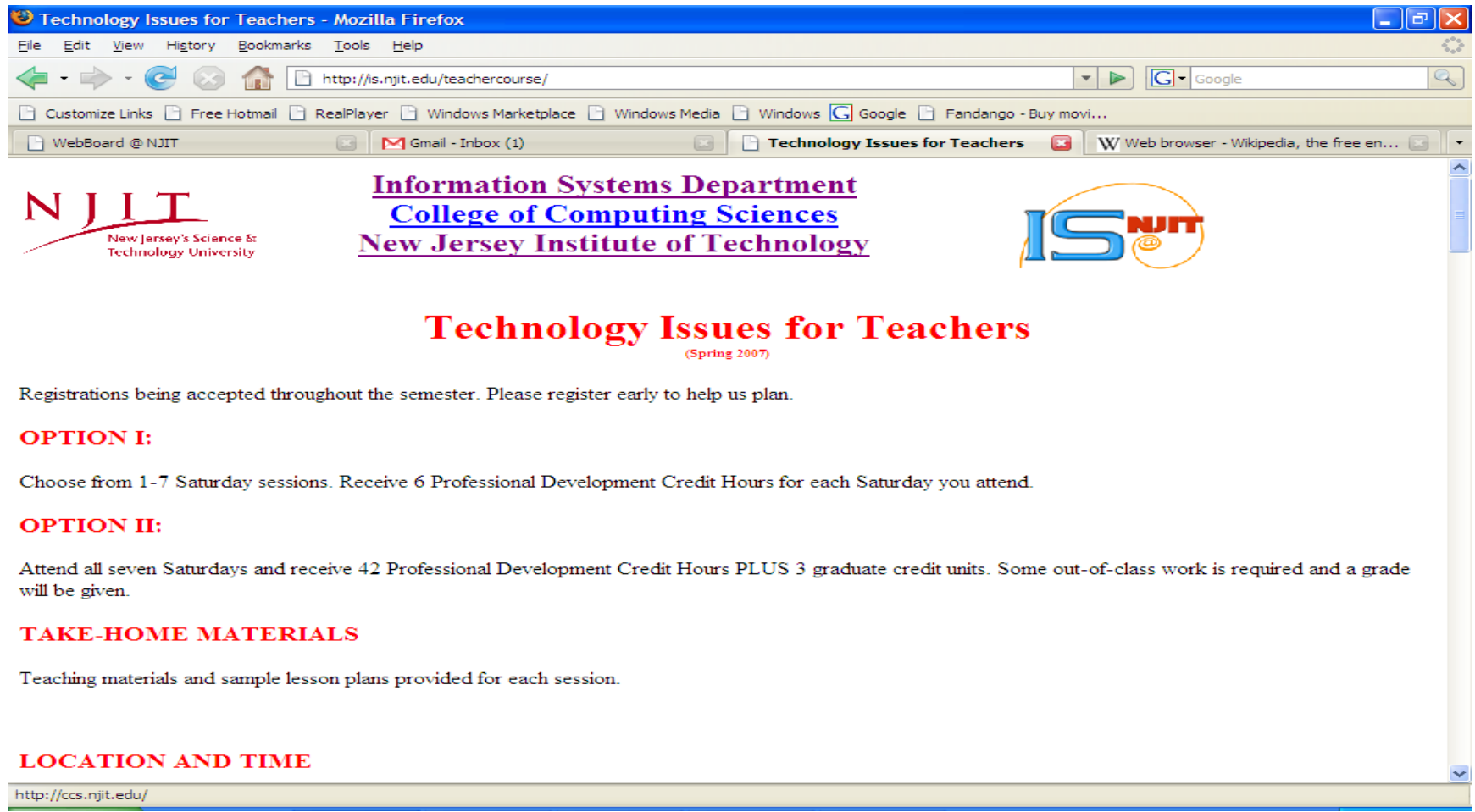

EFFECTIVE WEB SEARCHING



WEB BROWSERS – WHAT THEY ARE AND HOW TO USE THEM

- “A **web browser** is a software application that enables a user to display and interact with text, images, and other information typically located on a website on the World Wide Web or a local network.”(wikipedia.org)
- Web browsers allow a user to quickly access information on many web pages by traversing the links on the web page.

