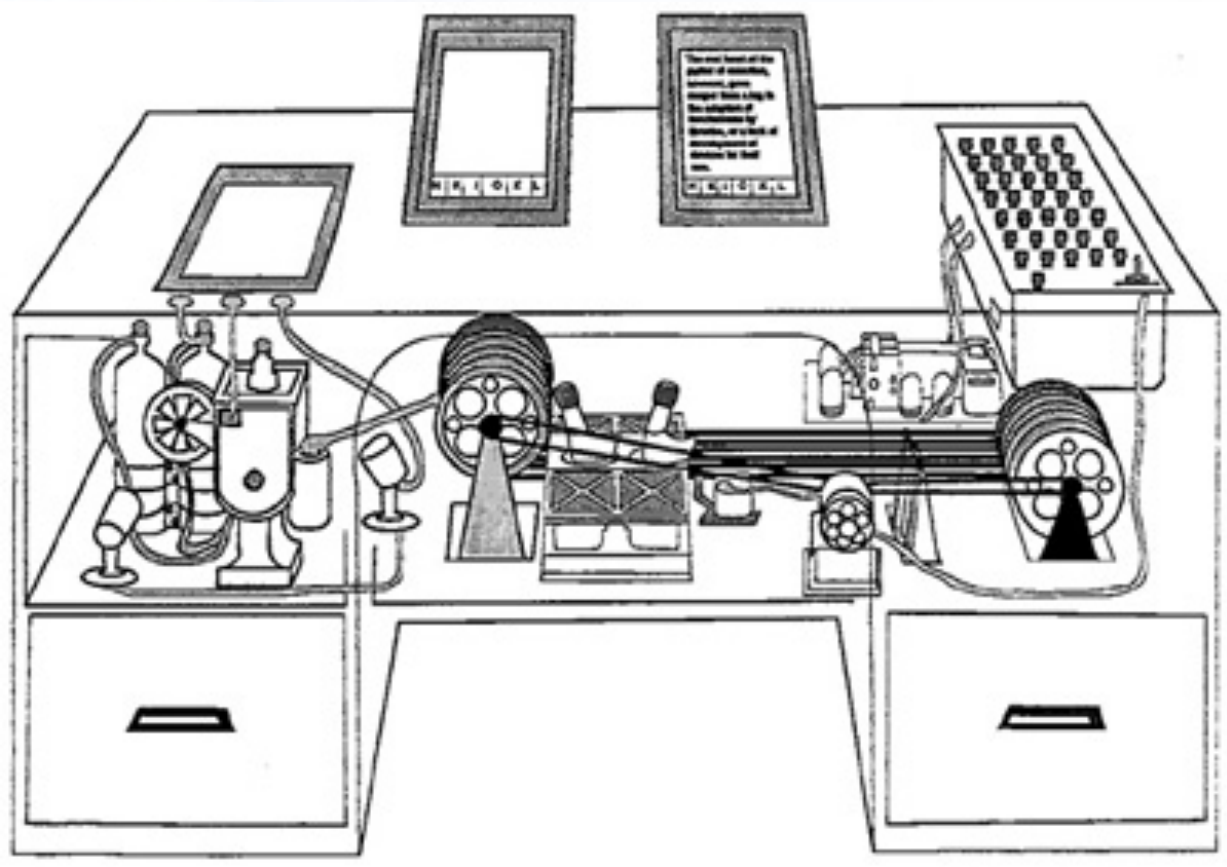
Introduction

- Bush's "As we may think"
  - Writes this at the end of WW II
  - ENIAC was the first computer, born to compute ballistic tables fast
  - Television just invented 5 years ago
  - Photography ("display technology") still a less than perfect process.
  - Microfilm ("storage technology") was a mature and stable technology.

Vannevar Bush (1890-1974)

- Director of the Office of Scientific Research and Development
  - lead 6000 scientists in R&D for WWII
- Predicted many technological advances
  - the “memex” is one whose spirit we are implementing
  - the purpose was to provide scientists the capability to exchange information; to have access to the totality of recorded information

Design for Memex (c. 1945)



Memex

- Integrated computer, keyboard, and desk
- "mechanized private file and library"
  - remove drudgery from information retrieval
  - suggested implementation was microfilm
  - various user operations are suggested
- Associative indexing was the main purpose
  - "the process of tying two items together is the important thing"
  - prelude to hypertext...

