
Searching for Information

INFO/CSE 100, Autumn 2004
Fluency in Information Technology

<http://www.cs.washington.edu/100>

Readings and References

- Reading
 - » *Fluency with Information Technology*
 - Chapter 5, Searching for Truth
- References
 - » Research 101, UW Libraries
 - **Research 101** is intended to help students learn to perform basic information research so they can tackle information problems anywhere.
 - <http://www.lib.washington.edu/uwill/research101/>

Searching for Information

- The Web and its search engines have made it much easier to quickly find the top-level information about a topic
 - » A simple Google or Yahoo search will turn up *something* on almost any topic
 - » That's enough if you're just trying to get a general idea about a topic
- But what if you really care about the accuracy?
 - » Look deeper, cross check, work with experts

Information hierarchies

- Experts in a particular field usually organize the information about the field into a structure
 - » If you understand the overall structure and ...
 - » you want some detail, then ...
 - » you can move quickly through the structure to the particular item of information that you seek
- This works very well
 - » IF you know the structure well enough to navigate within it.

Biological Classification

Table 9b-1: Hierarchical system of the biological classification of an organism.

Category	Name	Characteristics
Kingdom	Plantae	Organisms that usually have rigid cell walls and usually possess chlorophyll.
Subkingdom	Embryophyta	Plants forming embryos.
Phylum	Tracheophyta	Vascular plants.
Subphylum	Pterophytina	Generally large, conspicuous leaves, complex vascular system.
Class	Angiospermae	Flowering plants, seed enclosed in ovary.
Subclass	Dicotyledoneae	Embryo with two seed leaves.
Order	Sapindales	Soapberry order consisting of a number of trees and shrubs.
Family	Aceraceae	Maple family.
Genus	Acer	Maples and box elder.
Species	<u>Acer rubrum</u>	Red maple.

From PhysicalGeography.net, <http://www.physicalgeography.net/fundamentals/9b.html>

Library Classification Systems

- Within a library collection, materials are typically organized by subject. Librarians assign a call number based on a work's subject
 - » Library of Congress
 - Used in most college, university, and research libraries because it handles large collections
 - » Dewey Decimal
 - Used in most public and school libraries because it is more effective for smaller collections

From Research 101, <http://www.lib.washington.edu/uwill/research101/>

Library of Congress Classification

A -- GENERAL WORKS
 B -- PHILOSOPHY. PSYCHOLOGY. RELIGION
 C -- AUXILIARY SCIENCES OF HISTORY
 D -- HISTORY (GENERAL) AND HISTORY OF EUROPE
 E -- HISTORY: AMERICA
 F -- HISTORY: AMERICA
 G -- GEOGRAPHY. ANTHROPOLOGY. RECREATION
 H -- SOCIAL SCIENCES
 J -- POLITICAL SCIENCE
 K -- LAW
 L -- EDUCATION
 M -- MUSIC AND BOOKS ON MUSIC
 N -- FINE ARTS
 P -- LANGUAGE AND LITERATURE
 Q -- SCIENCE
 R -- MEDICINE
 S -- AGRICULTURE
 T -- TECHNOLOGY
 U -- MILITARY SCIENCE
 V -- NAVAL SCIENCE
 Z -- BIBLIOGRAPHY. LIBRARY SCIENCE. INFORMATION RESOURCES (GENERAL)

LIBRARY OF CONGRESS CLASSIFICATION OUTLINE



CLASS N - FINE ARTS



Subclass N Visual arts
 Subclass NA Architecture
 Subclass NB Sculpture
 Subclass NC Drawing.Design.Illustration
 Subclass ND Painting
 Subclass NE Print media
 Subclass NK Decorative arts
 Subclass NX Arts in general

Press 1 for sales, press 2 for ...

- Hello, thank you for calling ...
 - » press 1 for sales
 - » press 2 for frequently asked questions
 - » press 3 for account information
 - » press 4 customer support
- Poorly designed phone menu hierarchies are a royal pain
 - » dead ends are a waste of time
 - » multiple steps to get to the desired information
 - » designed to reduce costly customer interaction



Is the Web a hierarchy?

- Not a hierarchy, and that's good (to some)
 - » A major part of its amazing success is that relevant information is just a click away,
 - » That's why it's "The Web" not "The Tree"
- Not a hierarchy, and that's bad (to some)
 - » Any web page can link to any other web page
 - » This is causing heartburn among the gatekeepers
 - » You can link directly to a page on a web site without going through a top-level page



In the beginning

- When the Web was new
 - » there was no hierarchy and ...
 - » there was no way to find what you wanted other than careful detective work and keeping good notes
- Web detective games
 - » How quickly can you find the following factoid?
 - » Quite a challenge - those who knew their links could find information quickly

Now: Scan, then drill down

- Search engines do the top-level scanning for us
 - » Enter your search term in Google, Yahoo, etc
 - » You will get back a list of sites that are in some way related to your topic
- The links are entry points into the giant hairball that is the web
 - » It's still not a hierarchy, but you can get close to the high value sites quickly
 - » Drill down for information within those sites

Now: Select, then drill down

- Search engines are not the only way to get to the top of an information rich tree
 - » UW Administrative information?
 - UW Information Navigator page
 - » Scholarly information?
 - UW library has high-density information
 - » Tax forms and laws?
 - The IRS web site is the definitive source