AN EXAMPLE OF A WEB BROWSER (MOZILLA FIREFOX), DISPLAYING OUR COURSE WEB SITE



WEB BROWSERS: OVERVIEW OF BASICS

- URLs (Uniform Resource Locator)
 - Address of an Internet file
 - Example: <http://is.njit.edu/teachercourse/>
 - http is the protocol
 - is.njit.edu is the server or domain
 - Teachercourse.htm is the name of the page
- Most browsers allow you to:
 - Store “favorite places” or “bookmark” pages.
 - Block pop-ups

POPULAR WEB BROWSERS

- Internet Explorer
- Mozilla Firefox
- Opera

WEB BROWSERS: DEMONSTRATION AND PRACTICE

- Demonstration of
 - “surfing” to a web site
 - Clicking on links to traverse pages
 - Using the Back and Forward arrows
 - Using Bookmarks (or Favorite Places)
 - Using the drop-down history menu
- Now YOU try it!!

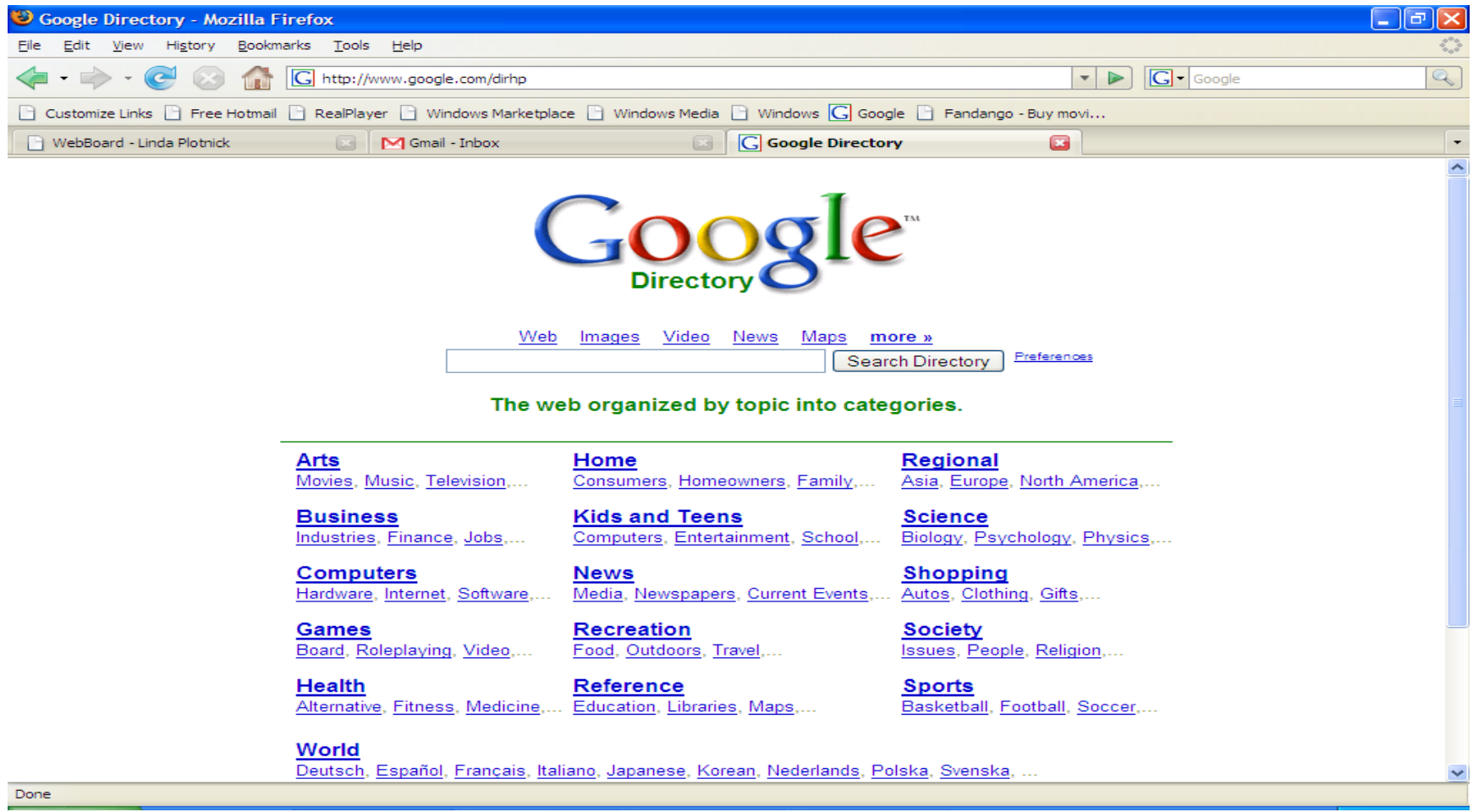
WHERE TO SEARCH ON THE WEB

- Directories
- Digital Libraries
- Search Engines

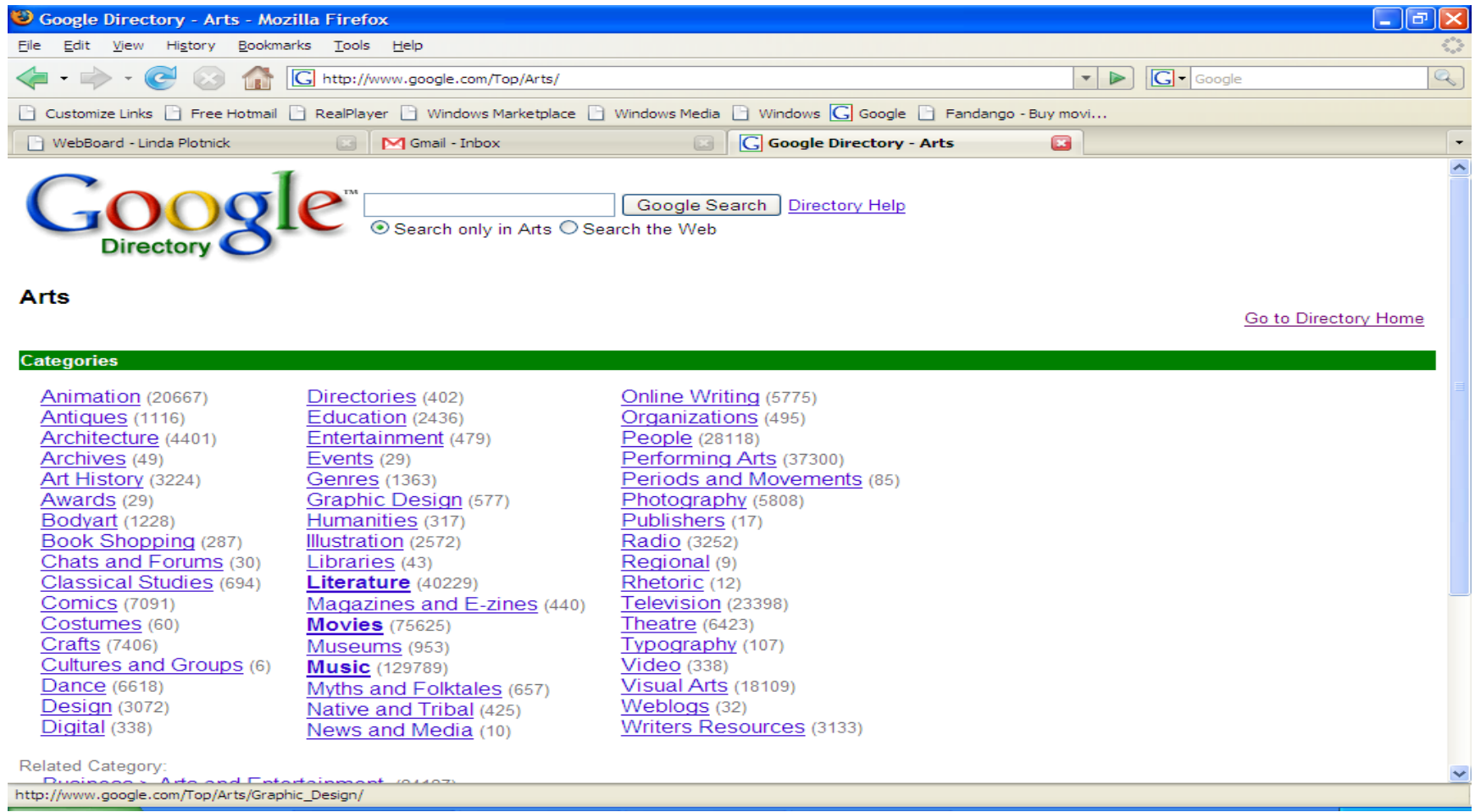
WHAT IS A DIRECTORY?

