

A Basic Guide to the Internet

By Laura Cohen
University at Albany Libraries

The Internet is a computer network made up of thousands of networks worldwide. No one knows exactly how many computers are connected to the Internet. It is certain, however, that these number in the millions.

No one is in charge of the Internet. There are organizations which develop technical aspects of this network and set standards for creating applications on it, but no governing body is in control. The Internet backbone, through which Internet traffic flows, is owned by private companies.

All computers on the Internet communicate with one another using the Transmission Control Protocol/Internet Protocol suite, abbreviated to TCP/IP. Computers on the Internet use a client/server architecture. This means that the remote server machine provides files and services to the user's local client machine. Software can be installed on a client computer to take advantage of the latest access technology.

An Internet user has access to a wide variety of services: electronic mail, file transfer, vast information resources, interest group membership, interactive collaboration, multimedia displays, real-time broadcasting, shopping opportunities, breaking news, and much more. The Internet consists primarily of a variety of access protocols. Many of these protocols feature programs that allow users to search for and retrieve material made available by the protocol.

COMPONENTS OF THE INTERNET

WORLD WIDE WEB

The World Wide Web (abbreviated as the Web or WWW) is a system of Internet servers that supports hypertext to access several Internet protocols on a single interface. Almost every protocol type available on the Internet is accessible on the Web. This includes e-mail, FTP, Telnet, and Usenet News. In addition to these, the World Wide Web has its own protocol: HyperText Transfer Protocol, or HTTP. These protocols will be explained later in this handout.

The World Wide Web provides a single interface for accessing all these protocols. This creates a convenient and user-friendly environment. It is no longer necessary to be conversant in these protocols within separate, command-level environments. The Web gathers together these protocols into a single system. Because of this feature, and because of the Web's ability to work with multimedia and advanced programming languages, the World Wide Web is the fastest-growing component of the Internet.

The operation of the Web relies primarily on hypertext as its means of information retrieval. HyperText is a document containing words that connect to other documents. These words are called links and are selectable by the user. A single hypertext document can contain links to many documents. In the context of the Web, words or graphics may serve as links to other

documents, images, video, and sound. Links may or may not follow a logical path, as each connection is programmed by the creator of the source document. Overall, the Web contains a complex virtual web of connections among a vast number of documents, graphics, videos, and sounds.

Producing hypertext for the Web is accomplished by creating documents with a language called HyperText Markup Language, or HTML. With HTML, tags are placed within the text to accomplish document formatting, visual features such as font size, italics and bold, and the creation of hypertext links. Graphics and multimedia may also be incorporated into an HTML document. HTML is an evolving language, with new tags being added as each upgrade of the language is developed and released. The World Wide Web Consortium (W3C), led by Web founder Tim Berners-Lee, coordinates the efforts of standardizing HTML. The W3C now calls the language XHTML and considers it to be an application of the XML language standard.

The World Wide Web consists of files, called pages or home pages, containing links to documents and resources throughout the Internet.

The Web provides a vast array of experiences including multimedia presentations, real-time collaboration, interactive pages, radio and television broadcasts, and the automatic "push" of information to a client computer. Programming languages such as Java, JavaScript, Visual Basic, Cold Fusion and XML are extending the capabilities of the Web. A growing amount of information on the Web is served dynamically from content stored in databases. The Web is therefore not a fixed entity, but one that is in a constant state of development and flux.

E-MAIL

Electronic mail, or e-mail, allows computer users locally and worldwide to exchange messages. Each user of e-mail has a mailbox address to which messages are sent. Messages sent through e-mail can arrive within a matter of seconds.

A powerful aspect of e-mail is the option to send electronic files to a person's e-mail address. Non-ASCII files, known as binary files, may be attached to e-mail messages. These files are referred to as MIME attachments. MIME stands for Multimedia Internet Mail Extension, and was developed to help e-mail software handle a variety of file types. For example, a document created in Microsoft Word can be attached to an e-mail message and retrieved by the recipient with the appropriate e-mail program. Many e-mail programs, including Eudora, Netscape Messenger, and Microsoft Outlook, offer the ability to read files written in HTML, which is itself a MIME type.

TELNET

Telnet is a program that allows you to log into computers on the Internet and use online databases, library catalogs, chat services, and more. There are no graphics in Telnet sessions, just text. To Telnet to a computer, you must know its address. This can consist of words (locis.loc.gov) or numbers (140.147.254.3). Some services require you to connect to a specific

port on the remote computer. In this case, type the port number after the Internet address.
Example: telnet nri.reston.va.us 185.

Telnet is available on the World Wide Web. Probably the most common Web-based resources available through Telnet are library catalogs. A link to a Telnet resource may look like any other link, but it will launch a Telnet session to make the connection. A Telnet program must be installed on your local computer and configured to your Web browser in order to work.

With the increasing popularity of the Web, Telnet has become less frequently used as a means of access to information on the Internet.

FTP

FTP stands for File Transfer Protocol. This is both a program and the method used to transfer files between computers. Anonymous FTP is an option that allows users to transfer files from thousands of host computers on the Internet to their personal computer account. FTP sites contain books, articles, software, games, images, sounds, multimedia, course work, data sets, and more.

If your computer is directly connected to the Internet via an Ethernet cable, you can use one of several PC software programs, such as WS_FTP for Windows, to conduct a file transfer.

FTP transfers can be performed on the World Wide Web without the need for special software. In this case, the Web browser will suffice. Whenever you download software from a Web site to your local machine, you are using FTP. You can also retrieve FTP files via search engines such as FtpFind, located at <http://www.ftpfind.com/>. This option is easiest because you do not need to know FTP program commands.

E-MAIL DISCUSSION GROUPS

One of the benefits of the Internet is the opportunity it offers to people worldwide to communicate via e-mail. The Internet is home to a large community of individuals who carry out active discussions organized around topic-oriented forums distributed by e-mail. These are administered by software programs. Probably the most common program is the listserv.

