Search Engine Basics

How do people search for information?

- Using a Search Engine
  - Put your search terms and hope for the best.
- Direct navigation
  - Enter the URL directly into the browser.
- Navigation within a Directory
  - Use a web portal as an entry point to the web.
- Information seeking on the web is not straightforward and people use a combination of techniques but turn often to search engines.

Web Search is not simply IR

- The scale of web search is way beyond traditional IR.
- The web is very dynamic.
- The web contains an enormous amount of duplication.
- The quality of web pages is not uniform.
- The range of topics on the web is open.
- The web is globally distributed.
- Users typical habits are different (short 2-3 words queries, inspect only top-10 pages).
- The web is hypertextual.

From Information Need to Answers

- Let’s say you discovered mice at home. Problem!
  - You would like to eliminate the problem.
  - How would you search for solutions?
- Formulate the query terms:
- You are a nice person, you would like to get rid of them in a politically correct way.
  - What changes in your search terms?
The classic search model

- User task
  - Get rid of mice in a politically correct way
- Info need
  - Misconception?
  - Info about removing mice without killing them
- Misformulation?
- Query
  - how trap mice alive
- Search engine
  - Search
- Results
- Collection

Types of Needs for Web Search

- Corpus: The publicly accessible Web
- Need: Retrieve high quality results relevant to the user’s need.
  - Characterize the particular need:
- Types of need
  - Informational – want to learn about something
  - Navigational – want to go to that page
  - Transactional – want to do something (web-mediated)
    - Access a service
    - Downloads
    - Shop
    - Gray areas
      - Find a good hub
      - Exploratory search “see what’s there”
- Examples:
  - Low hemoglobin
  - United Airlines
  - Tampere weather
  - Mars surface images
  - Nikon CoolPix
  - Car rental Finland
  - Abortion morality

Categories of Web searchers

Broder: A taxonomy of web search

- Informational _____ %
  - acquire some information about a topic from web pages.
- Navigational _____ %
  - find a site to start navigation from.
- Transactional _____ %
  - perform some activity mediated by a web site.

Think of your own searches. Do you agree?
How did Broder found out these categories?
How did he measure the percentages?

Search Engines: The players and the field

- The mechanics of a typical search.
- The search engine wars.
- Statistics from search engine logs.
- The architecture of a search engine.
- The query engine.
Mechanics: Results & ads come ranked

Search Engine Wars

- The battle for domination of the web search space!
- The competition is good news for users!
- Crucial: advertising is combined with search results!
- What if one of the search engines will manage to dominate the space?

Yahoo!

- Synonymous with the .com boom, once the best known brand on the web.
- Has very strong advertising and e-commerce partners

Lycos

- One of the pioneers of the field
- Introduced innovations in 1996 that inspired the creation of Google
To “Google” is synonymous with Web searching.

Has raised the bar on search quality

The most popular SE in the last few years.

Is innovative and dynamic.

Google

Google comp.sen-chess

Google

Ask Jeeves

Specializes in natural language question answering.

Search driven by Teoma.

Tries to differ...

Ask Jeeves

Cuil

Newer player

Claimed to have indexed 120B pages!

It did not rank!

bing (was: Live Search (was: MSN Search))

Successful third reincarnation of previous attempts

Was Synonymous with PC software.

Pyrhich victory in the browser wars with Netscape.

“Stop searching, start deciding”: Turned Google into copycat!
Other "search engines"...

How do you decide which is best?

How do you measure similarity in ranking?

How many people use the web? SEs?

Search Engines as Info Gatekeepers

- Search engines are becoming the primary entry point for discovering web pages.
- Ranking of web pages influences which pages users will view.
- Exclusion of a site from search engines will cut off the site from its intended audience.
- The privacy policy of a search engine is important.

Introna & Nissenbaum: Defining the Web: The Politics of Search Engines
Hindman et al: Googlearchy: How a few Heavily-Linked Sites Dominate Politics on the Web
Architecture of a Search Engine

The Web

Web spider

Indexer

Search

Indexes

Ad indexes

Google

Newer features: suggest

N-grams

Google trends

Google books Ngram Viewer