

## **BACK / FORWARD**

Buttons in most browsers' Tool Button Bar, upper left. BACK returns you to the document previously viewed. FORWARD goes to the next document, after you go BACK.

If it seems like the BACK button does not work, check if you are in a new browser window; some Web pages are programmed to open a new window when you click on some links. Each window has its own short-term search History. If this does not work, right click on the BACK button to select the page you want (some Web pages are programmed to disable BACK).

## **BLOG or WEB LOG**

A blog (short for "web log") is a type of web page that serves as a publicly accessible personal journal (or log) for an individual. Typically updated daily, blogs often reflect the personality of the author. Blog software usually has an archive of old blog postings. Many blogs can be searched for terms in the archive. Blogs have become a vibrant, fast-growing medium for communication in professional, political, news, trendy, and other specialized web communities. Many blogs provide RSS feeds, to which one can subscribe and receive alerts to new postings in selected blogs.

## **BOOKMARK/FAVORITES**

Way in browsers to store in your computer direct links to sites you wish to return to. Netscape, Mozilla, and Firefox use the term Bookmarks. The equivalent in Internet Explorer (IE) is called a "Favorite." To create a bookmark, click on BOOKMARKS or FAVORITES, then ADD. Or left-click on and drag the little bookmark icon to the place you want a new bookmark filed. To visit a bookmarked site, click on BOOKMARKS and select the site from the list.

You can download a bookmark file to diskette and install it on another computer. In most browsers now, you can do this with an Import... and Export... set of commands which can be found under FILE or in the Manage Bookmarks window's FILE.

## **BOOLEAN LOGIC**

Way to combine terms using "operators" such as "AND," "OR," "AND NOT" and sometimes "NEAR." AND requires all terms appear in a record. OR retrieves records with either term. AND NOT excludes terms. Parentheses may be used to sequence operations and group words. Always enclose terms joined by OR with parentheses.

## **BROWSE**

To follow links in a page, to shop around in a page, exploring what's there, a bit like window shopping. The opposite of browsing a page is searching it. When you search a page, you find a search box, enter terms, and find all occurrences of the terms throughout the site. When you browse, you have to guess which words on the page pertain to your interests. Searching is usually more efficient, but sometimes you find things by browsing that you might not find because you might not think of the "right" term to search by.

## **BROWSERS**

Browsers are software programs that enable you to view WWW documents. They "translate" HTML-encoded files into the text, images, sounds, and other features you see. Microsoft Internet Explorer (called simply IE), Mozilla, Firefox, Safari, and Opera are examples of "graphical" browsers that enable you to view text and images and many other WWW features. They are software that must be installed on your computer.

**CACHE**

In browsers, "cache" is used to identify a space where web pages you have visited are stored in your computer. A copy of documents you retrieve is stored in cache. When you use GO, BACK, or any other means to revisit a document, the browser first checks to see if it is in cache and will retrieve it from there because it is much faster than retrieving it from the server.

**CACHED LINK**

In search results from Google, Yahoo! Search, and some other search engines, there is usually a Cached link which allows you to view the version of a page that the search engine has stored in its database. The live page on the web might differ from this cached copy, because the cached copy dates from whenever the search engine's spider last visited the page and detected modified content. Use the cached link to see when a page was last crawled and, in Google, where your terms are and why you got a page when all of your search terms are not in it.

**CASE SENSITIVE**

Capital letters (upper case) retrieve only upper case. Most search tools are not case sensitive or only respond to initial capitals, as in proper names. It is always safe to key all lower case (no capitals), because lower case will always retrieve upper case.

**CGI**

"Common Gateway Interface," the most common way Web programs interact dynamically with users. Many search boxes and other applications that result in a page with content tailored to the user's search terms rely on CGI to process the data once it's submitted, to pass it to a background program in JAVA, JAVASCRIPT, or another programming language, and then to integrate the response into a display using HTML.

