
Searching the Web

Don't we already know how to do this?

2/14/05

CS120 The Information Era

1

What's the Deal?

- Searching the World Wide Web is something that Internet users do almost every day
- How effective are people at searching the web?
- How effective are YOU at searching the web?
- What sort of things do you search for online, and what tools do you use?

2/14/05

CS120 The Information Era

2

Let's See!

- Find out why zebras have never been domesticated. Is it impossible to train a zebra, is it just too much trouble, or is there some other reason?

2/14/05

CS120 The Information Era

3

Search Tools

- There are four main tools that can be used to help the searching process:
 1. Subject trees (directories)
 2. Clearinghouse
 3. General search engine
 4. Specialized search engine
- Let's examine each of these

2/14/05

CS120 The Information Era

4

1. Subject Trees (Directories)

- Category of topics
- Organized and maintained by humans
- Examples:
 - <http://dir.yahoo.com/>
 - <http://www.google.com/dirhp>
 - <http://www.about.com/>
- What do you think are some of the problems with subject trees?

2/14/05

CS120 The Information Era

5

1. Subject Trees

- Exercise:
 - List all the different clocks you can find in any of the subject trees listed on the previous slide
 - You have five minutes!
- For a comparison of various subject trees see
 - <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SubjDirectories.html>

2/14/05

CS120 The Information Era

6

2. Clearinghouse

- Collection of websites about a specific topic
- Have a small focus
- Maintained by humans
- The best way to find a clearinghouse on a specific topic is to use a clearinghouse index. An example is:
 - <http://www.winsor.edu/pages/sitepage.cfm?id=304>

2/14/05

CS120 The Information Era

7

3. General Search Engine

- How many search engines are there out there?
- Which are the main ones?

2/14/05

CS120 The Information Era

8

How do Search Engines Work?

- Spider visits webpages and follows the links on them
- Depending on the search engine, it either looks at specific text elements (such as headers, keywords, and/or the first several hundred words), or looks at all the text
- The data is stored in a database
- The data is then stored in a database and ranked

2/14/05

CS120 The Information Era

9

Some Ground Rules

- Know your search engine
- Never look beyond the first 20 or 30 hits
- Experiment with different keywords
- Don't expect the first query you use to be your last

2/14/05

CS120 The Information Era

10

Problem

- Find top 20 cities in California (by population)
- When did forks first become commonplace?

2/14/05

CS120 The Information Era

11

Meta Search Engines

- Searches multiple search engines at once
- Examples
 - <http://www.dogpile.com>
 - <http://www.ixquick.com>
 - <http://vivosimo.com>

2/14/05

CS120 The Information Era

12

Meta Search Engines

- Find a ranked list of the most popular cars in the United States

2/14/05

CS120 The Information Era

13

4. Specialized Search Engine

- Like a search engine but limited to Web pages that feature a specific topic
- Relies on a collection of documents picked by people
- Examples
 - <http://www.invisibleweb.com>
 - <http://search.com>
 - <http://www.searchengineguide.com>

2/14/05

CS120 The Information Era

14

Types of Questions

- All web queries fall under one of the following question types:
 - Voyager question
 - Deep thought
 - Joe Friday

2/14/05

CS120 The Information Era

15

Voyager Questions

- Open-ended, exploratory question
- You want to be educated
- Cover a lot of ground and require time
- Use subject tree or clearinghouse

2/14/05

CS120 The Information Era

16

Deep Thought

- Open ended, but more focused, goal oriented
- Subject tree or specialized search engine

2/14/05

CS120 The Information Era

17

Joe Friday

- Very specific question, and has a simple answer
- Take a minute or two
- Subject tree or general search engine

2/14/05

CS120 The Information Era

18

To Help you Search

- Use nouns and objects as key words
- 6-8 keywords
- Truncate words (plane instead of airplanes)
- Use synonyms with an OR statement
 - Cat OR Kitten
- Combine keywords into phrases with quotes for exact matches (“New York”)

2/14/05

CS120 The Information Era

19

Problem

- Using Google and Webcrawler, search for Pacific University and then “Pacific University”

2/14/05

CS120 The Information Era

20

To Help you Search

- Use title search (intitle:plane)
- Use site search (site:ww.amazon.com)
- Wildcard matching (quart*)
- Required/prohibited terms (fuzzy operators)
 - +:required terms
 - -: prohibited terms

2/14/05

CS120 The Information Era

21

Logical Expressions

- Almost the same as fuzzy operators
- Usually, the advanced search technique for search engines
- Boolean operators
 - Required terms: AND
 - Possible terms: or
 - Prohibited terms: NOT
- NEAR operator

2/14/05

CS120 The Information Era

22

General Tips

- Know your search engine
- Experiment with different keywords and queries
- Use different engines
- Search for titles
- If you don't find anything in the first 20-30 hits, give up
- Try adding synonyms

2/14/05

CS120 The Information Era

23

Assessing Credibility

- Know who the author is
- Verify identity
- Verify organization
- Find corresponding print document
- Look for experts
- Accurate writing and documentation
- Citations
- Stability

2/14/05

CS120 The Information Era

24

Problem

- Have a go at exercise 25 on page 363

2/14/05

CS120 The Information Era

25

Summary

- We've completed Chapter 6 in the book

2/14/05

CS120 The Information Era

26