A great variety of topics are covered by listservs, many of them academic in nature. When you subscribe to a listserv, messages from other subscribers are automatically sent to your electronic mailbox. You subscribe to a listserv by sending an e-mail message to a computer program called a listserver. Listservers are located on computer networks throughout the world. This program handles subscription information and distributes messages to and from subscribers. You must have a e-mail account to participate in a listserv discussion group. Visit Tile.net at <http://tile.net/> to see an example of a site that offers a searchable collection of e-mail discussion groups.

Majordomo and Listproc are two other programs that administer e-mail discussion groups. The commands for subscribing to and managing your list memberships are similar to those of listserv.

USENET NEWS

Usenet News is a global electronic bulletin board system in which millions of computer users exchange information on a vast range of topics. The major difference between Usenet News and e-mail discussion groups is the fact that Usenet messages are stored on central computers, and users must connect to these computers to read or download the messages posted to these groups. This is distinct from e-mail distribution, in which messages arrive in the electronic mailboxes of each list member.

Usenet itself is a set of machines that exchanges messages, or articles, from Usenet discussion forums, called newsgroups. Usenet administrators control their own sites, and decide which (if any) newsgroups to sponsor and which remote newsgroups to allow into the system.

There are thousands of Usenet newsgroups in existence. While many are academic in nature, numerous newsgroups are organized around recreational topics. Much serious computer-related work takes place in Usenet discussions. A small number of e-mail discussion groups also exist as Usenet newsgroups.

The Usenet newsgroups can be read by a variety of newsreader software programs. For example, the Netscape suite comes with a newsreader program called Messenger. Newsreaders are also available as standalone products.

FAQ, RFC, FYI

FAQ stands for Frequently Asked Questions. These are periodic postings to Usenet newsgroups that contain a wealth of information related to the topic of the newsgroup. Many FAQs are quite extensive. FAQs are available by subscribing to individual Usenet newsgroups. A Web-based collection of FAQ resources has been collected by The Internet FAQ Consortium and is available at <http://www.faqs.org/>.

RFC stands for Request for Comments. These are documents created by and distributed to the Internet community to help define the nuts and bolts of the Internet. They contain both technical specifications and general information.

FYI stands for For Your Information. These notes are a subset of RFCs and contain information of interest to new Internet users.

Links to indexes of all three of these information resources are available on the University Libraries Web site at <http://library.albany.edu/reference/faqs.html>.

CHAT & INSTANT MESSAGING

Chat programs allow users on the Internet to communicate with each other by typing in real time. They are sometimes included as a feature of a Web site, where users can log into the "chat room" to exchange comments and information about the topics addressed on the site. Chat may take

other, more wide-ranging forms. For example, America Online is well known for sponsoring a number of topical chat rooms.

Internet Relay Chat (IRC) is a service through which participants can communicate to each other on hundreds of channels. These channels are usually based on specific topics. While many topics are frivolous, substantive conversations are also taking place. To access IRC, you must use an IRC software program.

A variation of chat is the phenomenon of instant messaging. With instant messaging, a user on the Web can contact another user currently logged in and type a conversation. Most famous is America Online's Instant Messenger. ICQ, MSN and Yahoo are other commonly-used chat programs.

MUD/MUSH/MOO/MUCK/DUM/MUSE

MUD stands for Multi User Dimension. MUDs, and their variations listed above, are multi-user virtual reality games based on simulated worlds. Traditionally text based, graphical MUDs now exist. There are MUDs of all kinds on the Internet, and many can be joined free of charge. For more information, read one of the FAQs devoted to MUDs available at the FAQ site at <http://www.faqs.org/>.

UNDERSTANDING THE WORLD WIDE WEB

The World Wide Web is a system of Internet servers that supports hypertext to access several Internet protocols on a single interface. The World Wide Web is often abbreviated as the Web or WWW.

The World Wide Web was developed in 1989 by Tim Berners-Lee of the European Particle Physics Lab (CERN) in Switzerland. The initial purpose of the Web was to use networked hypertext to facilitate communication among its members, who were located in several countries. Word was soon spread beyond CERN, and a rapid growth in the number of both developers and users ensued. In addition to hypertext, the Web began to incorporate graphics, video, and sound. The use of the Web has now reached global proportions.

Almost every protocol type available on the Internet is accessible on the Web. **Internet protocols** are sets of rules that allow for intermachine communication on the Internet. The following major protocols are accessible on the Web:

E-mail (Simple Mail Transport Protocol or SMTP)

Distributes electronic messages and files to one or more electronic mailboxes

Telnet (Telnet Protocol)

Facilitates login to a computer host to execute commands

FTP (File Transfer Protocol)

Transfers text or binary files between an FTP server and client

Usenet (Network News Transfer Protocol or NNTP)

Distributes Usenet news articles derived from topical discussions on newsgroups

HTTP (HyperText Transfer Protocol)

Transmits hypertext over networks. This is the protocol of the WWW.

Many other protocols are available on the Web. To name just one example, the Voice over Internet Protocol (VoIP) allows users to place a telephone call over the Web.

The World Wide Web provides a single interface for accessing all these protocols. This creates a convenient and user-friendly environment. It is no longer necessary to be conversant in these protocols within separate, command-level environments. The Web gathers together these protocols into a single system. Because of this feature, and because of the Web's ability to work with multimedia and advanced programming languages, the World Wide Web is the most popular component of the Internet.

HYPERTEXT AND LINKS: THE MOTION OF THE WEB

The operation of the Web relies primarily on hypertext as its means of information retrieval. HyperText is a document containing words that connect to other documents. These words are called links and are selectable by the user. A single hypertext document can contain links to many documents. In the context of the Web, words or graphics may serve as links to other documents, images, video, and sound. Links may or may not follow a logical path, as each connection is programmed by the creator of the source document. Overall, the Web contains a complex virtual web of connections among a vast number of documents, graphics, videos, and sounds.

