UNIT 12   INTERNET SERVICES

Structure

12.0   OBJECTIVES

After reading this unit you will be able to:

● understand what is World Wide Web (W3);

● know the differences between search engines and meta-search engines;

● Get acquainted with various applications of a Web Browser; and

● understand the features and usage of e-mail.

12.1   INTRODUCTION

Why do people use the Internet? A few reasons could be: distance learning, looking up factual information, researching topics, searching libraries, networking with colleagues and friends, improving workplace productivity conducting business transactions worldwide, and for personal entertainment. The Internet is a medium

for communication and accessing the information in digital form from a vast resource of databases.

The Internet allows each of us to share information and resources such as government documents, electronic publications, online books, media publishing, human anatomical images, computer software, bibliographic and full-text databases, speeches, live concerts, animal sounds, and motion videos. The Internet's dynamic nature is derived from world citizens contributing their time, resources and energies to each other.

Typical users consult electronic resources at near and distant libraries, download computer shareware and software upgrades, read and print out government publications, make travel   arrangements and purchase goods and services. "What can the Internet do?" is a usual question. This question is never satisfactorily answered. The Internet allows us to be in contact with a lot of people. There are tools, such as electronic mail and news groups, to assist us in communicating with people who are interested in our topics. They are often more than eager to help out, providing answers to questions and engaging us in thoughtful discussions. A few usages of the Internet can be listed below:

- Retrieving information from reference sources like dictionaries, encyclopacdias, etc., required for our day today working from millions of such documents or files throughout the world.

- Retrieving information from databases of various large libraries like the Library of Congress, the British Library, the Indian Institute of Science library, the Parliament of India Library and several other large libraries.

- Searching databases of governments, individuals and organizations for files on topics of interest.

- To access electronic books, journals and other documents required to support any research work.

- Referring social and economic statistical data, such as census information, daily exchange rates, and government budgets and reports.

- Getting documents on fine arts and music, including digital images of art video and audio.

- Exchanging messages with any of the tens of millions of people across the world.

- Searching for and bringing to our computer shareware, freeware, and commercial software.

- Send or receive sound, animation and picture files from any place in the world.

- Setting up temporary or permanent discussion or work-oriented groups.

- Distributing or reading electronic newsletters, newspapers, bulletins and similar publications and products and services.

- To trade with people or other organizations.

Helper applications are external software programmes that allow Web users to view or hear multimedia presentations, regardless of platform. Also known as plug-ins, helper applications extend and enhance the capabilities of web browsers such as Netscape and   Internet Explorer, and are needed to handle many of the newer hypermedia such as streaming audio, vector graphics, three-dimensional multimedia and virtual worlds. Browsers hand data to appropriate helper applications such as RealAudio, Adobe Acrobat, QuickTime, Shockwave and others.

Web users move effortlessly from Internet site to Internet site by clicking on embedded links. Links or hyperlinks are transparent connections to and from various resources on the same server, or to external web sites near or far. Links are usually highlighted, flashing, underlined or emphasized in some way with different typefaces or colours. Some web sites feature clickable icons, bars or images. Links make it easy to move through the web and to explore. Some people have likened hyperlinks to the way our minds work, and as we may think.

12.2   WORLD WIDE WEB (WWW)

The World Wide Web- know as WWW, W3 or simply, the Web - is one of several Internet resource discovery tools developed to help people publish, organize and provide access to information on the Internet. The Web was first developed by Berners Lec in 1989 while working at CERN, European Particle Physics Laboratory in Switzerland, and has since become the most powerful, and popular, resource discovery toll on the Internet. The WWW can be defined as a hypertext, multimedia, distributed information system that provides links to hypertext documents, as well as to many other Internet tools and databases. There are several features that are unique to the Web that make it the most advanced information system to appear on the Internet to date.

12.2.1   Importance of the Web

The World Wide Web is important for libraries because it provides an extremely powerful method of organizing, and providing access to, information. It can provide one interface to a large variety of network information resources and systems. With the Web and its browsers libraries can:

- electronically publish anything that they now publish on paper

- provide access to in-house hypertext documents or to hypertext documents available out on the Internet

- create electronic orientation services with floor maps and descriptions of services

- provide access to Internet tools such TELNET, gopher, FTP, and WAIS through a single interface.

