Course Goals

- Gain an interdisciplinary understanding of search engines and related technologies.
  - How they work
  - How they affect communication
  - How they affect business
  - How they are changing our understanding of information and knowledge.

- Make the techy parts understandable for everyone.
Intro to the Internet & WWW

- Prof. Hearst

Internet Addresses

- The internet is a network on which each computer must have a unique address.
- The Internet uses IP addresses; for example, herald’s IP address is 128.32.226.90.
- Internet Protocol version 4 (IPv4) supports 32-bit dotted quad IP address format:
  - Four sets of numbers, each set ranging from 0 to 255.
  - UC Berkeley’s LAN addresses range from 128.32.0.0 to 128.32.255.255.
- Other addresses in the 5865 LAN include 128.32.226.49.
- Using this setup, there are approximately 4 billion possible unique IP addresses.
- Router software knows how to use the IP addresses to find the target computer.
John Battelle

• The Search

Search and Culture

• The Realization: My God....Google Knows What We Want...
• The Database of Intentions
• Ephemeral to Eternal
• First Use Case: Paid Search
Dr. Jan Pedersen

• The Four Dimensions of Search Quality

- Problem:
  - Ensure that what is indexed correctly reflects current state of the web
- Impossible to achieve exactly
  - Revisit vs Discovery
- Divide and Conquer
  - A few pages change continually
  - Most pages are relatively static
Dr. Dan Rose

- User Experience Issues in Web Search
Dr. Peter Norvig

- Google News, Print, Maps, & Earth
Dr. Sep Kamvar

- Personalization and Search
Dr. John Chuang

- Peer-to-peer Search
Sergey Brin

- Google and Life
Dr. Hal Varian

- Search advertising

Factors affecting revenue

\[
\text{Monetization (RPM)} = \frac{\text{Revenue}}{\text{Queries}} \times 1k
\]

\[
= \frac{\text{Revenue}}{\text{Clicks}} \times \frac{\text{Clicks}}{\text{Queries}}
\]

\[
= \frac{\text{Revenue}}{\text{Clicks}} \times \frac{\text{Queries w/ Ads}}{\text{Queries}} \times \frac{\text{Ads}}{\text{Queries w/ Ads}} \times \frac{\text{Clicks}}{\text{Ads}}
\]

\[
= \frac{\text{CPC}}{\text{Coverage}} \times \frac{\text{Depth}}{\text{CTR per Ad}}
\]

Price  Quantity  Quality
Jason Schultz

- Search and Intellectual Property

Are we being diverted or informed?
Dr. Sue Dumais

- Desktop Search
Dr. Mark Najork

- Web Spam

Examples of synthetic content

- Monetization
- Random words
- Well-formed sentences stitched together
- Links to keep crawlers going
Dr. Doug Tygar

- Search and Privacy
Dr. Alma Whitten

- Search Engine Architecture
Dr. Eric Brewer

- Reflections on Starting a Search Engine Company
Avi Rappaport

- Enterprise Search

Choose, Implement, Maintain

- Buy or download, don't build
  - Quality search and other features are non-trivial
  - Homegrown systems rarely satisfy

- Effort to implement varies wildly
  - Number of documents, complexity, interfaces
  - Resources, servers, network
  - Enterprise information needs

- Maintenance
  - Keep system working, index current, scale up
  - Add new data sources
  - Change as new needs appear
  - Log analysis
Jennifer English

- Enterprise Search

Can the user find the search box?

- Inconsistent placement.
- Size – text box usually too small to accommodate a reasonable query.
- Wording around search box/button is inconsistent.
- Text entry boxes are “ugly” – designers want them to be as small and inconspicuous as possible.
Bradley Horowitz

- Mass Media and Micro Media Search at Yahoo
Mary Hodder

- Searching the Live Web

The Blogosphere: what are you searching?
Dr. Geoff Nunberg

• How Search Shapes Cyberspace

Formal Correlates of Metrical Space:
The Book as a Public Presence

The newspaper reader, observing exact replicas of his own paper being consumed by his subway, barbershop, or residential neighbors, is continually reassured that the imagined world is visibly rooted in everyday life... creating that remarkable confidence of community in anonymity which is the hallmark of modern nations. —Benedict Anderson, Imagined Communities.
Dr. Marti Hearst

- Faceted Metadata in Search Interfaces
What is the Future of Search?
Administrivia
Final Projects

• Turn them in using online links

• HARD DEADLINE!

• Undergrads: due Friday Dec 9, 7pm
  http://www.sims.berkeley.edu/courses/is141/f05/discussion.html

• Grads: due Saturday Dec 10, 5pm
  http://www.sims.berkeley.edu/courses/is290-2/f05/assignments.html
Course Evaluations

- This is the SIMS form
- First page is instructor evaluation
- Back of page is course evaluation
- PLEASE CIRCLE THE CLASS YOU ARE IN AT THE TOP
  - 141 vs 290-2
- Instructor does NOT see these until after she turns in the grades.
- Turn in the form to a TA, who will then check you off for attendance.
  - TAs will not accept forms until 10 minutes after they are distributed.
Let’s Thank Our TAs!

Helen Kim and Fredrik Wallenberg
Thank you!

And Happy Searching!