Producing hypertext for the Web is accomplished by creating documents with a language called HyperText Markup Language, or HTML. With HTML, tags are placed within the text to accomplish document formatting, visual features such as font size, italics and bold, and the creation of hypertext links. Graphics may also be incorporated into an HTML document. HTML is an evolving language, with new tags being added as each upgrade of the language is developed and released. The World Wide Web Consortium, led by Tim Berners-Lee, coordinates the efforts of standardizing HTML. The W3C now calls the language XHTML and considers it to be an application of the XML language standard.

PAGES ON THE WEB

The World Wide Web consists of files, called pages or Web pages, containing information and links to resources throughout the Internet.

Web pages can be created by user activity. For example, if you visit a Web search engine and enter keywords on the topic of your choice, a page will be created containing the results of your

search. In fact, a growing amount of information found on the Web today is served from databases, creating temporary Web pages "on the fly" in response to user queries.

Access to Web pages may be accomplished by:

1. Entering an Internet address and retrieving a page directly
2. Browsing through pages and selecting links to move from one page to another
3. Searching through subject directories linked to organized collections of Web pages
4. Entering a search statement at a search engine to retrieve pages on the topic of your choice

RETRIEVING DOCUMENTS ON THE WEB: THE URL and DOMAIN NAME SYSTEM

URL stands for Uniform Resource Locator. The URL specifies the Internet address of a file stored on a host computer connected to the Internet. Every file on the Internet, no matter what its access protocol, has a unique URL. Web software programs use the URL to retrieve the file from the host computer and the directory in which it resides. This file is then displayed on the monitor connected to the user's local machine.

URLs are translated into numeric addresses using the **Domain Name System (DNS)**. The DNS is a worldwide system of servers that stores location pointers to Web sites. The numeric address, called the IP (Internet Protocol) address, is actually the "real" URL. Since numeric strings are difficult for humans to use, alphanumeric addresses are employed by end users. Once the translation is made by the DNS, the browser can contact the Web server and ask for a specific file located on its site.

Anatomy of a URL

This is the format of the URL:

protocol://host/path/filename

For example, this is a URL on the home page of the House Committee on Agriculture of the U.S. House of Representatives:

<http://www.house.gov/agriculture/schedule.htm>

This URL is typical of addresses hosted in domains in the United States.

Structure of this URL:

1. Protocol: **http**
2. Host computer name: **www**
3. Second-level domain name: **house**
4. Top-level domain name: **gov**
5. Directory name: **agriculture**
6. File name: **schedule.htm**,/li>

Note how much information about the content of the file is present in this well-constructed URL.

Other examples:

telnet://opac.albany.edu the University at Albany library text-based catalog

ftp://ftp.uu.net/graphics/picasso a file at an ftp site

Several top-level domains (TLDs) are common in the United States:

com	commercial enterprise
edu	educational institution
gov	U.S. government entity
mil	U.S. military entity
net	network access provder
org	usually nonprofit organizations

New domain names were approved in November 2000 by the Internet Corporation for Assigned Names and Numbers (ICANN): .biz, .museum, .info, .pro (for professionals) .name (for individuals), .aero (for the aerospace industry), and .coop (for cooperatives). These domain names are beginning to become available.

In addition, dozens of domain names have been assigned to identify and locate files stored on host computers in countries around the world. These are referred to as **two-letter Internet country codes**, and have been standardized by the International Standards Organization as ISO 3166. For example:

ch	Switzerland
de	Germany
jp	Japan
uk	United Kingdom

HOW TO ACCESS THE WORLD WIDE WEB: WEB BROWSERS

To access the World Wide Web, you must use a Web browser. A browser is a software program that allows users to access and navigate the World Wide Web. There are two types of browsers:

1. **Graphical:** Text, images, audio, and video are retrievable through a graphical software program such as Internet Explorer, Netscape Navigator and Opera. These browsers are available for both Windows-based and Macintosh computers. Navigation is accomplished by pointing and clicking with a mouse on highlighted words and graphics.
You can install a graphical browser such as Internet Explorer and Netscape Navigator in your Windows-based or Macintosh machine. Internet Explorer is available from the Microsoft Web site: <http://www.microsoft.com/>. Navigator is available for downloading on the Netscape Web site: <http://home.netscape.com/>. To use these programs to access the Web, you need a connection to the Internet. This is accomplish through such means as an ethernet connection, a dialup connection known as a SLPP or PPP, or a cable modem
2. **Text:** Lynx is a browser that provides access to the Web in text-only mode. Navigation is accomplished by highlighting emphasized words in the screen with the arrow up and down keys, and then pressing the forward arrow (or Enter) key to follow the link.
Extending the Browser: Plug-Ins

Software programs may be configured to a Web browser in order to enhance its capabilities. When the browser encounters a sound, image or video file, it hands off the data to other programs, called **plug-ins**, to run or display the file. Working in conjunction with plug-ins, browsers can offer a seamless multimedia experience. Many plug-ins are available for free.

File formats requiring plug-ins are known as **MIME types**. MIME stands for Multimedia Internet Mail Extension, and was originally developed to help e-mail software handle a variety of binary (non-ASCII) file attachments. The use of MIME has expanded to the Web. For example, the basic MIME type handled by Web browsers is text/html associated with the file extension .html.

A common plug-in utilized on the Web is the Adobe Acrobat Reader. The Acrobat Reader allows you to view documents created in Adobe's Portable Document Format (PDF). These documents are the MIME type "application/pdf" and are associated with the file extension .pdf. When the Acrobat Reader has been configured to your browser, the program will open and display the file requested when you click on a hyperlinked file name with the suffix .pdf. The latest versions of the Acrobat Reader allow for the viewing of documents within the browser window.

Web browsers are often standardized with a small suite of plug-ins, especially for playing multimedia content. Additional plug-ins may be obtained at the browser's Web site, at special download sites on the Web, or from the Web sites of the companies that created the programs. The number of available plug-ins is increasing rapidly.

Once a plug-in is configured to your browser, it will automatically launch when you choose to access a file type that it uses.

