



## Searching for the Right Stuff

And boy, is there a lot of stuff to search through!

The best way to narrow down the amount of stuff you have to look through is by searching in the right area to begin with

© Copyright 2000-2001, University of Washington



## Finding and Evaluating Information

- Key principle:
  - The best place to search for something is where it is likely to be found.
  - Question: Where is the best place to find tax information?
    - Used car information?
- What are the potential places that one would find the best information?
  - Library – not everything is on the Internet, and it never will be! – and the library specializes in organizing information.
  - Specialized databases – if you do search for information electronically, go to the places where the content has already been evaluated and authenticated. Often these electronic databases can be used, for free, at ....the Library!
- So why all this focus on the library?

© Copyright 2002-2003, University of Washington



## What do Humans do?

- We associate things with other things and we organize them (group them together)
- Humans have a natural tendency to organize (cluster) similar things together
  - When you were a child, did you put all of your Barbie dolls in one place, or all of your robot toys? How about the same color toys?
- Think about the subjects you study in school: Math, Social Sciences, Geography, Art, etc.
  - Each of these subjects is further divided, becoming more and more detailed, specialized
- Librarians tackle the problem of trying to organize and place information where people can most easily find it

© Copyright 2000-2001, University of Washington



## Organizing Electronic Things

- If we tend to group similar things together, then we also tend to look for things that are similar in the same places
  - Problem: Do any 2 people organize things in exactly the same way? Do any 2 groups of people?
- Now, go to the web where all of that stuff sits
  - NO organization
- But, individual web sites can try to organize their “stuff”. A well-designed and well-organized site will help you find the “stuff” faster

© Copyright 2002-2003, University of Washington

## A Well Organized Site

- A web site that wants to help users find relevant information fast is usually organized in a manner similar to a Library:
  - The site is divided into broad categories (subjects)
  - Each category in turn has sub categories
    - Familiar? Didn't we try to organize folders on Dante in a meaningful manner?
- If a user that visits a site (you) can quickly see how things are categorized then they will find what they need much quicker
  - This assumes they have started looking in the right spot...

© Copyright 2002-2003, University of Washington

University Libraries  
UNIVERSITY OF WASHINGTON

# INFORMATION GATEWAY

UW Home | HealthLinks | UW Tacoma | UW Bothell | Image Collections

**Find It**  
UW Libraries Catalog  
Top 20 Databases  
Databases & Catalogs  
E-Journals  
By Subject  
Reference Tools

**Get It**  
Borrowing/Delivery  
Renewals  
Course Reserves  
View Your Record

**About the Libraries**  
Contact Us  
Libraries & Hours  
General Information  
Giving to the Libraries  
Accessibility

**Services**  
For Faculty & Staff  
For Graduate Students  
For Undergraduates  
For Visitors  
My Gateway

**Help**  
Connecting  
Starting Points  
Research Guides  
Library Classes

**Alerts**  
News  
September 11  
Events & Exhibits  
Employment Opportunities  
Serials Review Project  
Suzzallo Renovation

Search | Gateway Index

Take our survey and help improve the Information Gateway!

<http://www.lib.washington.edu>

© Copyright 2002-2003, University of Washington

INSIDE Husky Athletics WITH BARBARA HEDGES

# WASHINGTON HUSKIES

OFFICIAL ATHLETIC SITE

SCHEDULES | ONLINE STORE | TICKETS | TRAVEL | DEPARTMENTS | NEWS CLIPS | AUDIO/VIDEO | TV-RADIO | CAMPAIGN | HUSKY EVENTS

**Husky Headlines** Choose Sport

**UW Battles No. 16 Portland To 1-1 Tie In Men's Soccer Action**  
Huskies enter Pac-10 play next weekend with an overall record of 3-3-3.

**Washington Launches Huskies Pass, Giving Fans the Ultimate In Husky Programming**  
Huskies join more than 60 top colleges and universities in exciting new digital media subscription service.

**UW Football Game Day 2002**

**Grey, Sumantir Advance To Qualifying Draw Of ITA All-American Championships**  
Husky pair to join teammates Carter and Klac in qualifying singles action Oct. 8-9.