**COOKIE**

A message from a WEB SERVER computer, sent to and stored by your browser on your computer. When your computer consults the originating server computer, the cookie is sent back to the server, allowing it to respond to you according to the cookie's contents. The main use for cookies is to provide customized Web pages according to a profile of your interests. When you log onto a "customize" type of invitation on a Web page and fill in your name and other information, this may result in a cookie on your computer which that Web page will access to appear to "know" you and provide what you want. If you fill out these forms, you may also receive e-mail and other solicitation independent of cookies.

**CRAWLER or WEBCRAWLER**

Same as Spider.

**DOMAIN, TOP LEVEL DOMAIN (TLD)**

Hierarchical scheme for indicating logical and sometimes geographical venue of a web-page from the network. In the US, common domains are .edu (education), .gov (government agency), .net (network related), .com (commercial), .org (nonprofit and research organizations). Outside the US, domains indicate country: ca (Canada), uk (United Kingdom), au (Australia), jp (Japan), fr (France), etc. Neither of these lists is exhaustive.

## **DOMAIN NAME, DOMAIN NAME SERVER (DNS)ENTRY**

Any of these terms refers to the initial part of a URL, down to the first /, where the domain and name of the host or SERVER computer are listed (most often in reversed order, name first, then domain). The domain name gives you who "published" a page, made it public by putting it on the Web.

A domain name is translated in huge tables standardized across the Internet into a numeric IP address unique the host computer sought. These tables are maintained on computers called "Domain Name Servers." Whenever you ask the browser to find a URL, the browser must consult the table on the domain name server that particular computer is networked to consult.

"Domain Name Server entry" frequently appears a browser error message when you try to enter a URL. If this lookup fails for any reason, the "lacks DNS entry" error occurs. The most common remedy is simply to try the URL again, when the domain name server is less busy, and it will find the entry (the corresponding numeric IP address).

## **DOWNLOAD**

To copy something from a primary source to a more peripheral one, as in saving something found on the Web (currently located on its server) to diskette or to a file on your local hard drive.

## **EXTENSION or FILE EXTENSION**

In Windows, DOS and some other operating systems, one or several letters at the end of a filename. Filename extensions usually follow a period (dot) and indicate the type of file. For example, *this.txt* denotes a plain text file, *that.htm* or *that.html* denotes an HTML file. Some common image extensions are *picture.jpg* or *picture.jpeg* or *picture.bmp* or *picture.gif*

## **FAVORITES**

In the Internet Explorer browser, a means to get back to a URL you like, similar to Bookmarks.

## **FEED READER**

A software package that enables you to easily read the XML code in which RSS feeds are written. Bloglines is currently the most popular feed reader but there are many competitors.

## **FIELD SEARCHING**

Ability to limit a search by requiring word or phrase to appear in a specific field of documents (e.g., title, url, link).

## **FIND**

Tool in most browsers to search for word(s) keyed in document in screen only. Useful to locate a term in a long document. Can be invoked by the keyboard command, Ctrl+F.

## **FRESHNESS**

How up-to-date a search engine database is, based primarily on how often its spiders recirculate around the Web and update their copies of the web pages they hold, and discover new ones. Also determined by how quickly they integrate new sites that web authors send to them. Two weeks is about as good as most search engines do, but some update certain selected web sites more frequently, even daily.

**FRAMES**

A format for web documents that divides the screen into segments, each with a scroll bar as if it were as "window" within the window. Usually, selecting a category of documents in one frame shows the contents of the category in another frame. To go BACK in a frame, position the cursor in the frame and press the right mouse button, and select "Back in frame" (or Forward). You can adjust frame dimensions by positioning the cursor over the border between frames and dragging the border up/down or right/left holding the mouse button down over the border.

**FTP**

File Transfer Protocol. Ability to transfer rapidly entire files from one computer to another, intact for viewing or other purposes.

**FUZZY AND**

In ranking of results, documents with all terms (Boolean AND) are ranked first, followed by documents containing any terms (Boolean OR) are retrieved. The farther down, the fewer the terms, although at least one should always be present.

