User Experience Issues in Web Search

Daniel E. Rose

September 19, 2005

Starting Assumption

The goal of a search engine is
to satisfy the user's information need

(In other words, to help the user with an information-seeking task.)

Corollary

The more info the user gives the system about his or her information need, the better job the system can do to satisfy it.

What do we know about information-seeking?

- *It's an iterative process*

What's the cure for AIDS?

AIDS is caused by HIV

What's the cure for HIV?

No cure, but treatments…

Best treatments for HIV?

What do we know about information-seeking?

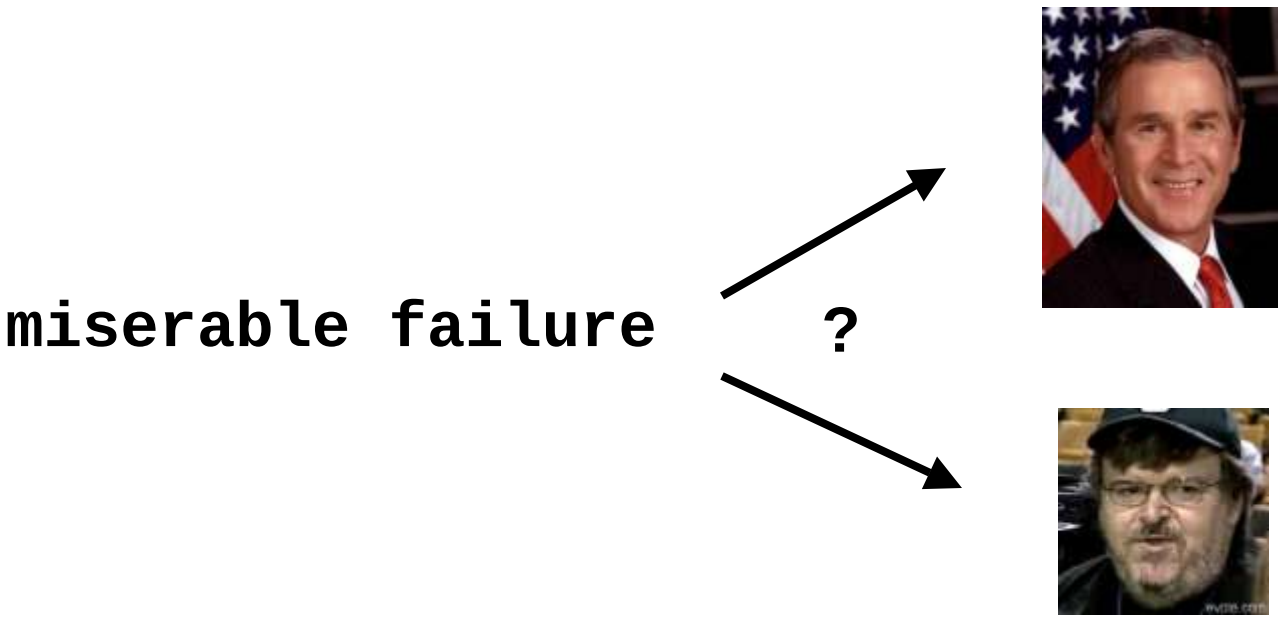
- Bates' berrypicking model:

  "Each new piece of information [users] encounter gives them new ideas and directions to follow, and, consequently, a new conception of the query."

Bates, M.J. (1989) "The Design of Browsing and Berrypicking Techniques for the Online Search Interface," *Online Review*, 13(5):407-24.

What do we know about information-seeking?

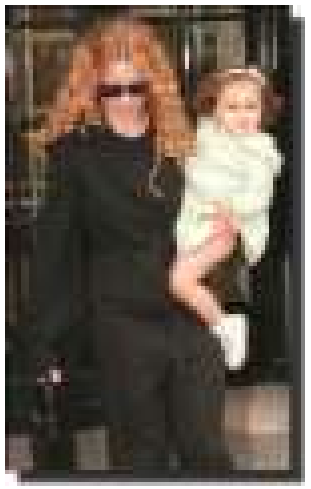
- *It is extremely subjective*



YAHOO!

