

Search Engines: Technology, Society, and Business

Course Summary

Marti Hearst

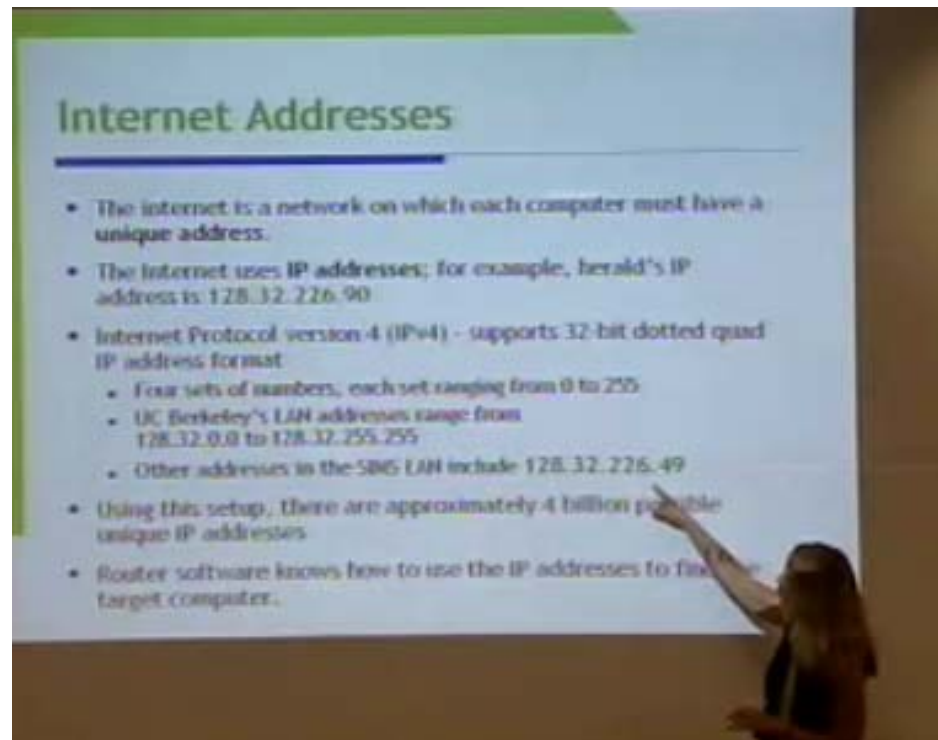
December 10, 2007

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



Dr. Jan Pedersen

- The Four Dimensions of Search Quality



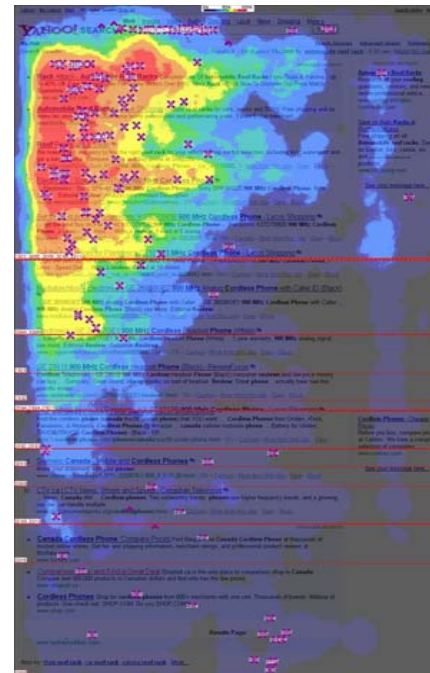
 Freshness

- 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

YAHOO!