**HEAD or HEADER** (of HTML document)

The top portion of the HTML source code behind Web pages, beginning with <HEAD> and ending with </HEAD>. It contains the Title, Description, Keywords fields and others that web page authors may use to describe the page. The title appears in the title bar of most browsers, but the other fields cannot be seen as part of the body of the page. To view the <HEAD> portion of web pages in your browser, click VIEW, Page Source. In Internet Explorer, click VIEW, Source. Some search engines will retrieve based on text in these fields.

**HISTORY, Search History**

Available by using the combined keystrokes CTRL + H, a more permanent record of sites you have visited/retrieved than GO. You can set how many days your browser retains history in Edit | Preferences, or in Tools | Options.

**HOST**

Computer that provides web-documents to clients or users.

**HTML**

Hypertext Markup Language. A standardized language of computer code, imbedded in "source" documents behind all Web documents, containing the textual content, images, links to other documents (and possibly other applications such as sound or motion), and formatting instructions for display on the screen. When you view a Web page, you are looking at the product of this code working behind the scenes in conjunction with your browser. Browsers are programmed to interpret HTML for display. HTML often imbeds within it other programming languages and applications such as SGML, XML, Javascript, CGI-script and more. It is possible to deliver or access and execute virtually any program via the WWW. You can see HTML by selecting the View pop-down menu tab, then "Document Source."

**HYPertext**

On the World Wide Web, the feature, built into HTML, that allows a text area, image, or other object to become a "link" (as if in a chain) that retrieves another computer file (another Web page, image, sound file, or other document) on the Internet. The range of possibilities is limited by the ability of the computer retrieving the outside file to view, play, or otherwise open the incoming file. It needs to have software that can interact with the imported file. Many software capabilities of this type are built into browsers or can be added as "plug-ins."

**INTERNET** (Upper case I)

The vast collection of interconnected networks that all use the TCP/IP protocols and that evolved from the ARPANET of the late 60's and early 70's. An "internet" (lower case i) is any computers connected to each other (a network), and are not part of the Internet unless the use TCP/IP protocols. An "intranet" is a private network inside a company or organization that uses the same kinds of software that you would find on the public Internet, but that is only for internal use. An intranet may be on the Internet or may simply be a network.

**IP Address or IP Number**

(Internet Protocol number or address). A unique number consisting of 4 parts separated by dots, e.g. 165.113.245.2

Every machine that is on the Internet has a unique IP address. If a machine does not have an IP number, it is not really on the Internet. Most machines also have one or more Domain Names that are easier for people to remember.

**ISP or Internet Service Provider**

A company that sells Internet connections via modem (examples: aol, Mindspring - thousands of ISPs to choose from; not easy to evaluate). Faster, more expensive Internet connectivity is available via cable, DSL, ISDN, or web-TV. Often these companies also provide Web page hosting service (free or relatively inexpensive web pages -- the origin of many personal pages).

**JAVA**

A network-oriented programming language invented by Sun Microsystems that is specifically designed for writing programs that can be safely downloaded to your computer through the Internet and immediately run without fear of viruses or other harm to our computer or files. Using small Java programs (called "Applets"), Web pages can include functions such as animations, calculators, and other fancy tricks. We can expect to see a huge variety of features added to the Web using Java, since you can write a Java program to do almost anything a regular computer program can do, and then include that Java program in a Web page.

**JAVAScript**

A simple programming language developed by Netscape to enable greater interactivity in Web pages. It shares some characteristics with JAVA but is independent. It interacts with HTML, enabling dynamic content and motion.

**KEYWORD(S)**

A word searched for in a search command. Keywords are searched in any order. Use spaces to separate keywords in simple keyword searching.

**LIMITING TO A FIELD**

Requiring that a keyword or phrase appear in a specific field of documents retrieved. Most often used to limit to the "Title" field in order to find documents primarily about one or more keywords. (Can be used for other fields.)