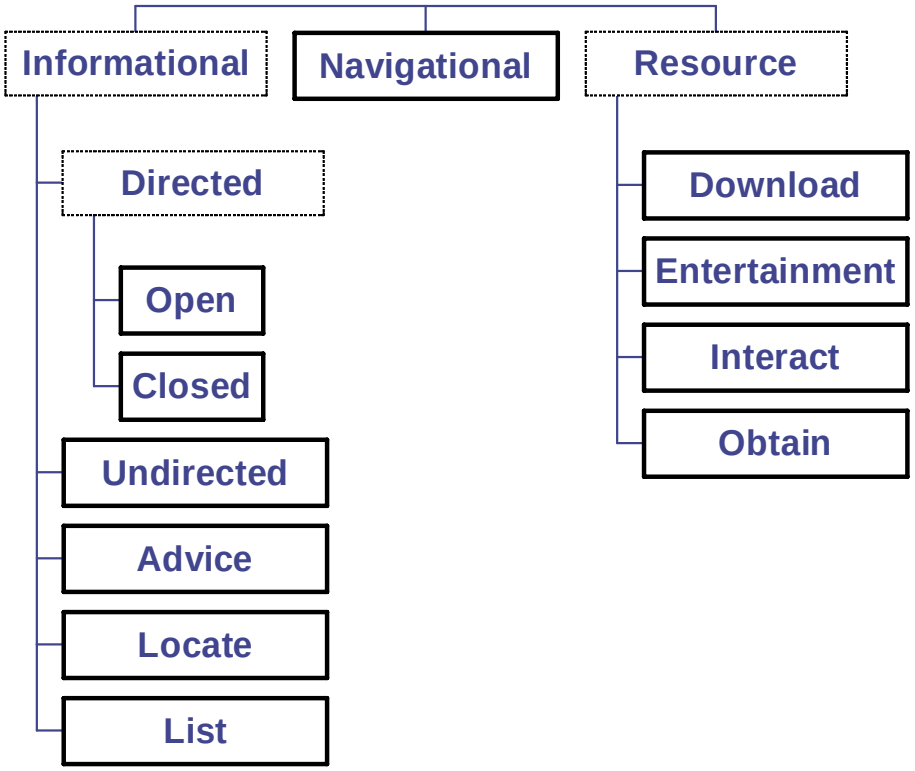
What do we know about information-seeking?

- *It depends on social and cultural context*
  - "pants" in UK vs. US
  - "madonna and child" for
    - art historian
    - pop music fan



What do we know about information-seeking?

- *It depends on your goal*



Rose, D.E. and Levinson D. (2004) "Understanding User Goals in Web Search," Proceedings of the 2004 World Wide Web Conference.

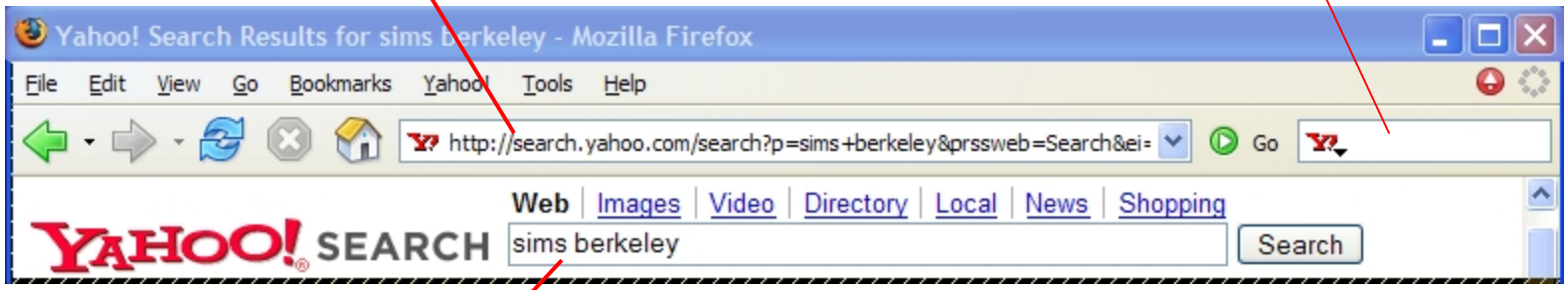
YAHOO!

User Input: Queries

Some Terminology

address bar

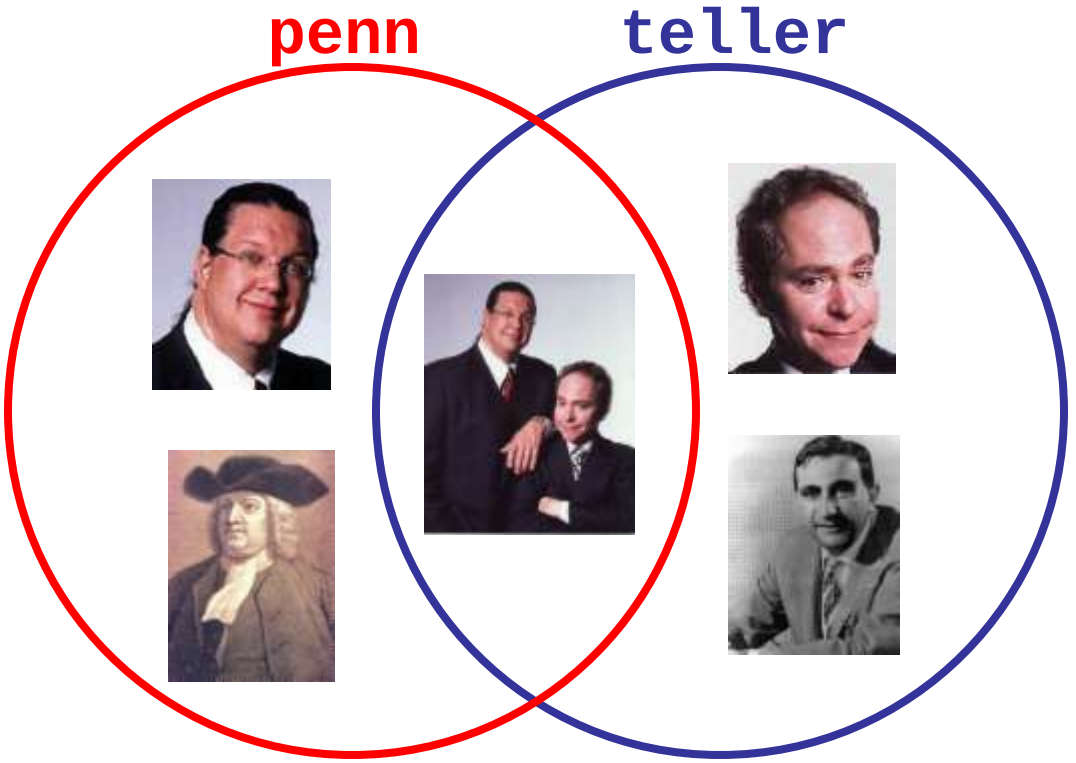
built-in query box
(some browsers)