BEYOND PLUG-INS: ACTIVE X

ActiveX is a technology developed by Microsoft which make plug-ins less necessary. ActiveX offers the opportunity to embed animated objects, data, and computer code on Web pages. A web browser supporting ActiveX can render most items encountered on a Web page. For example, Active X allows users to view three-dimensional VRML worlds in a Web browser without the use of a VRML plug-in. As another example of the power of ActiveX, this technology can allow you to view and edit PowerPoint presentations directly within your Web browser. ActiveX works best with Microsoft's Internet Explorer browser.

THE EXPERIENCE OF THE WEB

Today's World Wide Web presents an ever-diversified experience of multimedia, programming languages, and real-time communication. There is no question that it is a challenge to keep up with the rapid pace of developments. The following presents a brief description of some of the more important trends to watch.

Multimedia

The Web has become a broadcast medium. It is possible to listen to audio and video over the Web, both pre-recorded and live. For example, you can visit the sites of various news organizations and view the same videos shown on the nightly television news. Several plug-ins are available for viewing these videos. For example, Apple's Quick Time Player downloads files with the .mov extension and displays these as "movies" in a small window on your computer screen. Quick Time files can be quite large, and it may take patience to wait for the entire movie to download into your computer before you can view it.

The problem of slow download times has been answered by a revolutionary development in multimedia capability: streaming media. In this case, audio or video files are played as they are downloading, or streaming, into your computer. Only a small wait, called buffering, is necessary before the file begins to play. The Windows Media Player and RealPlayer plug-in play streaming audio and video files. Extensive files such as interviews, speeches and hearings work very well with these players. They are also ideal for the broadcast of real-time events. These may include press conferences, live radio and television broadcasts, concerts, etc. Many sites offer the option to use one player or the other.

Shockwave presents another multimedia experience. Shockwave allows for the creation and implementation of an entire multimedia display combining graphics, animation and sound.

Sound files, including music, may also be heard on the Web. It is not uncommon to visit a Web page and hear background music. Sound files are also available for downloading independent of Web page visits. Sound files of many types are supported by the Web with the appropriate plug-ins. The MP3 file format, and the choice of supporting plug-ins, is the latest music trend to sweep the Web. The famous Napster site allows for the exchange of MP3 files.

Live cams are another aspect of the multimedia experience available on the Web. Live cams are video cameras that send their data in real time to a Web server. These cams may appear in all kinds of locations, both serious and whimsical: an office, on top of a building, a scenic locale, a special event, and so on.

Programming Languages and Functions

The use of existing and new programming languages have extended the capabilities of the Web. What follows is a basic guide to a group of the more common languages and functions in use on the Web today.

CGI, Active Server Pages: CGI (Common Gateway Interface) refers to a specification by which programs can communicate with a Web server. A CGI program, or script, is any program designed to accept and return data that conforms to the CGI specification. The program can be written in any programming language, including C, Perl, and Visual Basic Script. A common use for a CGI script is to process an interactive form on a Web page. For example, you might fill out a form ordering a book through Interlibrary Loan. The script processes your information and sends it to a designated e-mail address in the Interlibrary Loan department.

Another type of dynamically generated Web page is called Active Server Pages (ASP). Developed by Microsoft, ASPs are HTML pages that include scripting and create interactive Web server applications. The scripts run on the server, rather than on the Web browser, to generate the HTML pages sent to browsers. Visual Basic and JScript (a subset of JavaScript) are often used for the scripting. ASPs end in the file extension .asp.

Java/Java Applets: Java is probably the most famous of the programming languages of the Web. Java is an object-oriented programming language similar to C++. Developed by Sun Microsystems, the aim of Java is to create programs that will be platform independent. The Java motto is, "Write once, run anywhere." A perfect Java program should work equally well on a PC, Macintosh, Unix, and so on, without any additional programming. This goal has yet to be realized. Java can be used to write applications for both Web and non-Web use.

Web-based Java applications are usually in the form of Java applets. These are small Java programs called from an HTML page that can be downloaded from a Web server and run on a Java-compatible Web browser. A few examples include live newsfeeds, moving images with sound, calculators, charts and spreadsheets, and interactive visual displays. Java applets can tend to load slowly, but programming improvements should lead to a shortened loading time.

JavaScript/JScript: JavaScript is a programming language created by Netscape Communications. Small programs written in this language are embedded within an HTML page, or called externally from the page, to enhance the page's the functionality. Examples of JavaScript include moving tickers, drop-down menus, real-time calendars and clocks, and mouse-over interactions. JScript is a similar language developed by Microsoft and works with the company's Internet Explorer browser.

VRML: VRML (Virtual Reality Modeling Language) allows for the creation of three-dimensional worlds. These may be linked from Web pages and displayed with a VRML viewer. Netscape Communicator comes with the Cosmo viewer for experiencing these three-dimensional worlds. One of the most interesting aspects of VRML is the option to "enter" the world and control your movements within the world.

XML: XML (eXtensible Markup Language) is a Web page creation language that enables designers to create their own customized tags to provide functionality not available with HTML. XML is a language of data structure and exchange, and allows developers to separate form from content. At present, this language is little used as Web browsers are only beginning to support it. In May 1999, however, the W3 Consortium announced that HTML 4.0 has been recast as an XML application called XHTML. This move will have a significant impact on the future of both XML and HTML.

Real-Time Communication

Text, audio and video communication can occur in real time on the Web. This capability allows people to conference and collaborate in real time. In general, the faster the Internet connection, the more successful the experience.

At its simplest, chat programs allow multiple users to type to each other in real time. Internet Relay Chat and America Online's Instant Messenger are prime examples of this type of program. The development of a messaging protocols is underway. Such a protocol would allow for the expansion of this capability throughout the Internet.

More enhanced real-time communication offers an audio and/or video component. CU-See Me is one of the most popular software programs of this type. Even more elaborate are programs that allow for true real-time collaboration. Microsoft's NetMeeting and Netscape's Conference (available with Communicator) are good examples of this.