**LINK**

The URL imbedded in another document, so that if you click on the highlighted text or button referring to the link, you retrieve the outside URL. If you search the field "link:", you retrieve on text in these imbedded URLs which you do not see in the documents.

**LINK "ROT"**

Term used to describe the frustrating and frequent problem caused by the constant changing in URLs. A Web page or search tool offers a link and when you click on it, you get an error message (e.g., "not available") or a page saying the site has moved to a new URL. Search engine spiders cannot keep up with the changes. URLs change frequently because the documents are moved to new computers, the file structure on the computer is reorganized, or sites are discontinued. If there is no referring link to the new URL, there is little you can do but try to search for the same or an equivalent site from scratch.

**LISTSERVERS**

A discussion group mechanism that permits you to subscribe and receive and participate in discussions via e-mail. Blogs and RSS feeds provide some of the communication functionality of listservers.

**META-SEARCH ENGINE**

Search engines that automatically submit your keyword search to several other search tools, and retrieve results from all their databases. Convenient time-savers for relatively simple keyword searches (one or two keywords or phrases in " ").

**NESTING**

A term used in Boolean searching to indicate the sequence in which operations are to be performed. Enclosing words in parentheses identifies a group or "nest." Groups can be within other groups. The operations will be performed from the innermost nest to the outmost, and then from left to right.

**NEWSGROUP**

A discussion group operated through the Internet. Not to be confused with LISTSERVERS which operate through e-mail.

**PERSONAL PAGE**

A web page created by an individual (as opposed to someone creating a page for an institution, business, organization, or other entity). Often personal pages contain valid and useful opinions, links to important resources, and significant facts. One of the greatest benefits of the Web is the freedom it as given almost anyone to put his or her ideas "out there." But frequently personal pages offer highly biased personal perspectives or ironical/satirical spoofs, which must be evaluated carefully. The presence in the page's URL of a personal name (such as "jbarker") and a ~ or % or the word "users" or "people" or "members" very frequently indicate a site offering personal pages.

**PACKET, PACKET JAM**

When you retrieve a document via the WWW, the document is sent in "packets" which fit in between other messages on the telecommunications lines, and then are reassembled when they arrive at your end. This occurs using TCP/IP protocol. The packets may be sent via different paths on the networks which carry the Internet. If any of these packets gets delayed, your document cannot be reassembled and displayed. This is called a "packet jam." You can often resolve packet jams by pressing STOP then RELOAD. RELOAD requests a fresh copy of the document, and it is likely to be sent without jamming.

**PDF or .pdf or pdf file**

Abbreviation for Portable Document Format, a file format developed by Adobe Systems, that is used to capture almost any kind of document with the formatting in the original. Viewing a PDF file requires Acrobat Reader, which is built into most browsers and can be downloaded free from Adobe.

**PHRASE**

More than one KEYWORD, searched exactly as keyed (all terms required to be in documents, in the order keyed). Enclosing keywords in quotations " " forms a phrase in AltaVista, , and some other search tools. Some times a phrase is called a "character string."

**PLUG-IN**

An application built into a browser or added to a browser to enable it to interact with a special file type (such as a movie, sound file, Word document, etc.)

**POPULARITY RANKING of search results**

Some search engines rank the order in which search results appear primarily by how many other sites link to each page (a kind of popularity vote based on the assumption that other pages would create a link to the "best" pages). Google is the best example of this.

**+REQUIRE or -REJECT A TERM OR PHRASE**

Insert + immediately before a term (no space) to limit search to documents containing a term. Insert - immediately before a term (no space) to exclude documents containing a term. Can be used immediately (no space) before the " " delimiting a phrase.

Functions partially like basic BOOLEAN LOGIC. If + precedes more than one term, they are required as with Boolean AND. If - is used, terms are excluded as with Boolean AND NOT. If neither + no - is used, the default is Boolean OR. However, full Boolean logic allows parentheses to group and sequence logical operations, and +/- do not.