search engine
query box

YAHOO!

More Terminology: Boolean Queries



penn OR teller

penn AND teller

People Don't Understand Boolean Logic!

- Lots of studies, esp. Wason's selection task.
- "If I search for cats AND dogs, I'll get all the pages about cats and all the pages about dogs, right?"

(Wrong!)

Wason, P.C. (1966) "Reasoning" in B.M. Foss (ed.) *New Horizons in Psychology I.*

An Old-Style Structured Query

```
assum! /5 risk
/p ic* snow*** snowfall
/s slip! fell fall***
```

WESTLAW Reference Manual, 3rd Edition, 1989.

Alternative: Best Match

- Also known as
  - partial match
  - Natural Language
  - Scored OR
- The more the pages match my query, the higher they'll rank – even if they don't include ALL the words.
- Featured in most general-purpose search tools in the early 90s (Verity Search 97, Applesearch) and most original web search engines (Infoseek, AltaVista, Excite, etc.) until about 1998.

Best Match Example

Query: **`teak outdoor patio furniture`**

Results:

1. “Patio World contains outdoor furniture in teak and redwood”
2. “Teak Warehouse – beautiful indoor and outdoor furniture.”
3. “Outdoor Furniture Showroom – patio chairs and benches in a variety of materials”
4. “Save the rainforest – don’t buy teak!”

How most web search engines interpret queries today

- Query words have implicit Boolean AND
- Quoted words must appear adjacent
- "-" before a word means "exclude pages containing that word"

User Behavior

What Are Users Doing?

- Not typing many words
  - Average query was 2.6 words long (in 2001), up from 2.4 words in 1997
- Moving toward e-commerce
  - less sex (down from 17% to 9%), more business (up from 13% to 25%)

Spink A., et al. “From E-Sex to E-Commerce: Web Search Changes,”*Computer*, March 2002.

Why are queries so short?

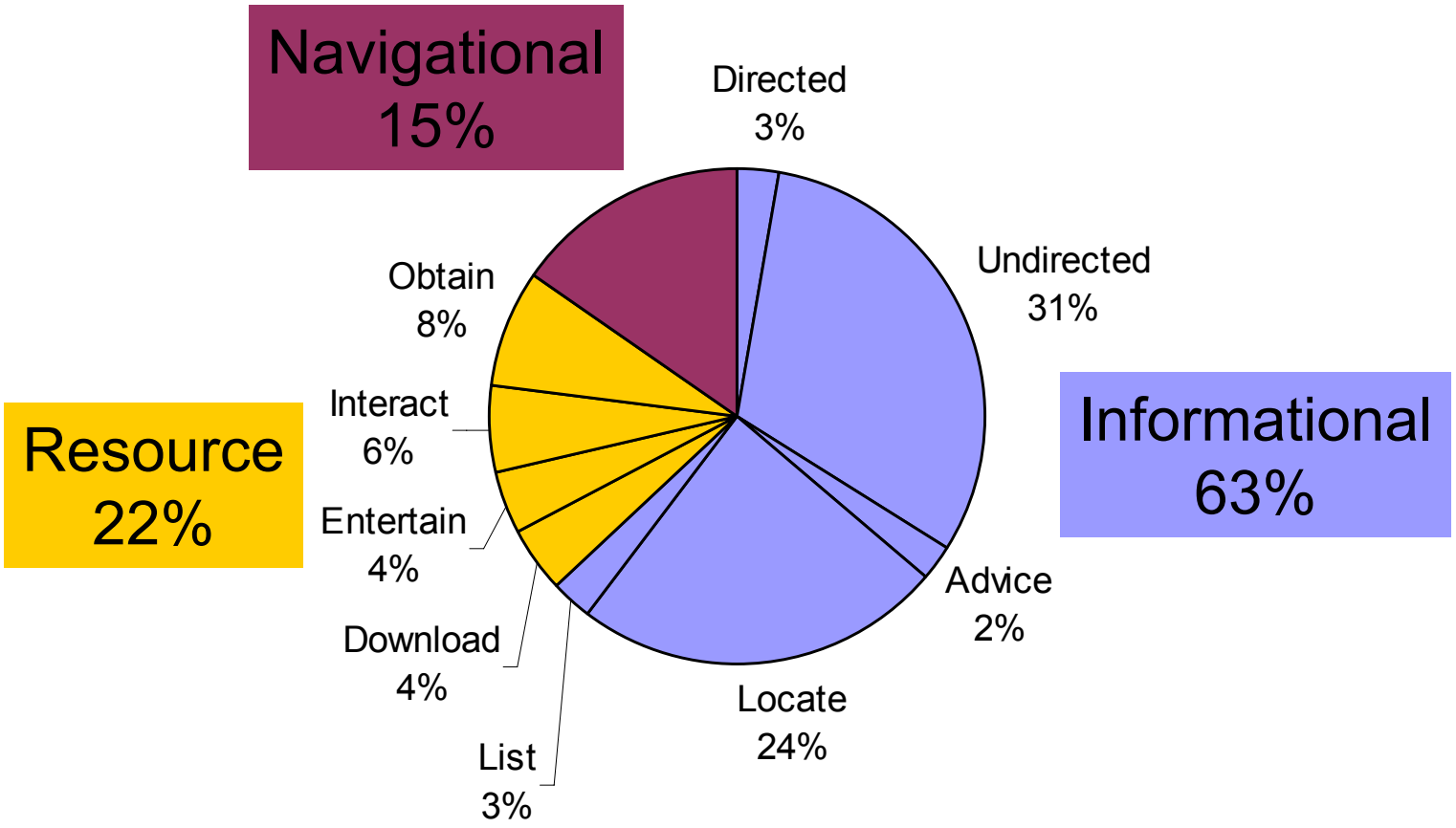
- Several possible reasons:
  - Users minimize effort
  - Users don't realize more info is better
  - Users learn that too many words means too few results (since implicit Boolean AND)
  - Query boxes are small