- create interfaces to in-house databases or bibliographies

- collect information from patrons through the forms feature.

With all its power, the Web is the most serious step yet towards creating electronic libraries. It provides a mechanism to present a wide variety of information resources to library users-and all Internet users-in a simple, efficient and effective manner.

i)   *The Web is a hypertext system:* In contrast to the hierarchical menu system used by earlier Internal tools such as Gopher, etc, the Web is a hypertext system, in which users move from one document to related documents through embedded links, such as a word or a phrase that, when selected, calls up another document on that topic. Instead of moving from menu to menu, as in Gopher, users of the Web can jump directly from document to document by clinking on hypertext links.

ii)   *The Web is a multimedia system:* Before graphical Web browsers (e.g., Mosaic and Netscape, etc), most of the information available on the Internet

was in the form of ASCII text - simple text devoid of any elements common to the printed page, such as large titles, italics, pictures, and other graphical content. With the advent of graphical browsers, however, the Web has become a multimedia system, combining many different types of media into one document. Specifically, web documents can contain:

- Normal text

- Features such as large fonts, bold, italics, indents

- Images such as pictures, graphics, fancy logos, illustrations

- Audio content such as sounds, music, commentary, voice messages

- Video content such as movie clips, animations, or computer generated simulations.

iii) *The Web is a distributed system:* Normally, the documents in a hypertext system reside in one place, such as on a hard drive or a CD-ROM. Documents on the Web, by contrast, can be located not only on a local machine, but also on any machine on the network whether next door, across a city, or around the world. From the perspective of the user, one set of related documents may appear to reside in one location, but in reality, the successive pages they read may have been requested from anywhere in the world. In this way, the Web is a distributed hypertext system because the Web documents are distributed throughout the network.

iv) *The Web incorporates other Internet tools:* The Web can provide links to other types of Internet tools, such as WAIS, Gopher, FTP and TELNET. One Web page on a particular topic, for example, can point to other relevant information resources on the network, regardless of whether that information is available on a gopher, through TELNET, or at an FTP site. In this way, the Web and its browsers become a method to seamlessly provide access to information available through many different Internet tools.

v) *The Web provides an interface to other database systems.* A particularly powerful features of the Web is that it can act as an interfaces to database systems connected to the Internet (e.g., WAIS, Z39.50 and library databases). Three elements are needed to create this interface. First, a Web browser which is used to collect information from the user through the forms processing feature - a method of creating interactive boxes on a Web page into which users can type information, or select among alternatives. Second, a database system of some sort, like a library catalogue. And third, a control gateway interface (CGI) that sits between the Web browser and database. It takes the information gathered from the Web browser and passes it on to the database. Once the request is processed, the CGI passes the results back to the Web browser in a format that it can display.

12.2.2　How does the Web Work?

The most important things required to understand the underlying mechanism of the Web are the client-server architecture, the Hypertext transfer protocol (HTTP), Hypertext Markup Language (HTML), and Universal Resource Locators (URLs). These concepts are described below.

i) **Client-server architecture**

Like most Internet tools, the Web adheres to the client-server model in which two separate software programmes- really, two complementary halves of one system-work together to perform some specific task. The software on the user's computer is the client, while the software on the remote computer is the server. In the case of the Web, the task is to explore hypertext documents and examine information resources. The Web client does the job of asking for, and displaying electronic documents, and the Web server does the job of storing and sending electronic documents to the client.

ii) **Hypertext transfer protocol**

In order to communicate through the network, clients and servers need a common language, and the common language within the Web is Hypertext Transfer Protocol, or HTTP. It is a fairly simple protocol, like the Gopher protocol, in which a document and its elements (e.g., text and images) are transferred through separate, brief network connections. Once all the needed parts of a document arrive, the connection is broken and there is no more interaction with the server until the next request is made.

iii) **Hypertext links-Uniform Resource Locators**

What about the links themselves? Links are based on a standard called a Uniform Resource Locator (URL). URLs contain all of the information needed for the client to find and retrieve a HTML document. An example of a URL is:

