

Searching the Web

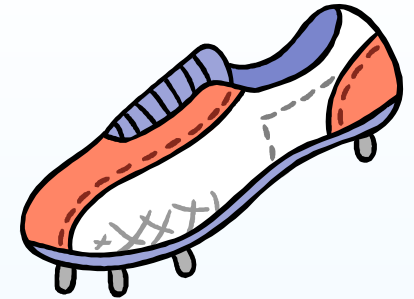
- The Problems
- Search Engines
- Directories
- Invisible Web
- How To Search



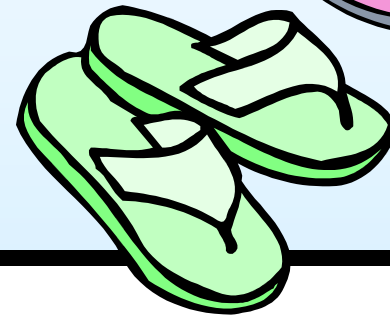
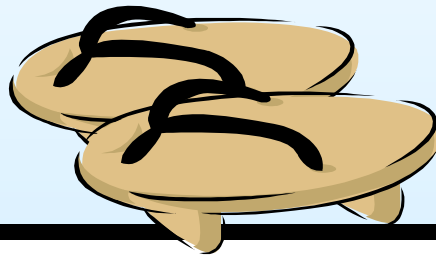
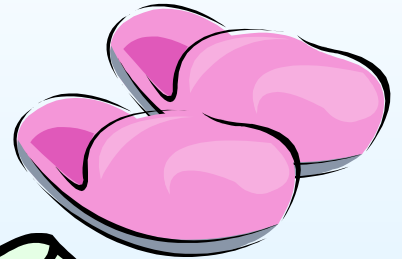
Problems