Does (query box) size matter?

- Yes! Belkin research:
  - A query box that held 5 lines of text yielded longer queries
  - Different instructions (describe information problem) yielded longer queries
- But, in a non-web, best match system

Belkin, N.J., et al. "Rutgers' TREC 2001 Interactive Track Experience," in Voorhees & Harmon, *The Tenth Text Retrieval Conference.*

Why are users searching?



Rose, D.E. and Levinson, D. (2004), "Understanding User Goals in Web Search";
Broder, A.(2002), "A Taxonomy of Web Search," SIGIR Forum 36(2).

YAHOO!

Do users know how to search?

- Hargittai has many examples of confusion:
  - Confusion about address bar:
    - **`my wallet has been stolen.com`**
    - **`www.new york times.com`**
  - Queries without spaces
    - **`presidentialcampaign2000`**
    - **`Princetonhistoricalsocietyvolunteer`**

"In email and Web addresses there are no spaces, so I tend not to use them in searches either."

Hargittai, E. (2004) "Classifying and Coding Online Actions," *Social Science Computer Review.* 22(2):210-227.

Do users know how to search?

- Quotation marks:
  - Only 16% of participants used quotation marks, many incorrectly
  - Some quoted single terms
  - Some quoted all terms for all queries
- Other operators?
  - **`lactose intolerance -recipes`**

Hargittai, E. (2004) "Classifying and Coding Online Actions," *Social Science Computer Review.* 22(2):210-227.

Novices vs. Experts

- Bhavnani
  - Looked at differences between expert and novice searchers doing medical search task:
    - Experts visited 3 sites, took 7 minutes, got all the info.
    - Novices used only Google, visited 13 sites, took 20 minutes, and missed lots of info.
- Hargittai
  - *"The ability to draw on a range of strategies and the agility to switch between them easily seems to be a key ingredient to successful and efficient Web navigation."*

Bhavnani, S. (2003), "Strategy Hubs: Next Generation Domain Portals With Search Procedures," *Proceedings of CHI 2003*.
Hargittai, E. (2004). "What Makes An Expert Searcher? Evidence from User Studies," *2004 World Wide Web Conference*.

Search is hard!

Vocabulary Problem

- People use different words for the same thing.
  - <20% chance of choosing same word
  - Even "best" word has 65-85% failure

"…The data show that no single access word, however well chosen, can be expected to cover more than a small proportion of users' attempts…"

Furnas, G.W., et al. (1987). "The Vocabulary Problem in Human-System Communication," *Communications of the ACM*, 30(11): 964-971.

