Readings and References

Searching for Information

INFO/CSE 100, Autumn 2004 Fluency in Information Technology

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

cse100-06-misinformation © 2004 University of Washington

• 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/

11-Oct-2004

cse100-06-misinformation © 2004 University of Washington

2

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

11-Oct-2004

11-Oct-2004

3

Information hierarchies

- Experts in a particular field usually organize the information about the field into a structure
 - $\, \ast \,$ If you understand the overall structure and $\ldots \,$
 - » 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	Acer rubrum	Red maple.

From PhysicalGeography.net, http://www.physicalgeography.net/fundamentals/9b.htm

11-Oct-2004

cse100-06-misinformation © 2004 University of Washington

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

n.edu/uwill/research101/	From Research 101, http://www.lib.washin		damentals/9b.html
6	cse100-06-misinformation © 2004 University of Washington	11-Oct-2004	5

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 **O** -- SCIENCE R -- MEDICINE S -- AGRICULTURE T -- TECHNOLOGY U -- MILITARY SCIENCE V -- NAVAL SCIENCE Z -- BIBLIOGRAPHY. LIBRARY SCIENCE. INFORMATION RESOURCES (GENERAL)



Subclass ND Painting

Subclass NE Print media

Subclass NK Decorative arts

Subclass NX Arts in general

7

Press 1 for sales, press 2 for ...

- Hello, thank you for calling ...
 - » press 1 for sales



9

11

- » 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

11-Oct-2004

cse100-06-misinformation © 2004 University of Washington

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

11-Oct-2004

cse100-06-misinformation © 2004 University of Washington

10

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

11

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

- 11-Oct-2004
- cse100-06-misinformation © 2004 University of Washington

14

Google Advanced

cse100-06-misinformation © 2004 University of Washington

Go	Advanced Search	Advanced Search Tips All About Google
Find results	with all of the words with the exact phrase with any of the words without the words	10 results 🔽 Google Search
anguage	Return pages written in	any language 💌
ile Format	Only 💌 return results of the file format	any format
Date	Return web pages updated in the	anytime 💌
Occurrences	Return results where my terms occur	anywhere in the page 💌
Domains	Only return results from the site or domain	e.g. google.com, .org <u>More info</u>
SafeSearch	No filtering O Filter using <u>SafeSearch</u>	

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

11-Oct-2004

15

13

See also http://www.google.com/help/refinesearch.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