Featured collaboration tools include:

- audio: conduct a telephone conversation on the Web
- video: view your audience
- file transfer: send files back and forth among participants
- chat: type in real time
- whiteboard: draw, mark up, and save images on a shared window or board
- document/application sharing: view and use a program on another's desktop machine
- collaborative Web browsing: visit Web pages together

Currently no standard exists that will work among all conferencing programs.

Push: Push refers to a technology that sends data to a program without the program's request. This is the opposite of the typical "pull" of the Web, in which the user clicks on a link to request a file from a server. With push, the data is sent automatically. Content is sent through a "channel." The early Web-based implementation of push was commercial. Push can also be used to deliver software upgrades to a desktop machine.

USING INTERNET EXPLORER 6

Internet Explorer is a Web browser produced by the Microsoft Corporation. This tutorial will cover the major features of version 6. For the sake of brevity, we will refer to the software as IE 6.

NOTE: Most of these features are also available on IE 5.5.

THE TOOLBARS

IE 6 has two toolbars at the top of of the browser window:

- Menu Bar: Contains menu items that open up dropdown lists for related options. Among the items are options for printing, customizing IE 6, copying and pasting text, managing Favorites, and accessing Help.
- Navigation Toolbar: Contains icons for a variety of features including navigating among Web pages, searching the Web using a selection of search tools, accessing and managing Favorites, viewing a History of visited pages, printing, and accessing email and newsgroups.
-

HOW TO ACCESS RESOURCES ON THE WEB WITH IE 6

1. If you have the URL (address) of a Web page

Type the URL to go directly to the page. IE 6 gives you two ways of doing this.

- Type the URL in the Address bar at the top of the screen. To accomplish this, click on the Address bar to highlight the current URL. Then type in the new URL and press the Enter key.
- Click on File/Open at the top left of the screen. A pop-up window will appear with a text entry window. Within that window, type the URL of the file you wish to retrieve. Press the Enter key.

2. If you are on a Web page

Click on

- words or images which change the shape of the mouse pointer from an arrow to a hand and display a URL on the bottom of the screen when the mouse pointer is placed over it
- the blue words on the display screen
- the purple words on the display screen (the purple color indicates that the resource has been recently accessed on your terminal)

Note: The color blue is generally the default color for text that contains a link, and purple is the default color for text representing a link that has been visited in the recent past. Nowadays, Web page creators are coloring their links in all sorts of ways. The best way to figure out which text represents a link is to position your mouse over the words and see if the pointer shape changes from an arrow to a hand. The hand represents a link.

3. If you want to use pre-installed links

IE 6 offers a collection of Web sites in its Favorites collection. Click on Favorites on either the text bar or the tool bar at the top of the screen to access these resources.

NAVIGATING THE WEB WITH IE 6

IE 6 allows you to move back and forth among the Web pages that you visit during a session.

To go back to previous sites:

- Click on the small Back left arrow on the navigation bar near the top left corner of your screen. Each time you click on this arrow, you will return to the next previous site that you visited. If you hold your mouse over the Back arrow, the title of the upcoming page will briefly appear.
- To skip farther back, click on the small black triangle to the right of the word Back. This will bring up a list of pages you have visited. Click on any one of these choices to return to the desired page. This is the equivalent of clicking on the Back arrow several times.

To move forward:

- When you have returned to previous sites with the Back arrow, you can go forward again by clicking on the small right-pointing arrow next to the Back arrow. If you hold your mouse over this arrow, the title of the upcoming page will briefly appear.

- To move farther ahead, click on the small black triangle to the right of the Forward arrow in the menu bar at the top of the screen. This presents a list of several sites you have visited. Click on any of the choices to return to the desired site. This is the equivalent of clicking on the Forward arrow several times.

ADDITIONAL TOOLBAR OPTIONS

- **Stop:** The circle containing the X will stop a page while it is in the process of loading. This is useful if a page is not successfully or speedily retrieving.
- **Refresh:** The square containing the two curved arrows re-retrieves the page you are currently viewing. This is useful if the page does not load successfully or completely.
- **Home:** The home icon takes you back to the page that was on the screen when you first started IE 6. You can customize your selection.
- **Search:** The search button opens up a function that uses one or more Web search tools. You can choose the search tool(s) you want as your default. You can also customize your search experience. After clicking on Search, choose the Customize option and make your selection. A pop-up window called "Customize Search Settings" will appear. If you choose to "Use Search Assistant" broad search topics will be displayed and the appropriate search tool will be queried. You can also opt to have IE 6 remember your last 10 searches so that you can easily repeat them. Also notice that you can click a button called "Autosearch settings." This allows you to choose the search tool you want when you use the Address bar as a search window. You can also customize this option on the "When searching" line. You can even choose to turn off the use of the Address bar as a search window. If you do this, all words you type into the Address bar will be interpreted as URLs.
- **Favorites:** Favorites are Web sites you have visited that you would like to store for easy access. You can add, delete and organize your Favorites.
 - To *add* the current Web page as a favorite, click on Favorites and then Add. To choose the folder where you want to store this listing, click on Create in and choose the folder you want. At this point, you also have the option to create a new folder.
 - To *delete* a Favorite, simply right click on the item and choose Delete. Or, you can choose Organize Favorites select the desired item, and click on the Delete button.
 - To *move* a favorite to another folder, click on Organize Favorites, select the desired item, and click on Move to folder. In the pop-up window, select the folder where you would like to store this listing.
- **History:** The history function allows you to view and select Web pages you have recently visited. You can sort your items by clicking on the black triangle to the right of the word View. You can sort by size, date, the number of times visited, and the order you have visited today.
- **Mail:** You can read email from this window. Choose the email software you wish to use by going back to the Menu Bar and choosing Tools/Internet Options/Programs.
- **Print:** Allows you to print the current page. This option will be explained in more detail below under Printing.