The Vocabulary Problem in Web Search

- Two people are unlikely to use the same word to describe the same thing…
- So, a web page author is unlikely to have used the same words as the user searching for the page's content
- But anchortext helps

Recognition vs. Recall

- Terminology:
  - Recall: Remembering something without seeing it.
  - Recognition: Identifying something that exists in memory.
- People are better at recognition than recall
  - Menus vs. Commands

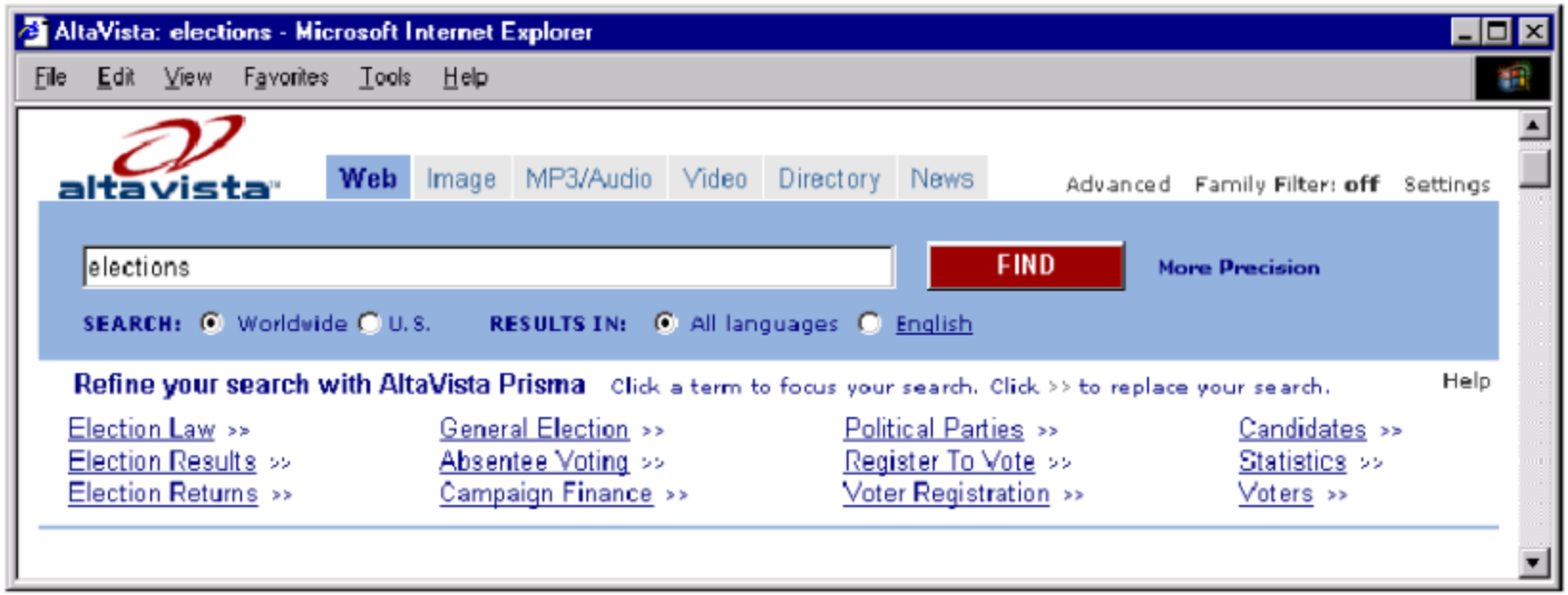
Klatzky, R.L. (1980). *Human Memory:Structures and Processes, 2nd ed*.

Recognition vs. Recall in Web Search

- Users may not remember the correct term, but could select it if they saw it.