- “A **web directory** is a directory on the World Wide Web. It specializes in linking to other web sites and categorizing those links.” (wikipedia.org)
- It is NOT a search engine.
- It lists web sites by category and subcategory.
- Inclusion in the directory is usually based on the entire web site not just on key words.
 - Sometimes site owners submit their sites for inclusion.

WEB DIRECTORIES – EX. GOOGLE DIRECTORY



GOOGLE DIRECTORY – CATEGORY ARTS



GOOGLE DIRECTORY – LET’S EXPLORE ARTS

- Go to www.google.com
- Click on **more**
- Click on **even more**
- Click on **directory**
- Choose the category “arts” by clicking
- As a class, let’s explore!!!

DIGITAL LIBRARIES

- “A **digital library** is a library in which a significant proportion of the resources are available in machine-readable format (as opposed to print or microform), accessible by means of computers.” (wikipedia)
- Some of the largest are completely digital.
- Can have open or restricted access.
- Many public libraries and universities have digital libraries of journal articles.

DIGITAL LIBRARIES - EXAMPLE

- National Science Digital Library
 - ❑ www.NSDL.org
 - ❑ Can browse by category, search, search by resource format and/or by grade level.
 - ❑ For STEM teachers
 - ❑ Has resources for teachers including lesson plans
 - Search for lesson plans within
 - ❑ General search
 - ❑ Pathways – “top picks” of sites with resources
 - ❑ Within browsing by topic
 - ❑ Let's try it out!!

WHAT IS A SEARCH ENGINE ?

■ Definition

1. A software program that searches a database and gathers and reports information that contains or is related to specified terms.
2. A website whose primary function is providing a search engine for gathering and reporting information available on the Internet or a portion of the Internet.

"search engine." *The American Heritage® Dictionary of the English Language, Fourth Edition*. Houghton Mifflin Company, 2004. 10 Feb. 2007. <Dictionary.com http://dictionary.reference.com/browse/search_engine>.

WHAT DOES A SEARCH ENGINE DO?

- What does a search engine NOT do?
 - Search engines do not actually search the web directly.
- What do search engines DO?
 - They search a database of the full text of web pages selected from the billions of web pages on the web.
 - BUT when you click on links provided in the search results, you retrieve the current version of the page
 - Some search engines allow you to view the cached page – i.e. the page that was stored in the database.

SEARCH ENGINES: HOW THEY WORK



- Search engine databases are built and contents selected by **spiders**.
 - Spiders are computer robot programs.
 - They find new pages to include in the database by following links in pages they already have (crawling the web).
 - New pages are sent to other programs that save them to the database.
- New pages can also be added by human request to the search engine company.

THE INVISIBLE WEB

- There are web pages that are excluded from general search engine databases either:
 - By policy or
 - Because the spiders cannot access them.
- Estimated to be two to three times bigger than the visible web.
 - Example
 - Academic papers in specialized, password protected databases.
 - Google Scholar finds only the citation to the paper.
 - Dynamically generated web pages for single use.
 - i.e. search results pages

WHAT IS A META SEARCH ENGINE?

- Meta search submit your query to several individual search engines.
 - You get back results from all the search engines
- Meta search engines do not own their own databases.
- Example: www.MetaCrawler.com searches
 - Google
 - Yahoo!
 - Msn Search
 - Ask
- www.clusty.com

POPULAR SEARCH ENGINES

- Google (www.google.com)
- Yahoo! Search (www.yahoo.com)
- Windows Live (www.live.com)
- Ask (www.ask.com)
- Excite (www.excite.com)
 - Partnership with AOL
- Clusty (www.clusty.com)

SEARCH RESULTS – WHO'S ON FIRST?

- Search results are ranked
- Each engine has its own closely guarded algorithm for ranking but usually depends on:
 - ❑ Its relevance to the words in the query
 - ❑ Its overall link popularity
 - ❑ Whether it is being penalized for excessive SEO
- SEO – search engine optimization
 - ❑ Techniques used by owners of pages to force a higher ranking.