Problems

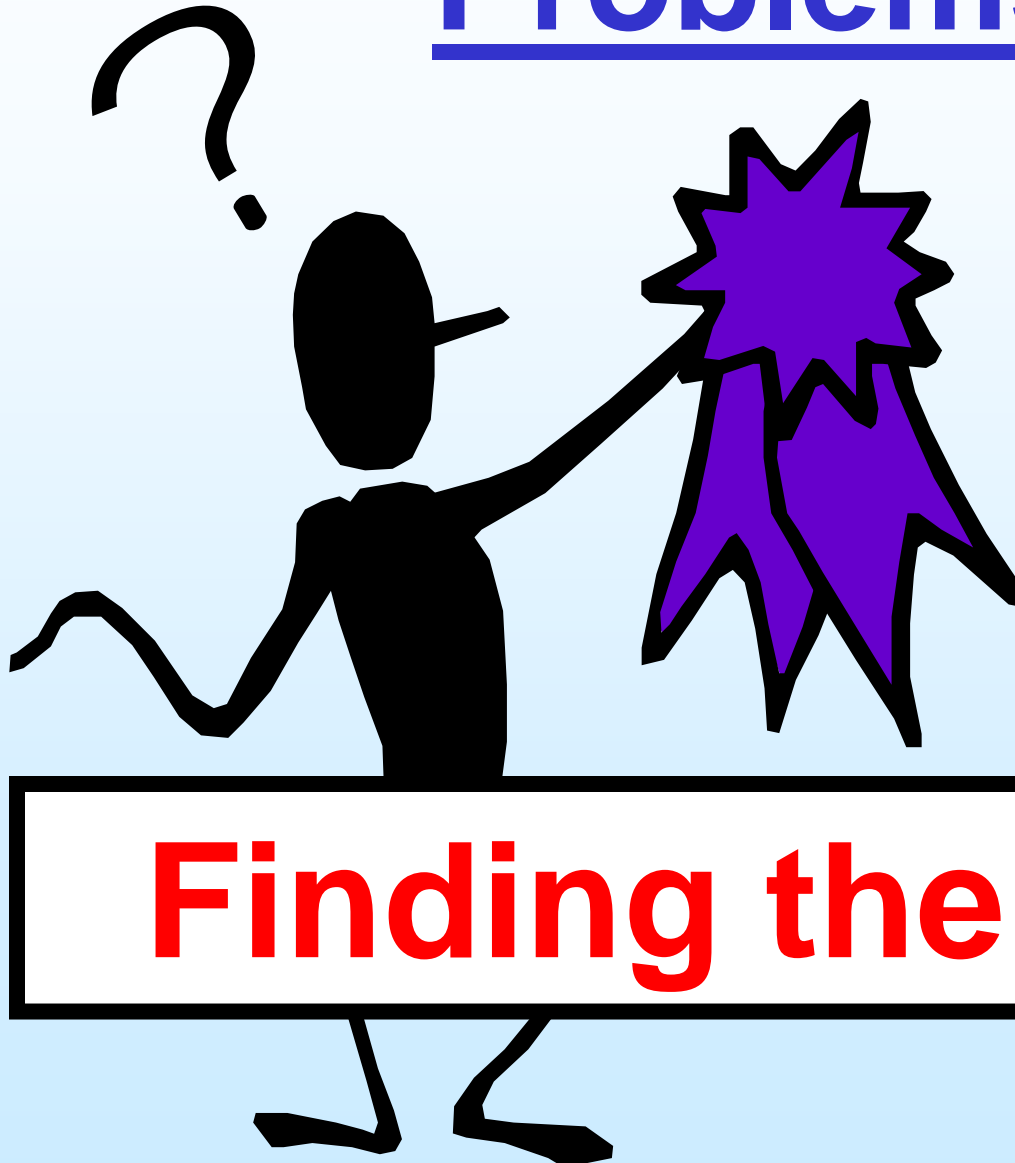


I NEED
RUNNING SHOES



Finding exactly
what you need.

Problems



Finding the best

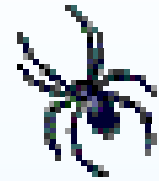
Other Problems

- **Search engines and directories miss a lot!**

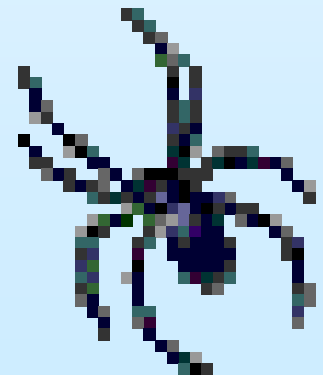
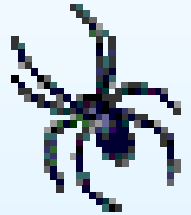
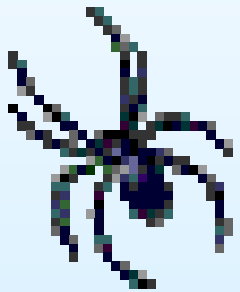
(Research by the NEC Research Institute in 1999 showed that the eleven top search robots together only covered about 42% of the Web)

- **Not all search engines search all formats** (.pdf for example)

Search Engines



- **Send out “spiders” to search Web pages** (URL and first lines or page)
- **Mechanical – no human touch**
- **Use keyword searching**
- **No browsing, no subject categories**



Search Engines

- **Consists of three parts:**
 - Spider**
 - Index – copy of each page gathered**
 - Program to search the index**
- **Results are usually ranked by relevancy (search terms in title, URL, heading, HTML META tag, or # of times term appears on page, term appearing in first part of page, page popularity, etc.)**

Recommended

★ **Best: Google**

Others:

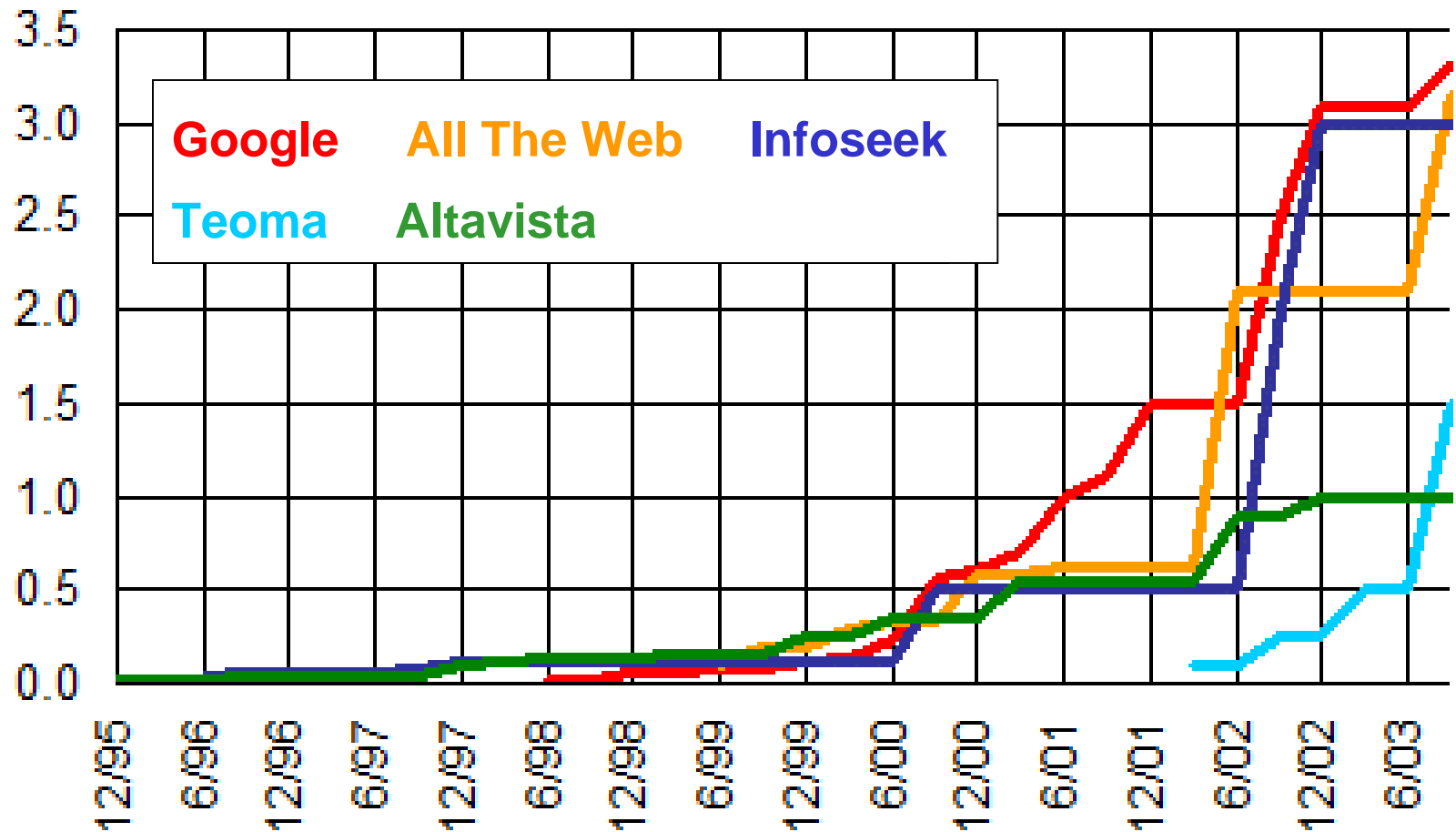
★ **Alta Vista Advanced Search**

★ **Alltheweb**

★ **Teoma**

★ **Northern Light**

Billions Of Textual Documents Indexed December 1995-September 2003



<http://searchenginewatch.com/reports/article.php/2156481#key>

Meta-Search Engines

- **Search several search engines at once**
- **Not recommended because they retrieve only about 10% of search results so may be very relevant, or may miss too much**
- **Good ones remove duplicate hits**

Meta- Search Engines

★ **Metacrawler**

★ **Ixquick**

★ **Copernic**

★ **SurfWax**

Subject Directories

- **Selected by humans (sometimes experts in a subject)**
- **May be evaluated and annotated**
- **May not be current**
- **Can browse and search using broad terms**

Recommended

★ **Librarians' Index**

★ **Infomine**

★ **AcademicInfo**

★ **About.com**

★ **MeL Internet Collection**

The Invisible Web

(Searchable Databases)

- **Specialized databases**
- **Are searched through a search box on the Web page**
- **2-3 times as big as the visible web**
- **Can be located through subject directories or keyword searching in some search engines**

Our Databases

- **Are available by subscription and need password or IP recognition**
- **Contain magazine articles, reference books, pictures, TV & radio scripts,...**
- **Are of high quality and reliable**
- **Are one of the first places to look when doing school research**

Michigan eLibrary Databases

- ★ **Gale Group (InfoTrac)**
- ★ **Michigan Newspapers**
- ★ **eLibrary Elementary**
- ★ **First Search**
- ★ **SIRS Discover Deluxe**
- ★ **LearnATest**
- ★ **netLibrary (e-books)**

How Do
You
Search?