One Possible Solution: Offering Document Concepts

- AltaVista Prisma

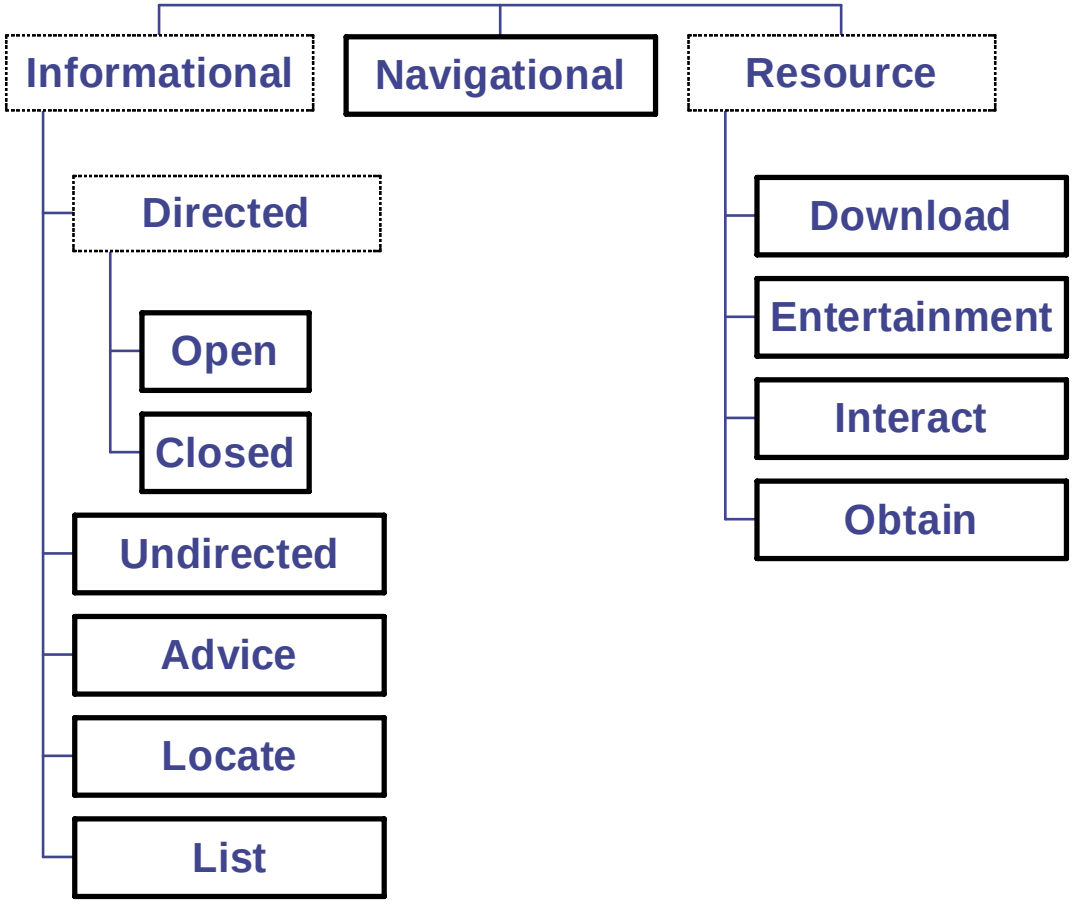


Anick, P. (2003). "Using Terminological Feedback for Web Search Refinement: A Log-based Study," *Proceedings of SIGIR 2003*.

YAHOO!

Web search is even harder!

Why harder?
All those different goals…



YAHOO!

Why Harder?
Web Corpus is Different

- Heterogeneous (in many ways)
  - Format, length, genre, authority, quality…
- Stability
  - Ongoing growth of content volume
  - Ongoing "linkrot" (2 year half-life)
  - Content change in same URL
  - *Don't know whether info is there*

Why Harder? Spammers

- Misleading content

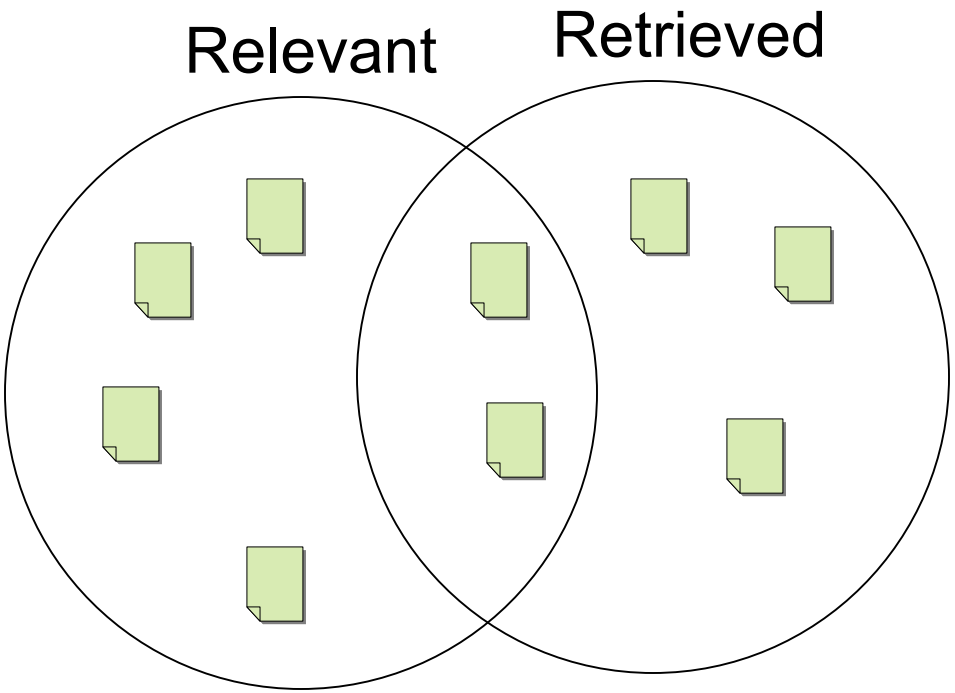
  Repeating terms, adding competitor's products, adding unrelated terms, cloaking…

- Misleading links

  Multiple domains, link farms, guestbook bots…

How to measure success of search user experience?

Still More Terminology



$$\text{RECALL} = \frac{|\,\text{Relevant} \cap \text{Retrieved}\,|}{\text{Relevant}}$$

$$\text{PRECISION} = \frac{|\,\text{Relevant} \cap \text{Retrieved}\,|}{\text{Retrieved}}$$

Measure relevance?

- Recall
  - Breaks down in shift from thousands to billions
  - Result set now bigger than entire TREC corpus
- Precision
  - What user goals are we trying to satisfy?
  - Harder to assess heterogeneous results
    - » Pages that link to good pages?
    - » Pages that had good content when indexed, but no longer?

Monitor user actions?

- What's the right number of clicks for
  - Navigational query?
  - Research?
  - Question?
  - Entertainment?
- Problem of unexpected content

Observe users?