SEARCH RESULTS – BEWARE THE SPONSORS

- Some sites are “sponsors” and pay to appear on a search page when a query has a specific keyword or phrase.
 - Google makes it clear which are sponsors and which are search results.
 - Other search engines don’t make it as clear even when they are labeled as “sponsored links”.
 - Example:
 - Google vs. AltaVista

GOOGLE SEARCH RESULT

- Go to *www.google.com* and search on the word **dolphin**
 - Find the sponsored links
- Go to *www.clusty.com* and search on the word **dolphin**
 - Find the sponsored links

SEARCHING – USING KEYWORDS

- Try this:
 - Go to *www.clusty.com* and search by typing in the words – **to be or not to be**
 - How many results did you get?
 - How relevant are they?
 - Now try it with *www.google.com*
 - What difference is there?
 - Some search engines are better than others at recognizing phrases that aren't specifically designated as phrases.

DOES WORD ORDER COUNT?

- Sometimes
 - Try this – go to www.google.com and type in (without any quotes) the words – **information systems**
 - How many results did you get?
 - Now search for – **systems information**
 - How many results did you get?
- Word order matters less today with most search engines than it did in the past.

WHAT TO DO? SIMPLE SEARCH TECHNIQUES

- Use quotes when word order matters or to indicate a phrase.
 - Try typing “to be or not to be” in a search in clusty.
 - Is there a difference?
- Use Boolean logic to indicate when key words must be included or excluded in the search.

BASIC SEARCH OPERATORS

- Google by default consider all the words in a search. However, here are some alternatives:
 - ❑ OR
 - Will return a page if at least one key word is found
 - ❑ NOT (-)
 - Will not return any page in which the keyword is found
 - ❑ Fill in the blanks (*)
 - placeholder for any unknown term(s) and then find the best matches.

BOOLEAN LOGIC - PRACTICE

- Go to www.google.com
 - ❑ Search for the following and note how many results are returned
 - Obama voted * on the * bill
 - ❑ 82,100,000 results
 - jazz OR Brecker
 - ❑ 1,070,000 results
 - Barry Bonds -trial
 - ❑ 6,200,000 results
 - ❑ Michael Brecker was a famous jazz saxophonist who recently passed away. Barry Bonds is a famous MLB player.

PROXIMITY

- Some search engines allow you to look for pairs of words in close proximity to each other.
- What if you wanted to look for information about bass guitars?
- See the number of results you get:
 - ❑ Go to www.altavista.com
 - ❑ Search for “bass guitar”
 - ❑ Search for bass OR guitar
 - ❑ Search for bass NEAR guitar
 - ❑ Search for bass AND guitar

FEATURES OF SEARCH ENGINES

- Search engines feature chart handout
 - Source
 - <http://www.searchengineshowdown.com/features/>
- Search engines by search features
 - Source
 - <http://www.searchengineshowdown.com/features/byfeature.shtml>
- Search engines reviews
 - <http://searchengineshowdown.com/reviews/>

MORE ADVANCED SEARCHING WITH GOOGLE

- Synonym search ~

- Example: search on ~food

- Returns matches to food, nutrition, cooking

- Number ranges in search ..

- search on digital camera..\$300

- Find digital cameras priced at or below \$300

- Search on “Babe Ruth” 1921..1935

- Find pages mentioning Babe Ruth between 1921 and 1935

MORE ADVANCED SEARCHING WITH GOOGLE

■ **intitle**

- ❑ Search for terms to occur in the official title of the page.
- ❑ Example: search for **intitle:ethics computing**
 - Finds pages that mention the word “ethics” in their title, and mention the word “computing” anywhere in the document (title or no).

■ **site**

- ❑ Search for terms in a particular site
- ❑ Example: search for **site:www.whitehouse.gov “global warming”**
 - Finds pages about global warming on the official White House site only.

MORE ADVANCED SEARCHING WITH GOOGLE

■ filetype

- specify the type of file to be returned
- Example: search for **filetype:ppt site:edu ethics**
 - Returns PowerPoint presentations on U.S. university websites about ethics

■ define

- Find web-based definitions
- Example: search for **define:cyberinfrastructure**
 - Returns ranked definitions for the term on the web.