Memex

- Information could come pre-associatively indexed, but the key point was user customization
  - WWW still does not provide that today
- Bush observes that tools change our way of doing, and expand the horizons before us
  - full impact of WWW and DLs still not known

What is a Digital Library (DL)?

- "a collection of information that is both digitized and organized" (Lesk)
  - there are numbers of alternate definitions, but this seems fair enough
  - no mention of architecture, implementation, content, etc.
- It is not just to reform the current library system, rather, we aim to
  - organize and access the "information overload"

Outline for today

- Introduction to libraries ✓
- Course administration
- Reading and writing research
- To think about

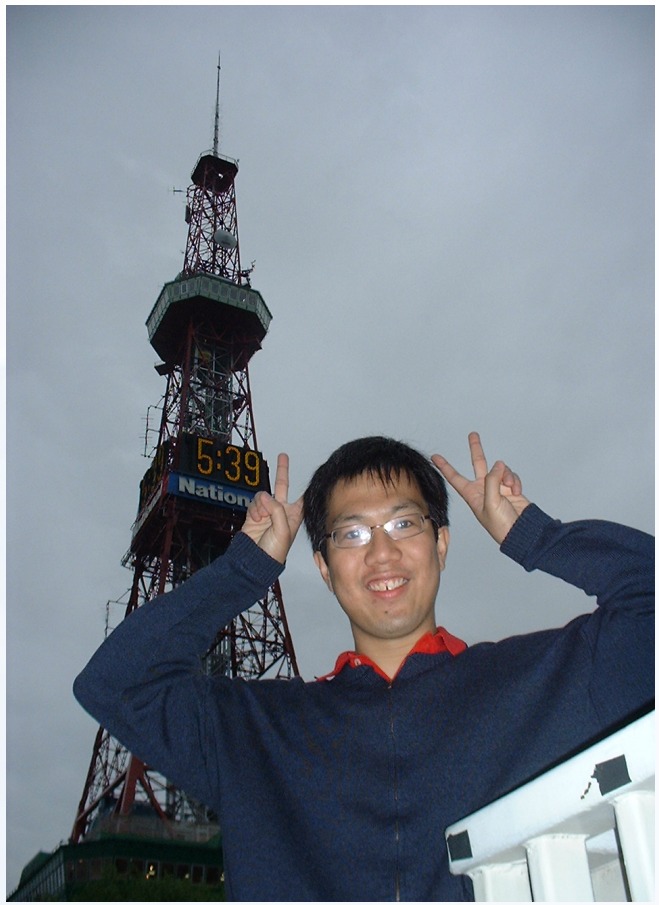
Course administration

- Teaching staff
- Web sites
- Objective
- Syllabus
- Assessment overview
- Survey paper and project

Any questions?

Teaching staff

- Lecturer:
  - Min-Yen Kan (“Min”)
  - kanmy@comp.nus.edu.sg
  - Office: AS6 05-12
    - 6516-1885
  - Hours: 2-4 pm Monday or by appointment
  - Interests:
    - taking care of baby. wifey and digital libraries!



At the TV tower in Sapporo, Japan (2003)

Course web sites

http://ivle.nus.edu.sg/

- Discussion forum
  - **Any questions related to the course should be raised on this forum**
  - **I expect you to talk amongst yourselves to answer questions, so will not answer questions here much.**
  - **Send me emails for urgent or personal matters**
- Announcements!
- Workbin: Lecture notes

http://www.comp.nus.edu.sg/~cs6242

- Grading specification
- Other supplementary content

Objective

- Building, using, presenting and maintaining large volumes of information
- Contrast computational approaches with traditional library science methods

Hey min, go over the website!

> http://www.comp.nus.edu.sg/~cs6242

Discussions

Class participation is very important. There are no “dumb” questions. You will only be penalized for “no” questions / comments.

Possibilities:

- Name tags
- Small group discussion and presentation

Final Exam

- 2 hour final (20%)
  - Calculation questions – that have an exact answer
  - Essay questions – many to look at tradeoffs in the digital library realm
    - **No necessarily right or wrong answers**

Literature survey

- Each student will pick an area of study to survey at least **4** papers in detail.

- Must be **interesting** to you
- Journal or conference papers from an authority list
- Limit to 6 pages
- Individual work only
- Give your perspective on area's future
- Add value by comparing strengths and weaknesses of different approaches.