**Husky Men's Golf Team Opens at Fall Preview**  
Washington Makes First Appearance at Top Fall Tourney

**THIS WEEK @ UW**  
free email newsletter

**Washington vs. California Postgame Quotes**

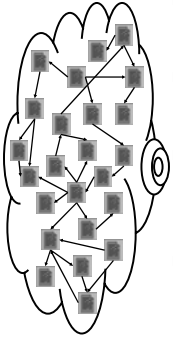
© Copyright 2000-2001, University of Washington

But what if you JUST DON'T KNOW?

Then the next step may be to go to do a search of information on the Web

© Copyright 2000-2001, University of Washington

## Structure of the Web



- The Web is just lots of documents on lots of machines
- Each document has a link to it from some other document
- If Page A links to Page B, you can get to page B from A (but not vice versa!)
- However, searching for information by going from page to page, one document at a time, would take the rest of eternity

© Copyright 2002-2003, University of Washington

## Directories: Yahoo!

- Not a search engine!
- A directory
  - Web pages have been organized into a hierarchical structure based on broad categories
  - Humans do the organizing, usually

<b>Arts &amp; Humanities</b> <a href="#">Literature, Photography...</a>	<b>News &amp; Media</b> <a href="#">Full Coverage, Newspapers, TV...</a>
<b>Business &amp; Economy</b> <a href="#">B2B, Finance, Shopping, Jobs...</a>	<b>Recreation &amp; Sports</b> <a href="#">Sports, Travel, Autos, Outdoors...</a>
<b>Computers &amp; Internet</b> <a href="#">Internet, WWW, Software, Games...</a>	<b>Reference</b> <a href="#">Libraries, Dictionaries, Quotations...</a>
<b>Education</b> <a href="#">College and University, K-12...</a>	<b>Regional</b> <a href="#">Countries, Regions, US States...</a>
<b>Entertainment</b> <a href="#">Picks, Movies, Humor, Music...</a>	<b>Science</b> <a href="#">Animals, Astronomy, Engineering...</a>
<b>Government</b> <a href="#">Elections, Military, Law, Taxes...</a>	<b>Social Science</b> <a href="#">Archaeology, Economics, Language...</a>
<b>Health</b> <a href="#">Medicine, Diseases, Drugs, Fitness...</a>	<b>Society &amp; Culture</b> <a href="#">People, Environment, Religion...</a>

© Copyright 2002-2003, University of Washington

## What is a Search Engine?

- A collection of programs designed to assist users in finding information
- Consists of four things (the book gives the 2 main ones):
  - A crawler (aka spider, robot)
  - A query processor
  - A user interface
  - A database

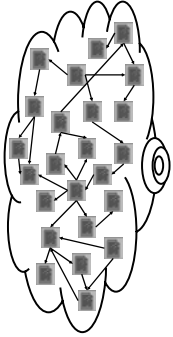
© Copyright 2002-2003, University of Washington

## Search Engine parts:

- A crawler does what the name implies:
  - "Crawls the Internet" building an index of URLs and key terms that are hopefully an indicator of the content of the page.
- A query processor takes a request from the user (search terms)
  - Interprets your question as best it can to return useful information
  - Retrieves the list of URLs associated with a given set of key word terms according to the index

© Copyright 2002-2003, University of Washington

## A Web Spider: See how they crawl!



- Start at a document
- Repeat these steps:
  - Store document
  - Extract all links
  - Index Terms
  - Follow every link
- Until: all links followed
- Redo when necessary
- Need to pick good start pages
  - Pages with multiple links

© Copyright 2002-2003, University of Washington

## The Search, literally, never ends

- Crawling is an ongoing process. Why?
  - To keep up with the millions of new pages that are added the the Web on a weekly basis
  - To go back and revisit sites and make sure that links aren't broken
- Lists of URL's are created on the spot for users
  - But the query processor doesn't have a robot search the web each time
  - Goes to the database of searched pages and matches search terms with index terms
- The effectiveness of a search engine will depend in part on how much of the Internet the crawlers have seen and how discriminating the index is.

© Copyright 2002-2003, University of Washington

## Web Searching

Knowing what you want and  
knowing how to ask for it are two  
different things

© Copyright 2000-2001, University of Washington

## Search Technique

- Search engines use basic logic to determine how to answer your question
- Boolean logic, in the form of Boolean operators, are the foundation of search logic:
  - AND
  - OR
  - NOT
- Many search engines now use "search math" instead of Boolean terms
  - +, -

© Copyright 2002-2003, University of Washington

## Choosing Effective Keywords

- Match your keyword to the specificity of your search need.
- Use the most specific word(s) available to describe your target.
- Avoid words with multiple meanings unless you can narrow them down with additional terms.