- Edit: You may edit the current page in the HTML editor of your choice. Choose the editor by going back to the Menu Bar and choosing Tools/Internet Options/Programs.
- Discuss: You may set a default Usenet newsgroup server.

USEFUL OPTIONS ON THE MENU BAR

The menu bar at the top of the screen includes some useful options. Here are a few highlights.

- File/New/Window: You can open up a second copy of IE 6 by using this feature. This allows you to visit more than one Web page at a time.
- File/Edit with...: You can edit the current Web page using the editor of your choice. Select the editor by going back to the Menu Bar and choosing Tools/Internet Options/Programs. Your choices will be determined by software installed on your computer.
- Edit/Find (on This Page): IE 6 allows you to do a text search of the document on your screen. Choose this option and type in the word or phrase you wish to search.
- Tools/Show Related Links: IE 6 will display pages that are related in content to the current page. This is a service of Alexa, a Web content and traffic analysis company.

The Tools menu offers you many ways to customize IE 6. This will be covered below under Customizing Internet Explorer.

SAVING WEB DOCUMENTS FOR LATER USE: HOW TO DOWNLOAD, EMAIL, AND PRINT

You can download to disk, email, or print the Web page on the IE 6 screen.

To DOWNLOAD

1. Click on File/Save As (top left of screen). A pop-up window will appear.
2. Save in: Choose the desired drive.
3. Save as type: Make sure you save the page to the file type that will be useful to you. If you save the page as a Web page, you will need a Web browser or HTML editor to view it. A text file (txt) can be viewed in a word processing program such as Word or WordPerfect.
4. Click on Save

To EMAIL

1. Click on File/Send (top left of screen).
2. You may send the current page as an email message, or you may insert the link to the current page within an email message. Once you make your selection, your email software will open. You can change the default software by going to the Menu Bar and choosing Tools/Internet Options/Programs. Your choices will be determined by software installed on your computer.

To PRINT

To PRINT THE ENTIRE DOCUMENT

1. Click on the Print icon on the Tool Bar
2. Click on OK

To PRINT SELECTED PAGES

1. Click on File/Print Preview (top left of screen)
2. Click through the pages using the navigation arrows and make a note of which pages you want to print
3. Click on Print (top left of screen)
4. Click on the circle next to "Pages"
5. Type in the pages separate by commas, e.g., 1, 5-6, 7, 9
6. OR, to print the page displayed in the Print Preview window, choose Current Page.
7. Click on OK.

THE RIGHT MOUSE BUTTON

The right mouse button offers a number of useful features if you are using a PC. To view the possibilities, press down on the right mouse button and hold it. Options will display in a pop-up window.

The following is a selected list of right mouse button options.

1. WHEN THE MOUSE POINTER IS ON THE SCREEN (but not on a link or an image)
 - Back: Moves back to the previously visited page in your history list (same as Back icon)
 - Forward: Moves forward to the next page in your history list (same as Forward icon)
 - Select All: Selects all the text on the page for copying and pasting
 - Create Shortcut: Creates a shortcut to the current Web page on your desktop
 - Add to Favorites: Adds the current Web page to your Favorites
 - View Source: Brings up the HTML source code of the current page
 - Encoding: Allows you to choose a language
 - Print: Prints the current document
 - Refresh: Reloads the current page from the server
2. WHEN THE MOUSE POINTER IS OVER A LINK
 - Open: Opens the page
 - Open in New Window: Opens the link in a new copy of IE 6
 - Save Target As: Saves the link as a file
 - Print Target: Prints the link
 - Copy Shortcut: Copies the URL to the Clipboard for pasting into a text editor or word processing program
 - Add to Favorites: Adds the selected page to your Favorites
3. WHEN THE MOUSE POINTER IS OVER AN IMAGE
 - Save Picture As: Saves the image to a disk drive of your choice
 - E-mail Picture: Opens your default email program and attaches the image to the message
 - Print Picture: Prints the image on your default printer
 - Set as Background: Uses the image as your desktop wallpaper
 - Set as Desktop Item: Sets the image as an Active Desktop item
 - Copy: Copies the image to the Clipboard for pasting into a graphics editing program
 - Add to Favorites: Adds the selected images to your Favorites

CUSTOMIZING INTERNET EXPLORER

IE 6 offers a number of customization options. This section will highlight some of the more useful features available under Tools/Internet Options on the Menu Bar.

Tools/Internet Options is divided into six tabs. Each one is explained below.

1. General

- **Home Page:** Specify the URL of the page you want to appear whenever you open IE 6, or whenever you click on the Home icon
- **Temporary Internet Files:** This option allows you to view the files in your browser's cache. The cache holds viewed Web pages for subsequent quick viewing. Retrieving a file from the cache is much faster than repeated trips to the remote Web server where the file originated. You can customize the Settings to decide how often to check for newer pages, to specify how much disk space to reserve for your cache, and to view files in the cache.
- **History:** This option customizes your access to pages you have visited with the History function. Here you can set the number of days to keep pages in your history.
- **Colors:** Choose colors for links, visited links, and link hovers (the color appearing when your mouse is over a link). You can also set a default text and background color.
- **Fonts:** Select the language script, the font displayed in Web pages, and the font displayed as plain text.
- **Languages:** Select the language that will display Web pages accessed with IE 6.
- **Accessibility:** Choose to ignore colors, font sizes and font styles on Web pages. You can also set a style sheet as the display template for all Web pages viewed with IE 6.

2. Security

Here you can set levels of security for individual Web pages. See IE 6's Help menus for more information.

3. Content

- **Content Advisor:** You can enable ratings of objectional content to control the pages that may be viewed with this browser.
- **Certificates:** This feature allows you to manage the identification certificates you may have. See the Help menus for more information.
- **Personal Information:** This consists of two options. AutoComplete will store entered Web address, information entered into forms, and usernames and passwords needed to access sites you have visited. When you are using your browser, previous entries will come up as choices so that you don't have to retype the information. This can make your work go much faster. You can customize these options, and delete your settings. My Profile offers a template for entering personal information. If a Web site requests this information, you can give permission for it to be used.

