

SEARCHING THE INTERNET

SEARCHING THE INTERNET

- The starting point for researching material on the internet is a **search engine**.
- A Search engine is a database made up of web page files assembled automatically.
- There are two kinds of search engines:
 - Individual search engines**
 - Metasearch engines**

SEARCH ENGINES

- Search engines do not really search the World Wide Web directly. Each one searches a database of web pages that it has harvested and cached. When you use a search engine, you are always searching a somewhat stale copy of the real web page. When you click on links provided in a search engine's search results, you retrieve the current version of the page.

SEARCH ENGINES

- Search engine databases are selected and built by computer robot programs called spiders. These "crawl" the web, finding pages for potential inclusion by following the links in the pages they already have in their database. They cannot use imagination or enter terms in search boxes that they find on the web.

SEARCH ENGINES

- If a web page is never linked from any other page, search engine spiders cannot find it. The only way a brand new page can get into a search engine is for other pages to link to it, or for a human to submit its URL for inclusion. All major search engines offer ways to do this.

SEARCH ENGINES

- After spiders find pages, they pass them on to another computer program for "indexing." This program identifies the text, links, and other content in the page and stores it in the search engine database's files so that the database can be searched by keyword and whatever more advanced approaches are offered, and the page will be found if your search matches its content.

SEARCH ENGINES

- Many web pages are excluded from most search engines by policy. The contents of most of the searchable databases mounted on the web, such as library catalogs and article database, are excluded because search engine spiders cannot access them. All this material is referred to as the "Invisible Web" -- what you don't see in search engine results.

SPIDERS

- **SPIDERS** Computer robot programs, referred to sometimes as "crawlers" or "knowledge-bots" or "knowbots" that are used by search engines to roam the World Wide Web via the Internet, visit sites and databases, and keep the search engine database of web pages up to date. They obtain new pages, update known pages, and delete obsolete ones. Their findings are then integrated into the "home" database. Most large search engines operate several robots all the time.

METASEARCH ENGINES

- In a meta-search engine, you submit keywords in its search box, and it transmits your search simultaneously to several individual search engines and their databases of web pages. Within a few seconds, you get back results from all the search engines queried. Meta-search engines do not own a database of Web pages; they send your search terms to the databases maintained by search engine companies.

SEARCHING STRATEGIES

- What do I want to do?
 1. Browse?
 2. Locate a specific piece of information?
 3. Retrieve everything I can on the subject?
- Your answer will determine how you conduct your search and what tools you will use.

SEARCHING STRATEGIES

1. If you're browsing and trying to determine what's available in your subject area, start out by selecting a subject directory like Yahoo! Then, enter your search keyword(s) into one of the metasearch engines, such as Vivisimo, just to see what's out there.
2. If you're looking for a specific piece of information, go to a major search engine such as Google, or to a specialized database.
3. If you want to retrieve everything you can on a subject, try the same search on several search engines. Also, don't forget to check resources off the Web, such as books, newspapers, journals and other print reference sources.

QUICK TIPS

- Use the plus (+) and minus (-) signs in front of words to force their inclusion and/or exclusion in searches.
EXAMPLE: +meat -potatoes
(NO space between the sign and the keyword)
- Use double quotation marks (" ") around phrases to ensure they are searched exactly as is, with the words side by side in the same order.
EXAMPLE: "bye bye miss american pie"
(Do NOT put quotation marks around a single word.)
- Put your most important keywords first in the string.
EXAMPLE: dog breed family pet choose

QUICK TIPS

- Type keywords and phrases in lower case to find both lower and upper case versions. Typing capital letters will usually return only an exact match.

EXAMPLE: *president* retrieves both *president* and *President*

- Use truncation (or stemming) and wildcards (e.g., *) to look for variations in spelling and word form.

EXAMPLE: *librar** returns *library*, *libraries*, *librarian*, etc.

EXAMPLE: *colo*r* returns *color* (American spelling) and *colour* (British spelling)

QUICK TIPS

- Combine phrases with keywords, using the double quotes and the plus (+) and/or minus (-) signs.

EXAMPLE: *+cowboys +"wild west" -football -dallas*
(In this case, if you use a keyword with a +sign, you must put the +sign in front of the phrase as well. When searching for a phrase alone, the +sign is not necessary.)

- When searching within a document for the location of your keyword(s), use the "find" command on that page.

QUICK TIPS

- Know the default (basic) settings your search engine uses (OR or AND). This will have an effect on how you configure your search statement because, if you don't use any signs (+, -, " "), the engine will default to its own settings .

QUICK TIPS

- Know whether or not the search engine you are using maintains a stop word list. If it does, don't use known stop words in your search statement. Also, consider trying your search on another engine that does not recognize stop words.

BOOLEAN OPERATORS

- Boolean logic takes its name from British mathematician George Boole (1815-1864), who wrote about a system of logic designed to produce better search results by formulating precise queries. He called it the "calculus of thought." From his writings, we have derived Boolean logic and its operators: AND, OR, and NOT, which we use to link words and phrases for more precise queries.

BOOLEAN "AND"

- The Boolean AND actually narrows your search by retrieving only documents that contain every one of the keywords you enter. The more terms you enter, the narrower your search becomes.

EXAMPLE: truth AND justice

EXAMPLE: truth AND justice AND ethics AND
congress

BOOLEAN "OR"

- The Boolean OR expands your search by returning documents in which either or both keywords appear. Since the OR operator is usually used for keywords that are similar or synonymous, the more keywords you enter, the more documents you will retrieve.

EXAMPLE: college OR university

EXAMPLE: college OR university OR institution
OR campus

BOOLEAN "NOT" / "AND NOT"

- The Boolean NOT or AND NOT (sometimes typed as ANDNOT) limits your search by returning only your first keyword but not the second, even if the first word appears in that document, too.

EXAMPLE: saturn AND NOT car

EXAMPLE: pepsi AND NOT coke

CREATING A SEARCH STATEMENT

- When structuring your query, keep the following tips in mind:
 1. Be specific
EXAMPLE: Hurricane Hugo
 2. Whenever possible, use nouns and objects as keywords
EXAMPLE: fiesta dinnerware plates cups saucers
 3. Put most important terms first in your keyword list; to ensure that they will be searched, put a +sign in front of each one
EXAMPLE: +hybrid +electric +gas +vehicles
 4. Use at least three keywords in your query
EXAMPLE: interaction vitamins drugs

CREATING A SEARCH STATEMENT

5. Combine keywords, whenever possible, into phrases
EXAMPLE: "search engine tutorial"
6. Avoid common words, e.g., water, unless they're part of a phrase
EXAMPLE: "bottled water"
7. Think about words you'd expect to find in the body of the page, and use them as keywords
EXAMPLE: anorexia bulimia eating disorder
8. Write down your search statement and revise it before you type it into a search engine query box
EXAMPLE: +"south carolina" +"financial aid" +applications +grants