How to use the WWW as a Research Tool

Dr. Chris Staff

About...

- Aims and Objectives

- Introduction
  - Why do we do research?
  - How do we present our findings?

About...

- Web-based & specialist search-engines
  - Google/Yahoo
  - Search-engines demystified
    - How do they work
      - Indexing
      - Searching

chris.staff@um.edu.mt

About...

- Web-based & specialist search-engines (contd...)
  - Tips and Tricks
    - What You Ask For Is What You Get
    - Using advanced settings
    - The search engine's search tips
    - Phrases
    - Site searches
    - OR and NOT
    - Searching for links!

chris.staff@um.edu.mt

About...

- Finding authoritative sources on the Internet
  - How do you know what you can trust?
  - Characteristics of trusted sources: references, references, references!
  - Subject specific search engines
  - Cyber-citation

About...

- Note taking and Literature reviews
- Proper citation of sources!
- Finding and using web-based tools to help with research activities
- Roles of the Library in the Internet Age

Aims and Objectives

- Why do we do research anyway?
- How do Web-based search engines work?
- How can we get the best results from Web-based search engines?
- How can we know we can trust the information we find?

Aims and Objectives

- How might we take notes of what we find, so we can use it properly?

- How do we reference or cite the sources of information we use?

- How can I find Web-based tools to help with data analysis?

- What role does the Library have these days?

How do *you* do research?

Log-book activity

- Either: write a short (500 word) literature review of common misconceptions about
  - the Internet
  - Computers
  - WWW
  - a combination of the above
- Or: keep a log-book of research you do for an assignment this semester (how you did it)

Log-book activity

- The log-book should be uploaded to your web site
  - which you'll learn in the next part of the ICT activity

Why do we do research?

- To support an academic report, which we write to:
  - report on current state-of-the-art
  - describe an alternative solution to some problem

Why do we do research?

- We should always describe the state-of-the-art!
  - To inform others of our findings, so they can trust our recommendations
  - To back up claims we make by referring to work of others
  - To show that we've considered a sufficient coverage of existing approaches
  - That we've acquired a certain depth of knowledge
  - That we can share that knowledge critically!

How to present findings

- A reader of a report needs to tell difference between and the relationship between:
  - The words of others that we are reporting
  - Our interpretation of the work of others
  - Our own contribution
- So it's important to cite our sources properly

Examples

Examples

- If the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

Examples

- If the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer (2007) claims that if the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

Examples

- If the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer (2007) claims that if the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer claims that if the world was flat, the oceans would fall off the edges of the earth (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round.

Examples

- If the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer (2007) claims that if the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer claims that if the world was flat, the oceans would fall off the edges of the earth (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer claims that if the world was flat, "water would be unable to remain in anything other than deep pools" (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round.

Examples

- If the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer (2007) claims that if the world was flat, the oceans would fall off the edges of the earth. The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer claims that if the world was flat, the oceans would fall off the edges of the earth (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer claims that if the world was flat, "water would be unable to remain in anything other than deep pools" (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round.

- Ebejer claims that if the world was flat, "water would be unable to remain in anything other than deep pools" (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round (Axisa, 2007).

- Compelling reasons for doing a proper research job, and reporting findings accurately and correctly
- We'll now look at how to find relevant and trustworthy material and how to use it and reference it in academic reports

How do Web-based search engines work?

- Web contains huge number of documents
  - Nobody knows exactly how many
  - No single search engine indexes all documents
  - Google probably indexes more than any other, but probably only 20% of Web-accessible docs

Web-based search engines...

- There is no comprehensive "directory" of web pages

- Search engines must "discover" pages to index

  - Start at some well-known highly linked pages, and follow links

  - How will Google find the web pages you create?

Web-based search engines...

- So search-engines crawl the Web looking for pages to index

- Indexing involves creating a *representation* of the page, usually of *term features* that occur in the document

  - term features include: frequency count, term location, Google's PageRank, etc.

Web-based search engines...