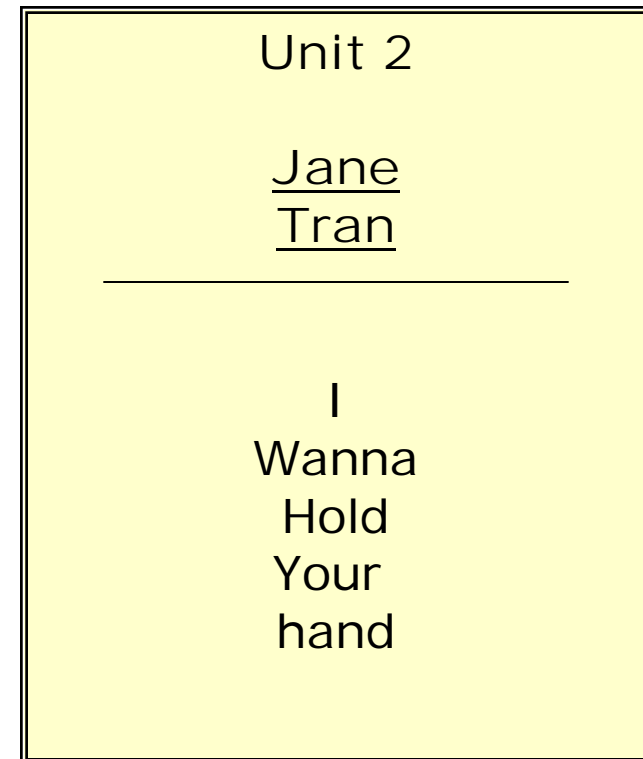
Dr. Daniel Russell

- User Experience Issues in Web Search



Class Exercise

- Students as web pages and a search engine
- Web pages:
 - Web site = where you live
 - Hyperlinks = who you know in class
 - Web page = Beatle's song title



Dr. Hal Varian

- Search advertising



Factors affecting revenue

$$\begin{aligned} \text{Monetization (RPM)} &= \frac{\text{Revenue}}{\text{Queries}} \times (1K) \\ &= \frac{\text{Revenue}}{\text{Clicks}} \times \frac{\text{Clicks}}{\text{Queries}} \\ &= \frac{\text{Revenue}}{\text{Clicks}} \times \frac{\text{Queries w/ Ads}}{\text{Queries}} \times \frac{\text{Ads}}{\text{Queries w/ Ads}} \times \frac{\text{Clicks}}{\text{Ads}} \\ &= \underbrace{\text{CPC}}_{\text{Price}} \times \underbrace{\text{Coverage} \times \text{Depth}}_{\text{Quantity}} \times \underbrace{\text{CTR per Ad}}_{\text{Quality}} \end{aligned}$$

Dr. Mark Najork

- Web Spam



Examples of synthetic content



Monetization

Random words


Well-formed sentences stitched together

Links to keep crawlers going

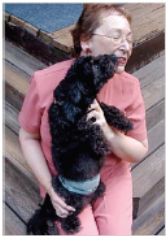
Chris Hoofnagle

- Privacy and Online Information



Search Strings 

- | AOL releases 20m queries based on 600k users to help researchers
 - Were trying to make routine access more efficient
- | Users are uniquely enumerated
- | Some easy to identify
 - Users vanity searched name, SSN
- | Many others identifiable based on searches unrelated to PII
 - Thelma Arnold, Lilburn, Ga



"Privacy & Search Engines," Hoofnagle Oct 21, 2007 8

Dr. Lynn Wilcox

- Multimedia Search



Jason Schultz

- Search and Intellectual Property



Are we being diverted or informed?

A collage of various consumer products including snacks, drinks, and household items. The products shown include brands like Oreo, Kit Kat, Quives, Carmen Crisp, Skippy, Chic Choc, Success Rice, Jumbo Pop, Bugles, and others. The collage is arranged in a grid-like fashion on a black background.

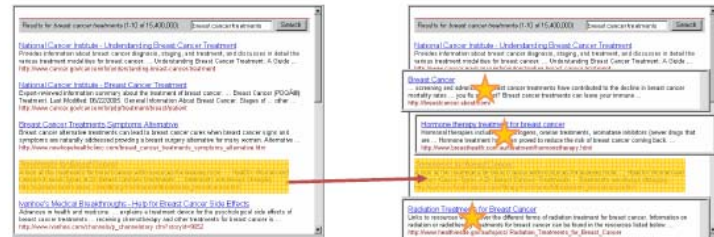
 Electronic Frontier Foundation

Dr. Jaime Teevan

- Personalization and Search



Ranking Results for Re-Finding



Natural Language Processing

- Prof. Hearst

How can a machine understand these differences?

- Get the cat with the gloves.



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

What was most surprising?

Final Projects

- Turn them in using online link
- <http://courses.ischool.berkeley.edu/i141/f07/assignments/projects.html>
- **HARD DEADLINE!**
- Due Sunday Dec 16, 9am
- Feel free to turn yours in early!

Course Evaluations

- This is the iSchool form
 - First page is instructor evaluation
 - Back of page is course evaluation
- I've also added another sheet with a few specific questions about future course format.
- 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!

Eun Kyoung Choe and Ani Sen

Thank you!

And Happy Searching!