**RELEVANCY RANKING of search results**

The most common method for determining the order in which search results are displayed. Each search tool uses its own unique algorithm. Most use "fuzzy and" combined with factors such as how often your terms occur in documents, whether they occur together as a phrase, and whether they are in title or how near the top of the text. Popularity is another ranking system.

**RSS or RSS feeds**

Short for "Really Simple Syncation" (a.k.a. Rich Site Summary or RDF Site Summary), refers to a group of XML based web-content distribution and republication (Web syndication) formats primarily used by news sites and weblogs (blogs). Any website can issue an RSS feed. By subscribing to an RSS feed, you are alerted to new additions to the feed since you last read it. In order to read RSS feeds, you must use a "feed reader," which formats the XML code into an easily readable format (feed readers are to XML and RSS feeds as web browsers are to HTML and web pages).

**SCRIPT**

A script is a type of programming language that can be used to fetch and display Web pages. There are many kinds and uses of scripts on the Web. They can be used to create all or part of a page, and communicate with searchable databases. Forms (boxes) and many interactive links, which respond differently depending on what you enter, all require some kind of script language. When you find a question mark (?) in the URL of a page, some kind of script command was used in generating and/or delivering that page. Most search engine spiders are instructed not to crawl pages from scripts, although it is usually technically possible for them to do so.

**SERVER, WEB SERVER**

A computer running that software, assigned an IP address, and connected to the Internet so that it can provide documents via the World Wide Web. Also called HOST computer. Web servers are the closest equivalent to what in the print world is called the "publisher" of a print document. An important difference is that most print publishers carefully edit the content and quality of their publications in an effort to market them and future publications. This convention is not required in the Web world, where anyone can be a publisher; careful evaluation of Web pages is therefore mandatory. Also called a "Host."

**SERVER-SIDE**

Something that operates on the "server" computer (providing the Web page), as opposed to the "client" computer (which is you or someone else viewing the Web page). Usually it is a program or command or procedure or other application causes dynamic pages or animation or other interaction.

**SHTML, usually seen as .shtml**

An file name extension that identifies web pages containing SSI commands.

**SITE or WEB-SITE**

This term is often used to mean "web page," but there is supposed to be a difference. A web page is a single entity, one URL, one file that you might find on the Web. A "site," properly speaking, is an location or gathering or center for a bunch of related pages linked to from that site. For example, the site for the present tutorial is the top-level page "Internet Resources." All of the pages associated with it branch out from there -- the web searching tutorial and all its pages, and more. Together they make up a "site." When we estimate there are 5 billion web pages on the Web, we do not mean "sites." There would be far fewer sites.

## **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. Even so, the Web is so enormous that it can take six months for spiders to cover it, resulting in a certain degree of "out-of-datedness" (link rot) in all the search engines.

## **SPONSOR (of a Web page or site)**

Many Web pages have organizations, businesses, institutions like universities or nonprofit foundations, or other interests which "sponsor" the page. Frequently you can find a link titled "Sponsors" or an "About us" link explaining who or what (if anyone) is sponsoring the page. Sometimes the advertisers on the page (banner ads, links, buttons to sites that sell or promote something) are "sponsors." *WHY is this important?* Sponsors and the funding they provide may, or may not, influence what can be said on the page or site -- can bias what you find, by excluding some opposing viewpoint or causing some other imbalanced information. The site is not bad because of sponsors, but you they should alert you to the need to evaluate a page or site very carefully.

## **SSI commands**

SSI stands for "server-side include," a type of HTML instruction telling a computer that serves Web pages to dynamically generate data, usually by inserting certain variable contents into a fixed template or boilerplate Web page. Used especially in database searches.

## **STEMMING**

In keyword searching, word endings are automatically removed (*lines* becomes *line*); searches are performed on the stem + common endings (*line* or *lines* retrieves *line*, *lines*, *line's*, *lines'*, *lining*, *lined*). Not very common as a practice, and not always disclosed. Can usually be avoided by placing a term in " ".