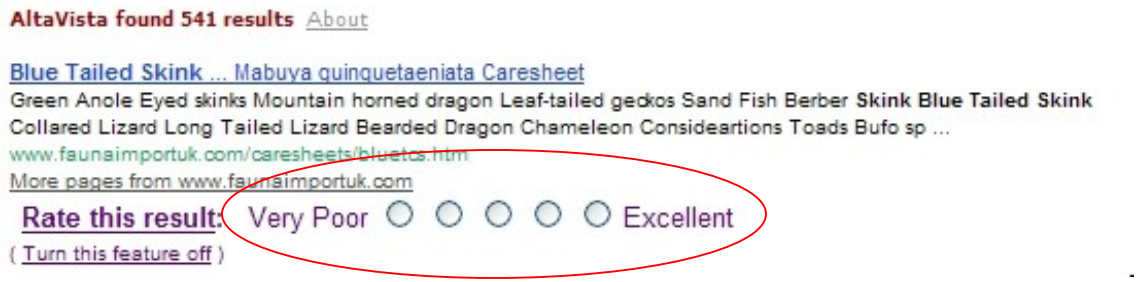
- Time on Task
  - Exploration/quality vs. speed
    - What if task is open-ended?
    - What if it's fun?

Ask users (through surveys)?

- Feedback from existing users

  Popup & inline rating forms.

  Incentive?



- Paying users

  Wrong incentive!

  (Discarded 55% in one AltaVista test.)

YAHOO!

Evolution of the Search UI?

AltaVista 1995

AltaVista | Advanced Search | Simple Search | Surprise | Help

Search the Web and Display the Results in Standard Form

"Fuzzy Logic" Submit

Tip: Do not use AND or OR to combine words, simply type a few words or phrases.

Word count: Fuzzy Logic: about 10000

Documents 1-10 of about 9000 matching some of the query terms, best matches first.

Hierarchical Fuzzy Logic Control of a Double Inverted Pendulum
Hierarchical Fuzzy Logic Control of a Double Inverted Pendulum. Reza Langari and Shuliang Lei Center for Fuzzy Logic and Intelligent Systems Research...
http://celebris.tamu.edu/project/dblpend/dip.htm - size 6K - 21 Nov 95

Project Management with Fuzzy Logic
Edit a Project. Please enter details of the project. For more information, please select help. Project ID: Project Name: Priority: Start Date: Duration:...
http://www.cs.bris.ac.uk/~wong/Proj/upproj.html - size 1K - 13 Oct 95

Project Management with Fuzzy Logic
Create a New Project. Please enter details of the project. For more information, please select help. Project ID: Project Name: Priority: Start Date:...
http://www.cs.bris.ac.uk/~wong/Proj/newproj.html - size 1K - 28 Sep 95

Project Management with Fuzzy Logic
Help. This program is design to give an accurate estimation of project risk analysis before and during the different phases of a project. It also allows...
http://www.cs.bris.ac.uk/~wong/Proj/help.html - size 3K - 13 Oct 95

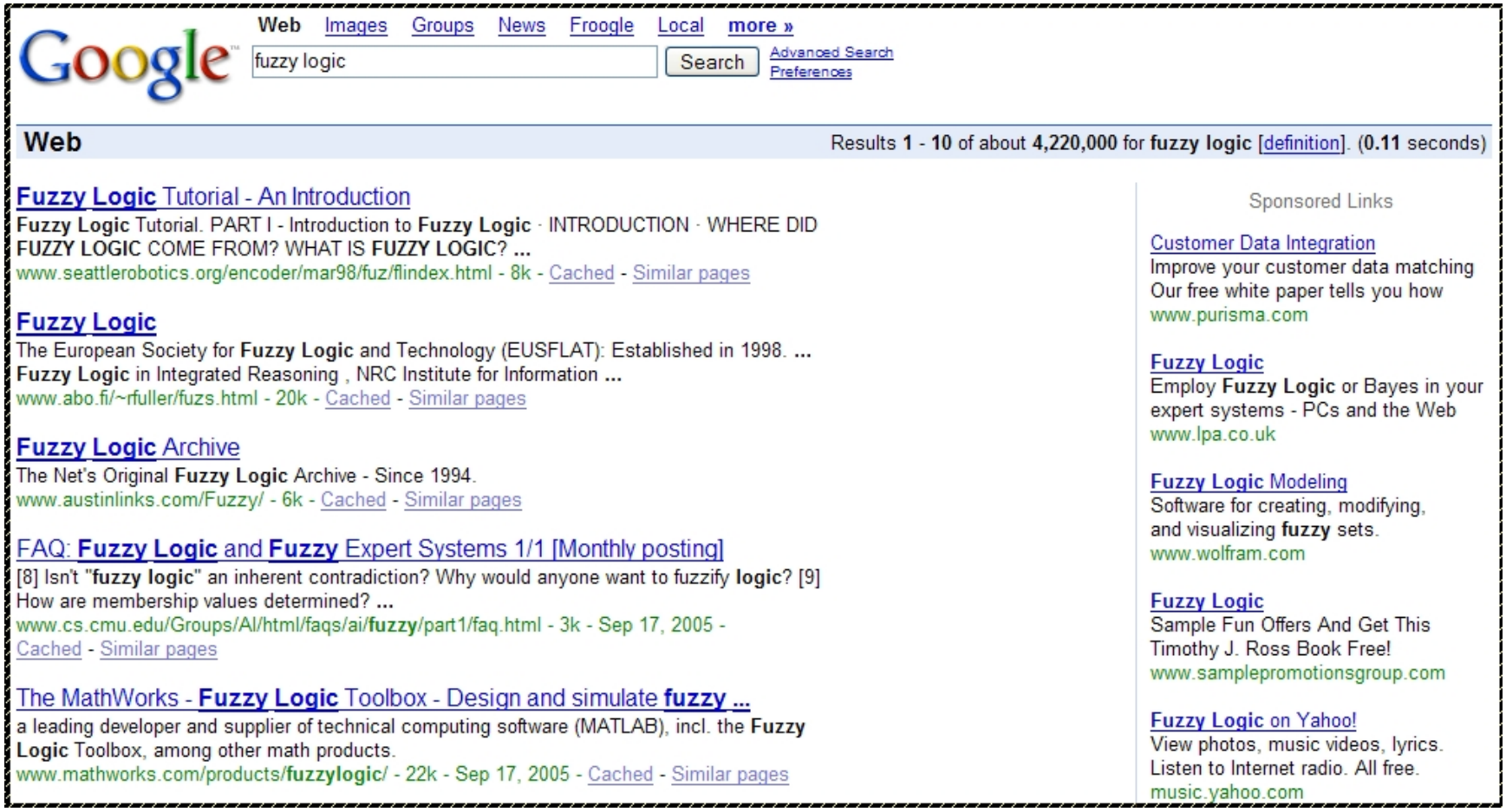
CENTRAL NOTICE - Fuzzy Logic
Apologies to fuzzy logicians. In a fuzzy world, you might only "partially" know stuff. So there would be much information that is neither...
http://www.notice.com/whyfuzzy.html - size 1K - 1 Apr 96