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/

Image: Seconds Multiple Seconds Multiple Image: Seconds Multiple Seconds Multip	1-001-2004		Jud Oniversity of Washington	17	11-00-2004		
File Edit Vew Nurgenn Related Mail Dat Eak Vinder Heip File Edit Vew Nurgenn Related Relation R	niversity of Washington Librar	ries - Opera					
Image: Search the Status Interview Interview<	File Edit View Navigation	Bookmarks Mail Chat Tools Window Help		_ #×	File Edit View Navigation	res Catalog Menu - Opera Bookmarks Mail Chat Tools Window Help	7
WASHINGTON Cifrarie WW Hore avaluation eed help with research? Interdiscont Scalage Resources by Interdiscont Scalage Resources by Interdiscont Scalage Resources brain Search the UW Libraries Catalog Resources by Subject Resources by Subject Resources by Subject Resources brain Resources brain Resources by Subject Resources by Subject Resources brain Resoures Brain Resoure	🚸 • • 🔊 🙆 🖄	<i>₽</i> ₽				<i>3</i> D	
with research Now a question anall dist phone Nore Search the UW Libraries Catalog Nescurces by Subject Reference Tools Nore Search the Search OW Home > UW Home > UW Home > UW Home > UW Home > Ourse from Off-Campus I head biologies Nore : Search this Stell Can Nore : Nore : Search this Stell Can Nore : Nore : Nore : Search this Stell Can Nore : Nore : Nore : Search this Stell Can Nore : Nore : Nore : Nore : Nore : Search this Stell Can Nore : <	WASHINGTON Libraries	Resources >>			WASHINGTON Libraries	UW Libraries Catalog Libraries Home Catalog Home Library Account Help	Off-Campus
W Home >> UW Home >> WB dothell Library >> WB dothell Library >> Y Tacoma Library >> Y Tacoma Library >> Y Connect from Off-Campus > Course Reserves > Interlibrary Lana > Interlibrary Lana > Reserve a Study Room > Reserve a Study Room > Author + > Library Course Reserves > Collections > More Spotlight About the Libraries > > Collections > Collections > Collections > Staff Directory </td <td>eed help with research? Have a question? ask us email chat phone</td> <td> UW Libraries Catalog Research Databases Resources by Subject Electronic Journals Reference Tools </td> <td>Search the UW Libraries Catalog Keyword Search More search options / Summit Catalog</td> <td></td> <td>Search the Catalog</td> <td>For: In: ord V Entire Collection V Search</td> <td></td>	eed help with research? Have a question? ask us email chat phone	 UW Libraries Catalog Research Databases Resources by Subject Electronic Journals Reference Tools 	Search the UW Libraries Catalog Keyword Search More search options / Summit Catalog		Search the Catalog	For: In: ord V Entire Collection V Search	
UW Borners Services >> UW Borners > Your Library Account + Your Library Account + Ourse Account + Ourse Account + Ourse Account + Connect from Off-Campus > Connect from Off-Campus > Course Reserves > Interlibrary Lann > Course Acserve a Study Room > Interlibrary Lann > More Spotlight About the Libraries > Collections <td></td> <td>More</td> <td></td> <td></td> <td>Company Company</td> <td>0</td> <td></td>		More			Company Company	0	
bit nterlibrary Loan • Interlibrary Loan • Reserve a Study Room • Reserve a Study Room • Reserve a Study Room • More • More • More • What's new in the Libraries : Fall, 2004 • Other Catalog Views: • Other Catalog views: • Unicode and East Asian Character Sets • Telnet Catalog • UW Law Library (also in Summit) • What's new in the Libraries : Fall, 2004 • Details on the new proxy service, website, changolicy and more. • Collections • Staff Directory • Collection and exact on policy and more. • Interlibration edu/seach**/ • Directory <	UW Home » UW Bothell Library » UW Tacoma Library » Health Sciences Library »	 Your Library Account Connect from Off-Campus Course Reserves 	How do I Get an article 🔽 Go		Searches: Author Title Keywor Call Nu Journal	Library of Congress Subjects ' By Course ' Medical Subjects ' By Profess ds ' Genra/Form mbers ' ISBN & Other Numbers searches	sor
	earch this site Go	 Interlibrary Loan Reserve a Study Room More About the Libraries >> 	Spotlight		Other Catal Vunicode Telnet	ng Views: and East Asian Character Sets Catalog Catalog Other Catalogs with UW resources: Summit: a catalog of regional librar VUW Law Library (also in Summit)	ies
		 Hours & Libraries Collections Staff Directory 	What's new in the Libraries : Fall, 2004 Details on the new proxy service, webs changes to the loan policy and more. The UW Libraries is participating in The Soctember Project - an interactional or	ite,	© 1998-2004 University of W	ashington Libraries	Q 100%
	1ttp://www.lib.washington.edu/		🖌 🚱 Google search	100% 🖌 🄜 -	Inttp://catalog.iib.washington.edu/ University a() (askington.edu/	search /	100%

inter e-journal title	Search Advanced Search		
na e-gournais by subject Aeronautics & Astronautics African Studies American Indian Studies Archaeology Archaeology Archaeology Archaeology Archaeology Biology Biology Biology Biology Biology Charlen's Literature Chimers Studies Chemisty Children's Literature Chimes Studies Cinema Studies Communication Computer Science Dance Dentistry Drama	East Asia (general) Earth & Space Sciences (Geology & Geophysics) Economics Education Education Educiation (general) English Environmental Sciences Fisheries Forench Geography Germanics Government Publications Grants Health Sciences History History of Science Indomational Studies Health Sciences History History of Science Indomational Studies Italian Japanese Studies Jewich Studies Korean Studies Law Lany	 Marine Studies Materials Science Mathematics & Applied Mathematics Medicince Musicia Engineering Musici Nursing Oceanography Pharmacy Phiarmacy <l< th=""><th>10</th></l<>	10
http://www.lib.washington.edu/types/ejournals/		V 🚱 I 🔍 100% V	•

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

- cse100-06-misinformation © 2004 University of Washington
- 22