## **STOP WORDS**

In database searching, "stop words" are small and frequently occurring words like *and*, *or*, *in*, *of* that are often ignored when keyed as search terms.

Sometimes putting them in quotes " " will allow you to search them.

Sometimes + immediately before them makes them searchable.

## **SUBJECT-BASED POPULARITY RANKING of search results**

A variation on popularity ranking in which the links in pages on the same subject are used to in ranking search results. Used by Teoma.

## **SUBJECT DIRECTORY**

An approach to Web documents by a lexicon of subject terms hierarchically grouped. May be browsed or searched by keywords. Subject directories are smaller than other searchable databases, because of the human involvement required to classify documents by subject.

## **SUB-SEARCHING**

Ability to search only within the results of a previous search. Enables you to refine search results, in effect making the computer "read" the search results for you selecting documents with terms you sub-search on. Can function much like RESULTS RANKING.

## TCP/IP

(Transmission Control Protocol/Internet Protocol) -- This is the suite of protocols that defines the Internet. Originally designed for the UNIX operating system, TCP/IP software is now available for every major kind of computer operating system. To be truly on the Internet, your computer must have TCP/IP software.

## TELNET

Internet service allowing one computer to log onto another, connecting as if not remote.

## THESAURUS

In some search tools, the terms you choose to search on can lead you to other terms you may not have thought of. Different search tools have different ways of presenting this information, sometimes with suggested words you may choose among and sometimes automatically. The terms are based on the terms in the results of your search, not on some dictionary-like thesaurus.

## TITLE (of a document)

The official title of a document from the "meta" field called title. The text of this meta title field may or may not also occur in the visible body of the document. It is what appears in the top bar of the window when you display the document and it is the title that appears in search engine results. The "meta" field called title is not mandatory in HTML coding. Sometimes you retrieve a document with "No Title" as its supposed title; this is caused when the meta-title field is left blank.

In Alta Vista and some other search tools, *title:* search also matches on the "meta" field, which contains document descriptors not displayed on the Web.

## TRUNCATION

In a search, the ability to enter the first part of a keyword, insert a symbol (usually \*), and accept any variant spellings or word endings, from the occurrence of the symbol forward. (E.g., *femini\** retrieves *feminine*, *feminism*, *feminism*, etc.)

## URL

Uniform Resource Locator. The unique address of any Web document. May be keyed in a browser's OPEN or LOCATION / GO TO box to retrieve a document. There is a logic the layout of a URL:

### ***Anatomy of a URL:***

Type of file (could say ftp:// or telnet://)	Domain name (computer file is on and its location on the Internet)	Path or directory on the computer to this file	Name of file, and its file extension (usually ending in .html or .htm)
http://	www.lib.berkeley.edu/	TeachingLib/Guides/Internet/	FindInfo.html

## USENET

Bulletinboard-like network featuring thousands of "newsgroups." Google incorporates the historic file of Usenet Newsgroups (back to 1981) into its Google Groups. Yahoo Groups offers a similar service, but does not include the old "Usenet Newsgroups." Blogs are replacing some of the need for this type of community sharing and information exchange.

**WORD VARIANTS**

Different word endings (such as *-ing*, *-s*, *es*, *-ism*, *-ist*, etc.) will be retrieved only if you allow for them in your search terms. One way to do this is TRUNCATION, but few systems accept truncation. Another way is to enter the variants either separated by BOOLEAN OR (and grouped in parentheses). In non-Boolean systems, enter the variant terms preceded with neither + nor -, because this will allow documents containing any of them to be retrieved.

**XHTML**

A variant of HTML. Stands for Extensible Hypertext Markup Language is a hybrid between HTML and XML that is more universally acceptable in Web pages and search engines than XML.

**XML**

Extensible Markup Language, a dilution for Web page use of SGML (Standard General Markup Language), which is not readily viewable in ordinary browsers and is difficult to apply to Web pages. XML is very useful (among other things) for pages emerging from databases and other applications where parts of the page are standardized and must reappear many times.