4. Connections

Here you can store the information about your Internet Service Provider, configure your LAN settings, or send your browser requests through a proxy server.

5. Programs

Here you can set the programs you want the browser to use for HTML editing, email, Usenet news, collaboration ("Internet Call"), your calendar and contact list.

6. Advanced

This screen offers a number of options in the categories of accessibility, browsing, HTTP settings, Microsoft VM (Virtual Machine), multimedia access, printing, searching and security. Set these options if you are comfortable with them.

Checklist of Internet Research Tips

1. The Internet is a self-publishing medium. It is not a library of evaluated publications selected by professionals. Rather, the Internet is a bulletin board containing everything from the definitive to the spurious. Everything, *everything* must be analyzed for its appropriateness for research use.
2. Be sure to try out a handful of sites when researching a topic on the Internet. Do not rely on only one site or one type of site.
3. Three major resources for locating Internet materials are the subject directory, the search engine, and content on the deep Web. These are useful for different types of queries. Be sure you understand the differences.

SUBJECT DIRECTORY

For general, research oriented queries, and when you want to view sites often recommended by experts, use a subject directory.

Definition: A subject directory is a service that offers a collection of links to Internet resources submitted by site creators or evaluators and organized into subject categories. Directory services use selection criteria for choosing links to include, though the selectivity varies among services. Most directories are searchable.

When using subject directories, keep in mind that:

- There are two basic types of directories: *academic and professional directories* often created and maintained by subject experts to support the needs of researchers, and directories contained on *commercial portals* that cater to the general public and are competing for traffic. Be sure you use the directory that appropriately meets your needs.
- Subject directories differ significantly in selectivity. Consider the policies of any directory that you visit. One challenge to this is the fact that not all directory services are willing to disclose either their policies or the names and qualifications of site reviewers. This is especially true of directories located on commercial portals.
- Many people don't make enough use of subject directories, but instead go straight to search engines. Keep in mind that academic subject directories contain carefully chosen and annotated lists of quality Internet sites. Don't overlook subject directories when searching for quality on the Internet.

INFOMINE, from the University of California, is a good example of an academic subject directory. Yahoo! is a good example of a commercial portal--but it should never be used for serious research. A more complete list of both types of directories may be found on the page Internet Subject Directories.

SEARCH ENGINE

For targeted, complex, and sometimes general queries, use a search engine.

Definition: A search engine is a searchable database of Internet files collected by a computer program (called a wanderer, crawler, robot, worm, spider). Indexing is created from the collected files, e.g., title, full text, size, URL, etc.

There is no selection criteria for the collection of files, though evaluation can be applied to the ranking of results.

A search engine might well be called a search engine service or a search service. As such, it consists of three components:

- Spider: Program that traverses the Web from link to link, identifying and reading pages
- Index: Database containing a copy of each Web page gathered by the spider
- Search engine mechanism: Software that enables users to query the index and that usually returns results in relevancy ranked order

HotBot is a good example of a search engine. A more complete list may be found on the page Internet Search Engines.

DEEP WEB

For targeted queries, when you are looking for non-textual information, use the DEEP WEB.

Definition: The deep Web consists of information stored in searchable databases mounted on the Web. Information stored in these databases is accessible by user query. These databases usually search a targeted topic or aspect of a topic, though entire Web sites may be contained within a database. Search engine spiders cannot or will not index this information.

The deep Web also consists of multimedia and image files, and files created in non-standard file types such as Portable Document Format (PDF). Many search services offer separate search options for locating these files. AltaVista, Excite, HotBot and Lycos are just a few examples of services that offer specialized media searches, while Google integrates searches of PDF files into its general search service.

When dealing with the deep Web, keep in mind that:

- Many databases on the Web are searchable from their own sites. Therefore, a good directory will link to these sites. In addition, search engine results will also turn up database sites. Once you connect to the site, you can then search the database.
- There are Web sites that specialize in collecting links to databases available on the Web. One such site is actually called The InvisibleWeb and links to 10,000 Web-

accessible databases. You may also want to visit other sites that collect links to Web databases.

- Topical coverage on the Invisible Web is extremely varied. This presents a challenge, since it is impossible to anticipate what might turn up in a database. In addition, this coverage will be fluid as databases proliferate on the Web.
- Information that is dynamically changing in content will appear on the Invisible Web. Examples include news, job postings, available airline flights, etc.
- Directories are a part of the deep Web. Examples include phone books and other "people finders," list of professionals such as doctors and lawyers, patents, dictionary definitions, company data, etc.
- Yahoo! is one of the most popular sites on the Web. It is one of the Web's largest commercial portals. But it is not a reliable or adequate research tool and should *not* be used for this purpose. Beware of its drawbacks:
 - Yahoo! is merely the passive recipient of sites submitted to it
 - Yahoo!'s staff does not carefully evaluate content when choosing to add items to the database; therefore scholarly sites are haphazardly mixed in with everything else
 - Yahoo! tends to index only the major landing page of a site; therefore, any significant subsidiary pages on a related or different topic may not show up on this site
 - The editors at Yahoo! don't have time to review all submitted sites. Many sites are not added to Yahoo! because there simply isn't enough time.
 - It is very helpful to understand the principles of Boolean search logic when using a search engine on the Web. This search logic is manifested in three distinct ways on Web search engines..
 - When you enter more than one word in a Web search engine, the space between the words has a logical meaning that directly affects your results. This is known as the default syntax. For example:

In Google, a search on the words

birds migration

means that you will get back documents that contain *both* the words birds and migration. This is because the space between the words defaults to the Boolean AND. Most search tools nowadays default to AND logic.

If you wish to use Boolean OR logic to search for something like

"global warming" "greenhouse effect"

you will need to select OR logic from a search template or use an advanced search option that gives you additional choices.