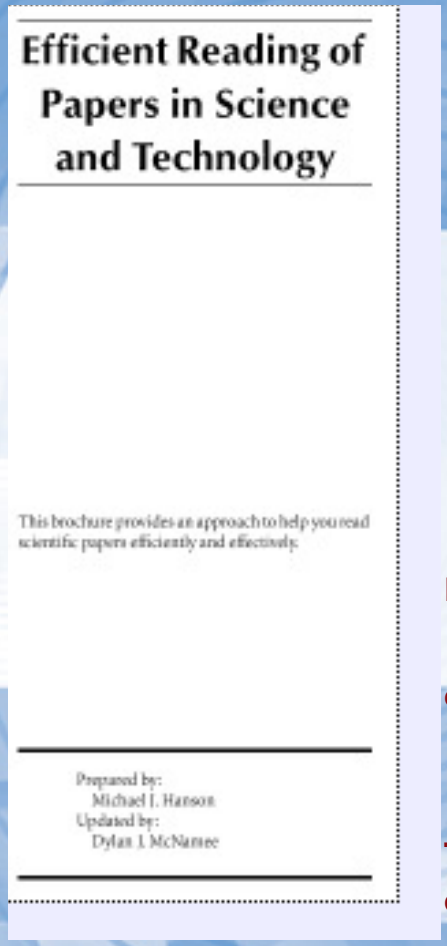
Final project

- Students will self-organize into groups for the final projects, shortly after the survey papers are due.

- Requires **original** work
- Cooperation and coordination
- Report as a conference submission
- Poster presentation to the public
- Sample topics on the web page

Outline for today

- Introduction to libraries ✓
- Course administration ✓
- Reading and writing research
- To think about



Reading and writing research papers

References:

> **http://www.cse.ogi.edu/~dylan/efficientReading.html**

> **ftp://fast.cs.utah.edu/pub/writing-papers.ps**

This section partially from Surendar Chandra of University of Notre Dame.

Why do you read a paper?

- Understand and learn new contributions
- However…
  - Not all papers are “good”
  - Not all papers are “interesting”
  - Not all papers are “worthwhile” for you
- You have to learn to identify a good paper and spend your time wisely
  1. Breadth
  2. Depth
  3. React

Reading a research paper

- What is this paper about?
  1. Read the title and the abstract

     If you still don't know what this paper is about, then this is a poorly-written paper.
  2. Read the conclusion

     Are you now sure you know what this paper is about? If not, throw it away.

3. Read the introduction
4. Read the section headings
5. Read tables and graphs and captions

Do not read a paper linearly!

How to read a paper

- See who wrote it, where it was published, when was it written (credibility)
- Skim references
  - Are authors are aware of relevant related work?
  - Do you know the work that they cite?
  - Do you know other work that they should have cited?

How to read a paper - depth

- Approach with scientific skepticism
- Read with context of other things that you've read in mind
  - It's only one part of the puzzle of a subject
- Examine the **assumptions**. Are they:
  - Reasonable?
  - What are the limitations of the work
    - **There are always limitations! Did they disclose them?**

How to read a paper - depth

- Examine the **methods**:
  - Did they measure what they claim?
  - Can they explain what they observed?
    - **Want an analysis of** why **the system behaves a certain way, not raw data.**
  - Did they have adequate controls?
  - Were tests carried out in a standard way? Were the performance metrics standard?
    - **If not, do they explain their metrics clearly?**

How to read a paper - depth

- Examine the **statistics**:
  "Lies, d*mned lies and statistics"
  - Appropriate statistical tests applied properly?
  - Did they do proper error analysis?
  - Are the results statistically significant?

How to read a paper - depth

- Examine the conclusions:
  - Do the conclusions follow logically from the experiments?
  - What other explanations are there for the observed effects ?
  - What other conclusions or correlations are in the data that were not pointed out?

How to read a paper - react

- Take notes
- Highlight major points
- React to the points in the paper
  - Place this work with your own experience
  - If you doubt a statement, note your objection
- **Summarize** what you read
  - Good practice: maintain your own bibliography of all papers that you ever read
  - Blog this!

How to write a research paper

- Write it such that anyone who reads it using the method we just discussed understands the idea
- Clearly explain what problem you are solving, why it is interesting and how your solution solves this interesting problem
- Be crisp. Explain what your contributions are, what your ideas are and what are others' ideas

Any questions?

Introduction to libraries ✓

Course administration ✓

Reading and writing research ✓

To think about for discussion

- What are the functions of a traditional library?
- Are these same functions in the **digital** library?
- How is the digital library different from:
  - Databases?
  - The WWW?