Archimedes Fuzzy Logic Development Tool
Archimedes Fuzzy Logic Development Tool. This product introduces many new functions for Fuzzy development: A graphical CAD surface to design the system...
http://www.mwmedia.com/tpvs/archimed/mtd/archfuzy.htm - size 1K - 19 Jul 95

Machine Learning Group, Neural Networks and Fuzzy Logic

YAHOO!

Google 2005



Google

Web Images Groups News Froogle Local more »

fuzzy logic Search Advanced Search Preferences

Web

Results **1** - **10** of about **4,220,000** for **fuzzy logic** [definition]. (**0.11** seconds)

Fuzzy Logic Tutorial - An Introduction
Fuzzy Logic Tutorial. PART I - Introduction to **Fuzzy Logic** · INTRODUCTION · WHERE DID **FUZZY LOGIC** COME FROM? WHAT IS **FUZZY LOGIC**? ...
www.seattlerobotics.org/encoder/mar98/fuz/flindex.html - 8k - Cached - Similar pages

Fuzzy Logic
The European Society for **Fuzzy Logic** and Technology (EUSFLAT): Established in 1998. ... **Fuzzy Logic** in Integrated Reasoning , NRC Institute for Information ...
www.abo.fi/~rfuller/fuzs.html - 20k - Cached - Similar pages

Fuzzy Logic Archive
The Net's Original **Fuzzy Logic** Archive - Since 1994.
www.austinlinks.com/Fuzzy/ - 6k - Cached - Similar pages

FAQ: **Fuzzy Logic** and **Fuzzy** Expert Systems 1/1 [Monthly posting]
[8] Isn't "**fuzzy logic**" an inherent contradiction? Why would anyone want to fuzzify **logic**? [9] How are membership values determined? ...
www.cs.cmu.edu/Groups/AI/html/faqs/ai/**fuzzy**/part1/faq.html - 3k - Sep 17, 2005 - Cached - Similar pages

The MathWorks - **Fuzzy Logic** Toolbox - Design and simulate **fuzzy** ...
a leading developer and supplier of technical computing software (MATLAB), incl. the **Fuzzy Logic** Toolbox, among other math products.
www.mathworks.com/products/**fuzzylogic**/ - 22k - Sep 17, 2005 - Cached - Similar pages

Sponsored Links

Customer Data Integration
Improve your customer data matching
Our free white paper tells you how
www.purisma.com

Fuzzy Logic
Employ **Fuzzy Logic** or Bayes in your expert systems - PCs and the Web
www.lpa.co.uk

Fuzzy Logic Modeling
Software for creating, modifying, and visualizing **fuzzy** sets.
www.wolfram.com

Fuzzy Logic
Sample Fun Offers And Get This Timothy J. Ross Book Free!
www.samplepromotionsgroup.com

Fuzzy Logic on Yahoo!
View photos, music videos, lyrics. Listen to Internet radio. All free.
music.yahoo.com

YAHOO!

Innovation?

YAHOO!

Instant Search

http://instant.search.yahoo.com

YAHOO!

Conclusion

The web search user experience does not reflect what we know about user behavior.


Questions?