- Once all page are indexed, we have an *inverted index* of all unique terms that occur in Web pages, together with list of web pages containing the term
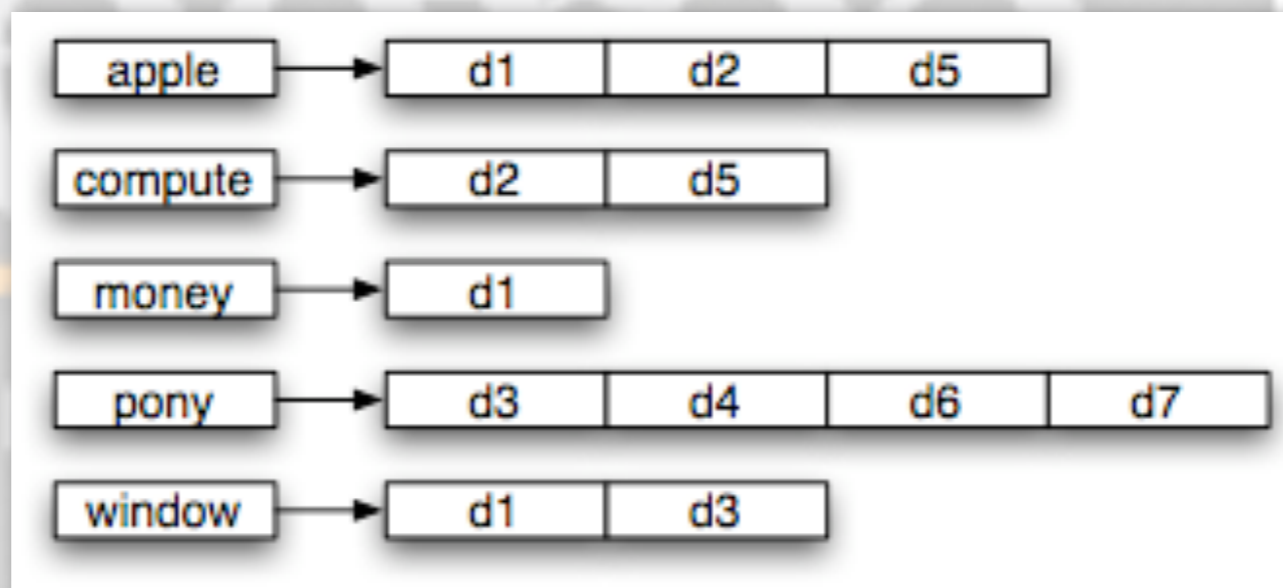
Getting best results from search engine

- You can use a search engine by:
  - Visiting its home page
  - Using your browser's search box
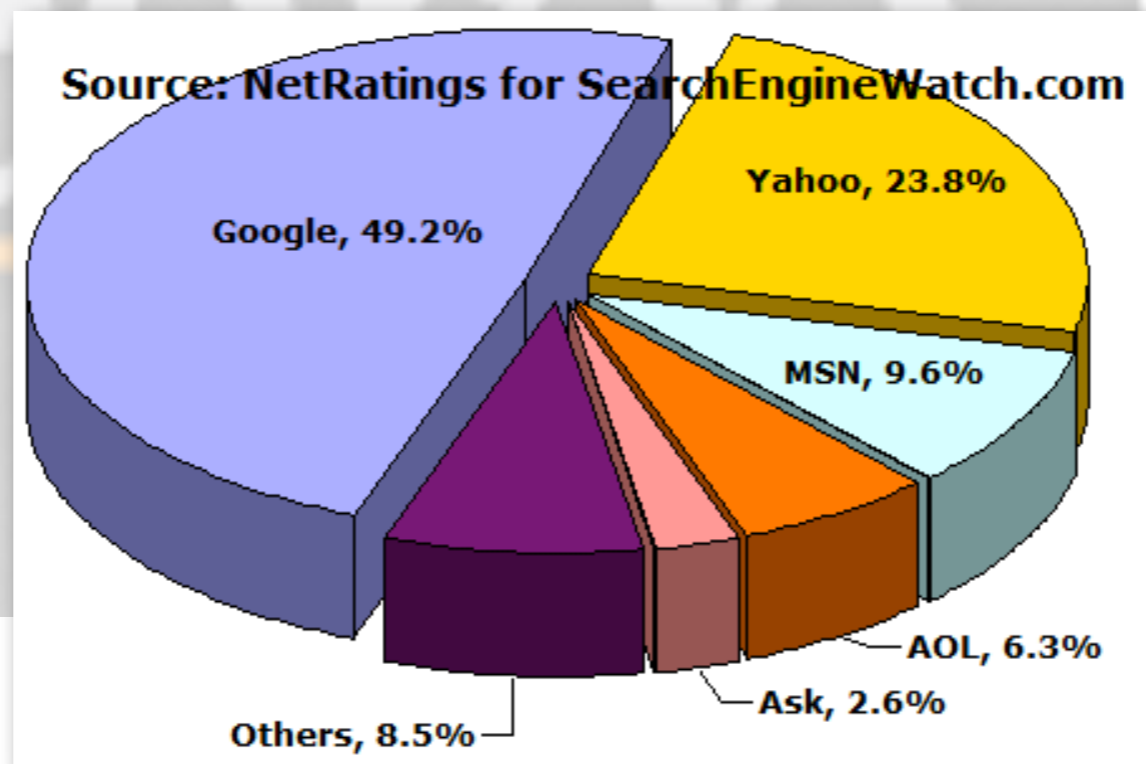 - Selecting words/phrases on a page and right-clicking

Activity 1

- Assume that you want to find out about common computer, Internet, and Web misconceptions...