MORE ADVANCED SEARCHING WITH GOOGLE

■ Calculator and Conversions

□ Mathematical functions

- +, -, *, %, of, nth root of, sqrt(nn), ^ for exponentiation
 - Example: search for **sqrt((125+25)-50)**

□ Advanced Math

- Trig: sin, cos, tan, sec, csc, cot, etc.
- Inverse trig: arcsin, arcos, artan, etc.
- Hyperbolic trig: sinh, cosh, tanh, etc.
- Logarithms
- Factorials
- Etc.
 - Example: search for **arcsin(sin(1))**

MORE ADVANCED SEARCHING WITH GOOGLE

■ Calculator and Conversions

□ Units of measure and conversions

- Mass: kilogram or kg, grams or g, pounds or lbs, etc.
- Length: meters or m, miles, feet, inches, Angstroms, etc.
- Volume: gallons, liters or l, bushels, teaspoons, pints, etc.
- Area: square miles, square kilometers, acres, etc.
- Time: day, seconds or s, years, centuries, fortnights, etc.
- Electricity: volts, amps, ohms, etc.
- Energy: Calories, British thermal units or BTU, joules, etc.
- Power: watt, kilowatts, horsepower or hp, etc.

MORE ADVANCED SEARCHING WITH GOOGLE

- Calculator and Conversions, continued
 - Units of Measure and Conversion, continued
 - Information: bits, bytes, kbytes, etc.
 - Quantity: dozen, baker's dozen, percent, gross, etc.
 - Numbering System: decimal, hexadecimal or hex, etc.
 - Quick Facts: currency in countries, population, biographies.
 - Example: search for
 - 1 lb in carats returns the number of carats in a pound
 - currency in Panama returns the unit of currency in Panama.

ADVANCED SEARCHING IN GOOGLE

- Google has advanced searching.
- Go to www.google.com and click on “Advanced search”.
 - Form to specify search parameters and limitations
 - Search for pages that link to a specified page
 - Search for pages that are similar to a specified page.
- Let's explore together!!

USING SEARCH ENGINES TO DETECT PLAGIARISM

- There are applications that can detect plagiarism.
 - Example – turnitin.com
- When can't you use them?
 - They are usually subscription based – your institution may not have one available.
 - The papers submitted by students are not submitted electronically.

USING SEARCH ENGINES TO DETECT PLAGIARISM

■ Scenario

- ❑ Your student has submitted a paper and your judgment tells you that it might be plagiarized.
- ❑ The writing is at a much higher level than the student usually produces.
- ❑ You cannot use a plagiarism detection program.
 - Use search techniques to ascertain if the paper or sections of the paper were plagiarized.

DETECTING PLAGIARISM – AN EXAMPLE

- The following paragraph in a paper your student hands in is suspicious.
- Try to find if it is taken from a page available on the web:
 - *Up until now we have been talking at length about how the IS has to be designed to meet the organization's goals. That has meant, in this current environment, to design across organizational lines for processes not for individual functional entities in the organization. Yet, the decoupling of the subsystems would seem to me to defeat that design for process goal. Is it the integration that IS can provide between subsystems that keeps the IS process oriented? IS acts as a coordinator, provides an interface. I think that, for example, the clustering of subsystems described would enable a process subsystem to be developed.*

DETECTING PLAGIARISM

■ Hints

- ❑ Use quotes around suspicious phrases and/or sentences.
- ❑ Try different string lengths.
- ❑ Remember – there are other sources than the web.
 - Not finding evidence on the web is not a guarantee that the work is original.
- ❑ Students may plagiarize complete sections from one source or many sources.

EVALUATING SEARCH RESULTS

- Is it relevant?
- Who wrote the page?
 - Is the author or organization identified?
 - What are their credentials?
 - Is the article likely to be biased?
 - Example – is it an article on the dangers of smoking written by a tobacco company?
- What type of domain does it come from?
 - .edu
 - .com
 - .gov
 - .mil
 - .org

EVALUATING SEARCH RESULTS

- Who published the page?
 - Usually the name in the URL between http:// and the first /
 - Ex. http://www.nytimes.com/
 - Is this a reputable entity?
 - If there is an “About” page – it can provide information on the entity publishing the page.
- Is the information current?
 - Can you identify when the page was created or the last update made?

EVALUATING SEARCH RESULTS

■ Documentation

- Are the claims documented with references?
 - Footnotes and links

■ Delivery

- Are there grammatical errors?
- Are there obvious factual errors?

EVALUATING SEARCH RESULTS

- What do other websites and people say?
 - Use <http://alexa.com> to see gather information on
 - Traffic rank
 - Subjective reviews
 - Statistics including sites that link to the page
 - Look up the page in a reputable directory.
 - Perform a search on the author or organization.
- Evaluate the page using the form found at http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/EvalForm_General_Barker.pdf