Search Engines

- No one controls or assigns hierarchy locations to the pages published on the WWW ... it is totally decentralized
- To find pages, search engines crawl Web
 - » Two parts
 - Crawler visits Web pages building an index of the content
 - Query processor checks user requests against the index, reports on known pages

Only a fraction of the Web's content is crawled

Google Advanced

Google™ Advanced Search

Advanced Search Tips | All About Google

Find results with all of the words [input] 10 results [dropdown] Google Search

with the exact phrase [input]

with any of the words [input]

without the words [input]

Language Return pages written in [any language dropdown]

File Format [Only dropdown] return results of the file format [any format dropdown]

Date Return web pages updated in the [anytime dropdown]

Occurrences Return results where my terms occur [anywhere in the page dropdown]

Domains [Only dropdown] return results from the site or domain [input] e.g. google.com, .org [More info](#)

SafeSearch No filtering Filter using [SafeSearch](#)

Boolean Queries

- Search Engine words are independent
 - » The words don't have to occur together on the page
- To be explicit about occurrences use Boolean queries and quotes
 - » Logical Operators: AND, OR, NOT
 - monet AND water AND lilies
 - “van gogh” OR gauguin
 - vermeer AND girl AND NOT pearl

See also <http://www.google.com/help/refinerearch.html>

Queries

- Searching strategies ...
 - » Limit by top level domains or format
 - » Find terms most specific to topic
 - » Look elsewhere for key words
 - » Use exact phrase only when universal
 - » If too many hits, requery
 - » Once found, ask if site is best source

11-Oct-2004

cse100-06-misinformation © 2004 University of Washington

17

Selected Sites

- For some searches, there are information rich areas in which to start. No need to Google.
- Scholarly material
 - » UW Library
 - » Library of Congress
- Government information
 - » <http://access.wa.gov/>
 - » <http://www.firstgov.gov/>

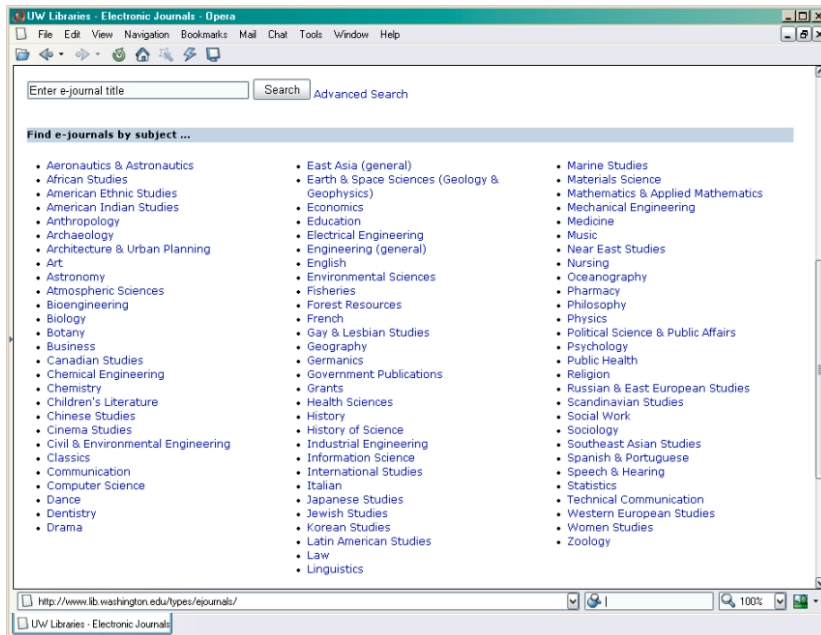
11-Oct-2004

cse100-06-misinformation © 2004 University of Washington

18

The screenshot shows the University of Washington Libraries homepage. The browser window title is "University of Washington Libraries - Opera". The page features a navigation menu on the left with links to "Resources" (UW Libraries Catalog, Research Databases, etc.), "Services" (Your Library Account, Connect from Off-Campus, etc.), and "About the Libraries" (Hours & Libraries, Collections, etc.). A central search box is titled "Search the UW Libraries Catalog" and includes a "Keyword" input field and a "Search" button. Below the search box are sections for "How do I..." (with a "Get an article" dropdown) and "Spotlight" (with news items about proxy services and international access). The footer contains the URL "http://www.lib.washington.edu/" and a "Google search" button.

The screenshot shows the University of Washington Libraries Catalog Menu. The browser window title is "University of Washington Libraries Catalog Menu - Opera". The page features a navigation menu at the top with links to "Libraries Home", "Catalog Home", "Your Library Account", and "Help". A search box is titled "Search the Catalog" and includes a "Keyword" input field and a "Search" button. Below the search box are sections for "Or Select an advanced search screen below:" (with links to Author, Title, Keywords, Call Numbers, Journal searches, Library of Congress Subjects, Medical Subjects, Genre/Form, ISBN & Other Numbers) and "Other Catalog Views:" (with links to Unicode and East Asian Character Sets, Telnet Catalog). The footer contains the URL "http://catalog.lib.washington.edu/search?/" and a "Google search" button.



The Web and reality

- The Web contains only a fraction of the information available to us
 - » It provides a great tool for getting started with a topic
 - » It provides rapid access to many significant collections of information
- The real world contains much, much more!
 - » Use the Web as a map, then go exploring
 - » Go to the library, talk to a librarian, read the primary sources, talk to the experts in the field