- Look for some information now...

Advanced Search

- We'll look at Google's and Yahoo's advanced search options and short-cuts
  - Other search engines have similar advanced options
- Google + Yahoo account for 73% of all queries!

Source: NetRatings for SearchEngineWatch.com
Google, 49.2%
Yahoo, 23.8%
MSN, 9.6%
AOL, 6.3%
Ask, 2.6%
Others, 8.5%

chris.staff@um.edu.mt

YAHOO! SEARCH

Advanced Web Search

You can use the options on this page to create a very specific search. Just fill in the fields you need for your current search.

[Yahoo! Search]

Show results with

all of these words	[ ]	[any part of the page ▼]
the exact phrase	[ ]	[any part of the page ▼]
any of these words	[ ]	[any part of the page ▼]
none of these words	[ ]	[any part of the page ▼]

Tip: Use these options to look for an exact phrase or to exclude pages containing certain words. You can also limit your search to certain parts of pages.

Updated [anytime ▼]

Site/Domain
- ⦿ Any domain
- ○ Only **.com** domains ○ Only **.edu** domains
- ○ Only **.gov** domains ○ Only **.org** domains

- ○ only search in this domain/site: [   ]

Tip: You can search for results in a specific website (e.g. yahoo.com) or top-level domains (e.g. .com, .org, .gov).

Creative Commons Search
BETA

- ☐ Search only for **Creative Commons** licensed content
  - ☐ Find content I can use for commercial purposes
  - ☐ Find content I can modify, adapt, or build upon

Tip: This special Yahoo! Search finds pages that have content with a Creative Commons license. Learn more...

ⓒ creative commons

File Format Only find results that are: [all formats ▼]

chris.staff@um.edu.mt   25 of 48

Activity 2

- Go to Google or Yahoo and type
  - common misconceptions into Exact Phrase field
  - www computers internet into Any
  - http://www.google.com
  - http://search.yahoo.com/
- see if you get better results than before

Advanced query operators

- Normally, words entered into the simple query field are ANDed, and term proximity is not explicit

- But the advanced search options are not too great either
  - What if we want to search for www or world wide web?
  - What happens if we just add all these into Any?

chris.staff@um.edu.mt

Advanced query operators

- We can create a phrase by placing "" around the text
  - But it is ignored if you put it in the Any field!
  - We must use the simple query field (from search engine home page or search box in browser) and also use Boolean operators
  - e.g., "common misconceptions" computers OR internet OR www OR "world wide web"

Advanced query operators

- We can use "" in the simple query box to combine phrases in ways that cannot be done in the "the exact phrase" field
  - "world wide web" OR www
  - "common misconception" "world wide web" OR www -evolution

Advanced query operators

- Google's help on advanced options is at http://www.google.com/intl/en/help/refinesearch.html

- Yahoo's is at http://help.yahoo.com/l/us/yahoo/search/

- Also, can find information about other operators, and a 'cheat sheet'

Activity 3

- Read the information on Google's (or your favourite search engine's) advanced search.

- Construct a query to find the home page of the Department of Artificial Intelligence at the University of Malta (um.edu.mt). As the department has recently split from the Department of Computer Science and AI, the query should specify that "Computer Science" should not appear in the results.

Trusting the information we find

- Visit http://uncyclopedia.org/wiki/Eiffel_Tower
- Can you trust the information there?
  - Why?
  - Why not?

Trust

- Web has no "editor" who vets information that's published

- Information does not have an independently assigned trust mark

- But for an academic report, it is essential that we use trustworthy information only

  - Otherwise, the verity of our own work will be questioned!

Trust

- Nowadays, can usually find on-line journals, conference proceedings, trusted collections, for any discipline (though you may need to pay for access!)

- These are trusted, because publications are peer-reviewed - so use them!

- For others, use the link: operator in Google/Yahoo to see what others who link to the site say about it

  - Combine it with -site: to exclude self-references

Sources to avoid

- Avoid relying on the following sources:
  - Wikipedia!
    - and similar...
    - Correct information in Wikipedia is likely to also be found in a trusted source, so use the trusted source
  - Magazines
  - Blogs
  - Newspapers articles (unless necessary)
  - Personal home pages...

Citation and Cyber-Citation

- Need to distinguish between words of others, our interpretation of the words of others, and our own contribution
- Ebejer claims that if the world was flat, "water would be unable to remain in anything other than deep pools" (Ebejer, 2007). The fact that Malta is surrounded by water is proof that the world is round (Axisa, 2007).
- Anything not followed by a citation is assumed to be ours

Citation and Cyber-Citation

- The full publishing details of each referenced work should be listed in a references section

- Full details include author details, year of publication, where published, etc.