- When using Web search engines, a de facto search language has emerged especially for basic search (i.e., main screen) interfaces. When in doubt, use the following syntax:
 - + for mandatory words: +birds +migration
 - phrases within double quotations: "human rights"

1. If you are looking to search without the hassle of perfecting a technique, try Ask Jeeves. This search service find answers to questions stated in plain English. Ask Jeeves works best with questions that have a precise factual answer.
2. Search engines offer numerous features that help you hone in on what you want. For a review of some of these features, and the search engines that support them, see [How to Choose a Search Engine or Directory](#).
3. Search engines return results in a schematic order. Most search engines use various criteria to construct a term relevancy rating of each hit and will present your search results in this order. Criteria can include: search terms in the title, URL, first heading, HTML META tag; number of times search terms appear in the document; search terms appearing early in the document; search terms appearing close together; etc. This is known as "on the page" ranking and applies to many long-standing **first generation search services**.
4. One of the most interesting developments in search engine technology is the organization of search results by peer ranking, concept, site and domain, rather than by relevancy. This type of ranking looks at "off the page" information to determine the order of your search results. Search engines that employ this alternative may be thought of as **second generation search services**. For example:
 5. [Google](#) ranks by the number of links from pages ranked high by the service
 6. [Teoma](#) ranks by the number of links from pages on the same topic for one of its results options
 7. [Guidbeam](#) sorts results into categories representing concepts derived from your search
8. A more detailed look at second generation search services may be found in the tutorial [Second Generation Searching on the Web](#).
9. Take advantage of second generation tools that return results in a *horizontal* presentation. Most search tools return results in one long, vertical list. In contrast to this, there is a group of search tools that use concept processing to return results in a horizontal organization. With these tools, you can first review concept categories retrieved by your search before examining the results within particular categories. This can make it easier to zero in on the aspects of your topic that interest you. Examples of these tools include [Guidbeam](#), [Query Server](#), and [Vivisimo](#).
10. Don't be impressed--or even necessarily worried-- by a large number of hits in response to a well-formulated search. Often multiple pages are returned from a single site because they all contain your search terms. [AltaVista](#), [Excite](#) and [HotBot](#) are among those that avoid this by a technique called *results grouping*, whereby all the results from one site are clustered together into one result. You are then given the opportunity to view all the retrieved pages from that site if you choose. With these engines, you may get a smaller number of results from a search, but each result is coming from a different site.
11. If you have too many search results, or results that are not relevant:
 - Add concept words
 - Use vocabulary that is specific to your topic, e.g., *Honda* rather than *cars*.
 - Link appropriate terms with the Boolean AND (+) so that each term is required to appear in the record
 - Use term proximity operators if they are available
 - Narrow your search to individual parts of the Web page such as title, first page level, etc.
 - Use the Boolean NOT to keep out records containing terms you don't want

12. If you have too few search results:

- Drop off the least important concept(s) to broaden your subject
- Use more general vocabulary
- Add alternate terms or spellings for individual concepts and connect with the Boolean OR
- Try the option available on some engines to find related documents to one or more of your relevant hits. [Google](#), [ProFusion](#) and [Raging Search](#) all offer this type of feature.

13. Meta search engines simultaneously search multiple search engines. They are also referred to as parallel search engines, multithreaded search engines, or or mega search engines. These are useful when:

- you have an obscure topic
- you are not having luck finding anything when you search
- your search is not complex
- you want to retrieve a relatively small number of relevant results

There are *two* types of meta search engines:

1. One type searches a number of engines and does not collate the results. This means you must look through a separate list of results from each engine that was searched; you will often see the same result more than once. Some engines require you to visit each site to view your results, while others will fetch the results back to their own sites. When results are brought back to the site, a certain limitation is placed on what is allowed to be retrieved (more on this in the next point). With this type of meta search engine, you can retrieve comprehensive, and sometimes overwhelming, results. An example of this type of engine is [Dogpile](#).

2. The other type is more common and returns a single list of results, often with the duplicate hits removed. This type of meta engine always brings the results back to its own site for viewing. In these cases, the engine retrieves a certain maximum number of documents from the individual engines it has searched, cut off after a certain point as the search is processed. The cut-off may be determined by the number of documents retrieved, or by the amount of time the meta engine spends at the other sites. Some of these services give the user a certain degree of control over these factors. All of this has two implications:

- These meta search engines return only a portion of the documents available to be retrieved from the individual engines they have searched
- Results retrieved by these engines can be highly relevant, since they are usually grabbing the first items from the relevancy-ranked list of hits returned by the individual search engines

The better meta search engines remove duplicate files and give you some information along with the document title. To see a list of meta search engines, visit [Internet Search Engines](#).

14. Keep in mind that search engines do not index all the documents available on the Web. For example, most search engines cannot index files to password-protected sites, behind firewalls, or configured by the host server to be left alone. Still other Web pages may not be picked up if they are not linked to other pages, and are therefore missed by a search engine spider as it crawls from one page to the next. Search engines rarely contain the

most recent documents posted to the Internet; do not look for yesterday's news on a search engine.

15. Finally, watch for converging content. Many well-known sites now contain information from an array of sources. This can increase the usefulness of search sites, but also create confusion in terms of the information source. For example:

- Spider gathered index: The mechanism for searching a spider-gathered index is the feature people usually associate with a search engine.
- Results from other search services: It is increasingly common for a search engine to return results from other services with which it has partnered. Examples include [Ask Jeeves](#), [AOL Search](#) and many others. Each partner service offers an enhancement over the more traditional term ranked results. This represents an interesting combination of first and second generation search technologies appearing on the same site.
- Directory: Many search services offer a directory on their sites. This directory may be a name brand such as LookSmart or the Open Directory Project, or a directory compiled by a site's own editors. Results from the directory may appear automatically with results from the spider-crawled Web, or the directory may be searched or browsed separately.
- Deep Web: Many search services offer the option to search databases offering specific content. Included may be news, business, shopping, multimedia files, and so on. These databases constitute a small subset of the deep Web.