Be specific

- **Search for:**

Windows 98 bugs (*not just Windows*)

- **Or try searching for the exact problem:**

I can't install a USB device in Windows 98

Use lower case

- If a Web page uses capital letters or capitalized words, a Web search may be case sensitive and will only return words that match exactly the capitalization.
- Lower case words will match words no matter how they're capitalized.

Use Phrase Searching

- **Putting your words in quotation marks will make the search engine look for those words in the exact order you enter them.**

How do you search?

school dress code

acid rain

child abuse

gun control

affirmative action

domestic violence

“school dress code”

“acid rain”

“child abuse”

“gun control”

“domestic violence”

“affirmative action”

Use Phrase Searching

To search for a proper name or title:

- Use quotation marks
“michigan state university”
- Capitalize the initial letters of a name
Michigan State University
- Use the underscore
michigan_state_university

How do you search?

**You want to find the Web site for
the World Health Organization**

- *World Health Organization*
- *“world health organization”*
- *world_health_organization*



Use the + sign

- The + sign means all the words must appear on the Web page.

+ windows +98 +bugs

(Note: There is no space after the + sign.)

- The Boolean Operator AND may also work.

How do you search?

You want to reserve a camping space in California's Yosemite National Park.

**+yosemite +national +park
+camping +reservations**



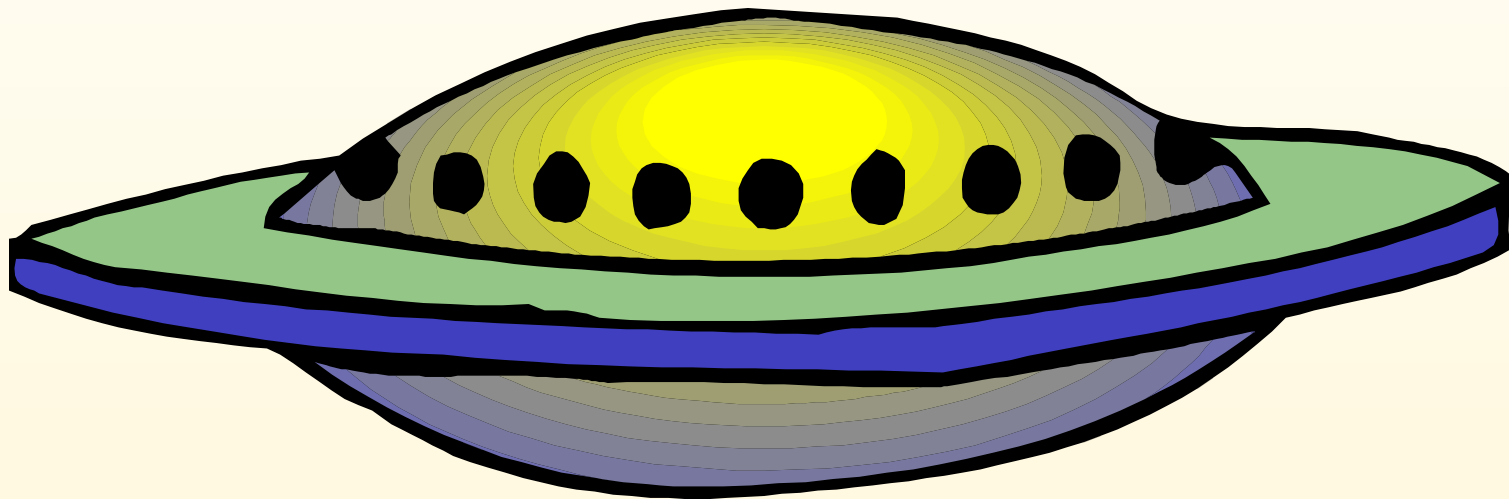
Use the - sign

- The – sign means those words will NOT appear on the Web page.
- Search for **Windows -98 -3.1**
(Note: There is no space after the - sign.)
- The Boolean Operator NOT may also work.

How do you search?

You are a fan of the original Star Trek series and don't want pages about Voyager, Deep Space Nine or Star Trek: The Next Generation.