© Copyright 2002-2003, University of Washington

## Boolean Operators

- AND (+): Narrows results --> Dog AND Cat
  - Pages are returned ONLY if Dog and Cat appear in the index for that page
- OR: Widens results --> Dog OR Cat
  - Pages are returned if just Dog is in the index, or if just Cat is in the index, or if both terms appear in the index for that page
- NOT (-) Narrows results --> Dog NOT Cat
  - Dog is in the index, and Cat is NOT.
  - No pages with Cat will be returned, even if Dog is in them

© Copyright 2002-2003, University of Washington

## Narrowing Your Results

- Add more terms.
- Add more specific terms.
- Use Boolean/Math qualifiers
  - Parenthesis have the same effect in Search Engines as they do in math

Example: (Dog OR Cat) AND Bird

Will find all pages where **Dog** or **Cat** or both are in the page, then look to make sure **Bird** is also on those pages and return those results

© Copyright 2002-2003, University of Washington

## Grouping

- Search for exact phrases, like titles, with quotation marks
  - "Fellowship of the Ring"
- Using " " around a phrase in many search engines will limit results to pages indexed with the phrase -->
  - "Fellowship of the Ring"

it will not retrieve

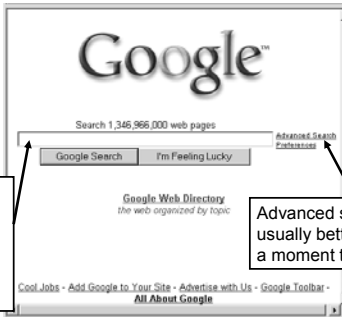
**"Fellowship, Rings and community"**

but will retrieve

**"Fellowship of the Ring should get an Oscar!"**

© Copyright 2002-2003, University of Washington

## A bit about Google

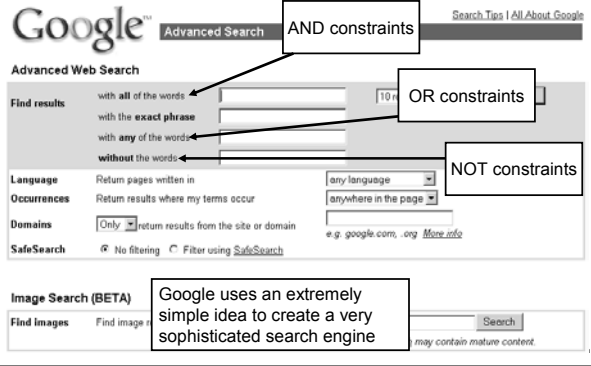


The "basic" search, seen here, is often effective just based on their rankings

Advanced search is usually better if you spend a moment thinking about it

© Copyright 2002-2003, University of Washington

## Advanced Search



Google uses an extremely simple idea to create a very sophisticated search engine

## Google's Sophistication

- Google is a search engine that indexes a page by the words used in the anchor tags of pages that link to it

favorite, Thai, place, ...

The Cuisine of Siam

Molly's Home Page

Check out my **favorite Thai place**

LEMON GRASS:

restaurant, Thai, Lemon, Grass, Chate, ...

Springfield Restaurant Review

ETHNIC – **Thai Lemon Grass Restaurant** (lemongrass.com) is a well-appointed eatery run by **Chate Koolerang**. Their specialty..

© Copyright 2002-2003, University of Washington

## More Google Sophistication

- Popularity is also a key to Google's rankings
- If page A links to page B, then that is considered a vote by page A for page B
- If page A is also a very popular site that many other sites link to, then page A's vote is worth more
- How can WE get a sense of how popular a web site is? Do a simple link search in Google to see how many sites link to the site that interests us.
  - link:www.washington.edu/

© Copyright 2002-2003, University of Washington



## Summary

- Search where you are most likely to find the information
- Good sites will have effective navigation that you can easily figure out
- Local searches in good sites can quickly find candidate pages
- Search Engines build indexes to assist in searching the web
- When doing a search of two or more words or phrases, specify whether
  - Both words **MUST** be present: AND, +
  - At least one of the words must be present: OR
  - The word(s) must **NOT** be present: NOT, -