- See http://www.jiscpas.ac.uk/referencing.php

- Difference between a references section and a bibliography...

Taking notes

- When we take notes, we need to remember which are the words of others, so we can write it in quotation marks

- How much do we need to quote?
  - Unusual terminology
  - Phrases (two or more words!)
    - Keep longer quotations to a minimum - paraphrase instead

- Take a note of the URL too!

Examples

[13] Ahu Sieg, Bamshad Mobasher, Steve Lytinen, Robin Burke, Concept Based Query Enhancement in the ARCH Search Agent (2003)

```
@misc{ sieg03concept,
 author = "A. Sieg and B. Mobasher and S. Lytinen and R. Burke",
 title = "Concept based query enhancement in the {ARCH} search agent",
 text = "A. Sieg, B. Mobasher, S. Lytinen, and R. Burke. Concept based query enhancement
   in the arch search agent. In Proceedings of the 4th International Conference
   on Internet Computing, Las Vegas, NV, 2003.",
 year = "2003",
 url = "citeseer.ist.psu.edu/sieg03concept.html" }
```

An "interactive query formulation aid" "based on conceptual queries"
Uses "domain knowledge inherent in Web-based classification hierarchies such as Yahoo to add just those terms likely to improve the match with the user's interest"
Users submit a query to ARCH which then displays either the appropriate portion of the hierarchy or else the entire concept hierarchy.

[14] Using Semantic Treemaps to categorize and visualize bookmark files (2002)
                            Ying Feng, Katy Börner

```
@inproceedings{ feng02using,
 author = "Y. Feng and K. Brner",
 title = "Using semantic treemaps to categorize and visualize bookmark files",
 text = "Feng Y. & Brner K. Using semantic treemaps to categorize and visualize
   bookmark files. In Proceeding of SPIE - Visualization and Data Analysis.
   Volume 4665, January 2002, San Jose, CA, USA, pp. 218-227.",
 year = "2002",
 url = "citeseer.ist.psu.edu/feng02using.html" }
```

"analyses, categorizes, and visualizes the semantic relationships of web pages in personal bookmark or favorites collections based on their semantic similarity"

The Literature Review

- Describe approaches in parallel, rather than in sequence

- Avoid writing a list of detailed descriptions

- Instead, look for similarities between approaches, or common approaches, and write *statements* followed by a list of references that support the statement

- This gives a unique view of the literature (because it's based on the papers *you* read), and reduces the chances of "accidental" plagiarism

Data Collection and Analysis

- Finding and using Web-based collections and analysis
  - How are you collecting data?
  - Do you need a questionnaire?
  - Do you need to conduct a survey?
  - Do you need to analyse your data?

Data Collection and Analysis

- Lots of Web-based guides to questionnaire writing and determining survey (population) sizes...
  - http://www.sysurvey.com/tips/arsham.htm
  - http://www.deyalexander.com.au/resources/uxd/surveys.html
- ... and Web-based analytical tools
  - http://statpages.org/, http://davidmlane.com/hyperstat/, and http://statpages.org/javasta4.html

Role of the Library

- Unlike a library, the Web is always "open"
- All Web-based publications are accessible without having to leave your computer!
- But, significant amounts of copyrighted material are not available on-line
  - Not necessarily critical for your discipline
  - Major search engines, on-line bookstores trying to find solution

Role of the Library

- Unlike Web, Library is an organised collection of information
  - Experts have decided in which sub-collection information should be placed
  - Can find one relevant publication through OPAC and browse through 'nearest neighbours'
- Librarians can help you find information, and guide you through relevant collection!
- Librarians are trained to update collections
- Libraries may host special collections, e.g., Melitensia

Conclusions

- We do research so others can learn from us
- We need to back up all the claims we make
  - Either by providing data as evidence or by referring to others who substantiate the claim
- Using a search engine's advanced search features can help us locate better quality information

Conclusions

- It's important that we cite our findings adequately, so a reader can easily tell if the claim is ours or somebody else's
  - and if somebody else's, whether we are using their words or our own words to describe them
- Taking notes carefully can help us avoid plagiarism, and to write a good literature review
- On-line tools can help with data collection and analysis.

chris.staff@um.edu.mt

Conclusions

- Specialist collections, and the library, are better sources of reliable and trustworthy information than places like wikipedia, newspapers and magazines, and blogs!

Building a Web Site

- Next ICT degreePlus activity is to show you how to build your own Web Site.

- Facilitated by Marie Louise Scifo and Christian Tabone

chris.staff@um.edu.mt