**star trek – voyager – deep – space –
next -generation**



Combine symbols

“star trek” –voyager –”deep space nine” –
”next generation”

Use Boolean Operators

- **AND** is used to include all the words
(Celtic **AND** legends)
- **NOT** is used to exclude words
(Celtic **NOT** basketball)
- **OR** is used to find the synonyms of a word
(legends **OR** myths)

(Note: Boolean Operators must be capitalized)

Use Wild Cards...

- Wild cards let you substitute symbols for letters in words if you're not sure how to spell a word, if the word can be spelled in different ways, or if you want to search all forms of a word.
- The wildcards are *, ?, and !

Use * (truncation)

- An asterisk (*) stands for any number of characters, including none.
- Searching **swim*** would search for **swim, swimmer, swimming, swims...**

(Note: Some search engines automatically search all forms of a word.)

(Note: The asterisk can be used within a word.)

Use ?

- The question mark (?) stands for exactly one character when you are unsure of how to spell a word.
- Not sure if it's relevance or relevation? Use **relev?nce**

(Note: Use can use multiple question marks in a row, but each one stands for just one letter.)

Use !

- The exclamation point (!) stands for one or no characters and is useful if you want to search singular and plural, but not other forms of a word.
- Search **product!** for **product** and **products** but not **productive** or **productivity**

Use OR

- Use for synonyms – **cars OR automobiles**
- Use for variant spellings – **theater OR theatre** (could also use truncation)

Synonyms?

- **Air = Atmosphere**
- **Dirt = Soil**
- **Bushes = Shrubbery**
- **Creek = Stream**

How do you search?



What is the history
of the dog breed
chihuahua?

1. **history AND dog* AND chih*a**
2. **+history +dog* +chihuahua –city**
3. **(history OR pedigree OR lineage)
AND (dog* OR canine* OR breed*)
AND (chihuahua OR “mexican
hairless”)**

Other tricks

Try Another

No search engine indexes all the Web. If you're not finding what you want, try another search engine.

(Start with Google!)

Other tricks

- **Use advanced search**
- **Use help features**

Google Advanced Search

Google Advanced Search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Google™ [Advanced Search](#) [Advanced Search Tips](#) | [All About Google](#)

Find results with **all** of the words 10 results

with the **exact phrase**

with **at least one** of the words

without the words

Language Return pages written in

File Format return results of the file format

Date Return web pages updated in the

Occurrences Return results where my terms occur

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

SafeSearch No filtering Filter using [SafeSearch](#)

Google Advanced Search

- Think of all the words that would always appear on the perfect page. Put them in the **“With All Words”** box.
- Think of any synonyms which might be used and put them in the **“With Any of the Words”** box.
- Think of all the things you don’t want. Put those words in the **“Without”** box.

How do you search?

You're interesting in the legendary lost continent of Atlantis. There have been several movies with Atlantis in the title, but you don't want information on them. You also don't want anything on the space shuttle Atlantis.

- **WITH ALL:** *Atlantis continent*
- **WITHOUT:** *shuttle film movie*



Google Advanced Search

- Instead of using quotation marks, use Google's Advanced Search box **“With the Exact Phrase”**

How do you search?

You may be interested in attending San Diego State University. How can you search for it without getting information on other universities?

- **EXACT PHRASE:** *San Diego State University*



Google Advanced Search

Google Advanced Search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Date	Return web pages updated in the	anytime
Occurrences	Return results where my terms occur	anywhere in the page
Domain	<input type="text" value="Only"/> return results from the site or domain	<input type="text"/>
		e.g. google.com, .org More info
SafeSearch	<input checked="" type="radio"/> No filtering <input type="radio"/> Filter using SafeSearch	

Page-Specific Search

Similar	Find pages similar to the page	<input type="text"/>	Search
		e.g. www.google.com/help.html	
Links	Find pages that link to the page	<input type="text"/>	Search

Topic-Specific Searches

Google Advanced Search

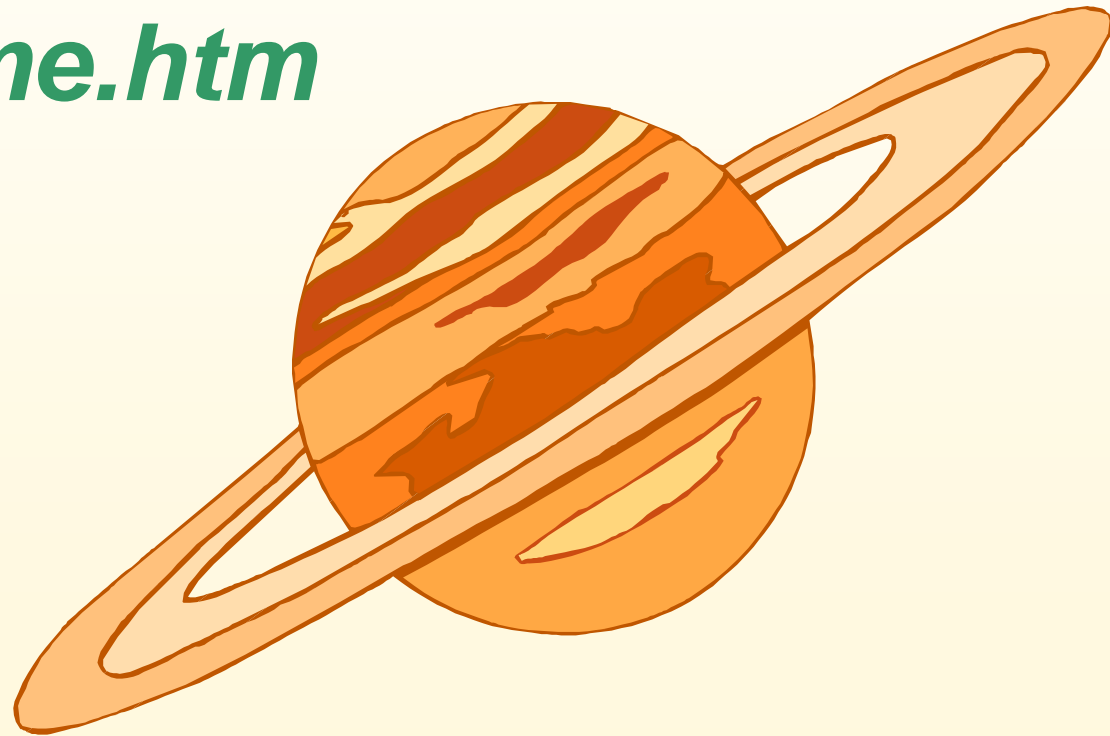
- You can use **“Similar To”** to find pages that are linked to a page you like.
- You can use **“Linked To”** to find pages that link to the page you’re using (useful in evaluating Web sites).

How do you search?

You found NASA's *Welcome to the Planets Web* page helpful and would like to see more planet Web sites.

- **SIMILAR:**

*[http://pds.jpl.nasa.gov/planets/
welcome.htm](http://pds.jpl.nasa.gov/planets/welcome.htm)*



Resources

- **To compare search engine features:**
<http://www.searchenginewatch.com/facts/ataglance.html>
- **OR**
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html> - Phrase
- **OR**
<http://infopeople.org/search/chart.html>

Resources

- **Tutorials on using Boolean and other searching techniques:**

<http://lib.nmsu.edu/instruction/eval.html>

<http://library.albany.edu/internet/boolean.html>

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>

Resources

- **Sample Searches:**

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/BestStuffExercises.pdf>

- **To become a good Google searcher:**

<http://www.edrenplanners.com/infolit/searchactivities/google.html>

- **Comparing good & bad Web sites:**

<http://lib.nmsu.edu/instruction/evalexpl.html>

Bibliography

- **Search Engine Watch**
<http://searchenginewatch.com>
- **Tilburg University Web Tutorial**
<http://www.tilburguniversity.nl/services/library/instruction/www/onlinecourse/>
- **The Good, the Bad, and the Ugly**
<http://lib.nmsu.edu/instruction/eval.html>
- **UC Berkeley's Internet tutorials**
<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>

Bibliography

- **The WebQuest Page**
<http://webquest.sdsu.edu/>
- **Grammar of the Internet Information Literacy**
<http://www.edrenplanners.com/infolit/>
- **University at Albany's Internet Tutorials**
<http://library.albany.edu/internet/>



Happy Searching!

Lapeer East High School Library

April 2005