<div align="center">http://www.zodiac.ca/htdocs/home.html</div>

It has four parts: 1) the protocol used to connect to the remote server. In this example, the protocol is HTTP, the protocol used to connect to Web servers. The protocol could also be Gopher, FTP, or TELNET, indicating that the link is to one of these Internet tools, 2) the Internet address of the server where the document resides. In this case, the address is www.zodiac.ca', 3) the directory on the server where the document is located, called the document path. In this case, the path is /htdocs', and 4) the filename of the document itself. In the example, it is home.html, where the html extension indicates that the document is marked up with HTML.

12.2.3 Web Servers

Web architecture contains two classes of computers: servers, which store information and clients which access it. Together, these are the two ends of the Web, each with its own supporting software. Server software runs exclusively on server machines, handling the storage and transmission of documents. In contrast, client software such as Netscape Internet Explorer, etc. runs on the end-users' computers accessing, translating and displaying documents. The two ends communicate through the Internet using a protocol known as Hypertext Transfer Protocol (HTTP).

A web server is a software package that processes HTML documents for viewing by Web browsers. The server enables users on other sites to access documents that you have chosen to make available to them.

Web servers can be run from any hardware platform. There are servers that are specifically designed for Macintosh computers, PCs, Silicon Graphics, and various other platforms, including the Amiga. The most important software is the web server itself. Just as it can be run under a number of hardware platforms, it can also run

under several operating systems, including MS Windows, Windows NT, Unix, and VAX's VMS.

The web server is responsible for document storage and retrieval. It sends the document requested (or an error message) back to the requesting client. The client interprets and presents the document. The client is responsible for document presentation. The language that web clients and servers use to communicate with each other is called the Hypertext Transfer Protocol (HTTP). All web clients and servers must be able to speak HTTP in order to send and receive hypermedia documents. For this reason, web servers are often called HTTP servers, or HTTP Daemons (HTTPD).

There are many different web servers, running on many platforms. In this example you are using a browser (Netscape, or Mosaic, or some other browser) and you click on a reference to Webmaster Magazine Online. The browser is able to figure out that what you really want (in web terminology) is the object the web knows as http://www.cio.com/WebMaster/wmhome.html-the Web address (or URL) for WebMaster Magazine Online. WWW documents transmitted between client and server are written in a text-formatting language, called Hypertext Markup Language (HTML) that describes the structure and character formatting data of documents. The key to remember is that HTML is a language, neither an application nor a software package. It is simply a data-set of text and instructions that requires additional client software to be used.

12.2.4   Web Browsers

A web browser is the software programme you use to access the World Wide Web, the graphical portion of the Internet. The first browser, called NCSA Mosaic, was developed at the National Center for Supercomputing Applications in the early '90s. The easy-to-use point-and-click interface helped popularize the Web, although few then could imagine the explosive growth that would soon occur. The two hottest graphical browsers battling for the top spot are Microsoft Internet Explorer and Netscape Navigator. Both are fast and both have integrated audio and video. Microsoft Internet Explorer can be downloaded at no charge. All of Netscape's products that are available via download, including Netscape Navigator, are free to anyone who qualifies as an educational user.

A web browser is a client programme that uses the Hypertext Transfer Protocol (HTTP) to make requests of Web servers throughout the Internet on behalf of the browser user. The software that enables you to go from one resource to another by following hyperlinks is known as a web browser. The basic capabilities of a browser are to retrieve documents from the web, jump to links specified in the retrieved document, and save and print the retrieved documents.

12.2.5 Mark up Languages

The most common method of creating a web documents is through the use of Markup Languages. Most of these are created by adding a set of formatting code to ASCII text to show fonts, justifications, links, etc. on the web. A markup language is a language that uses tags to indicate a change in presentation style or a change in content type. A mark up language describes explicitly. In sum, HTML describes explicitly the parts of the hypertext multimedia document, and the web client uses this information to construct the document on the user's screen.

12.2.5.1 Standard Generalized Markup Language (SGML)

SGML (Standard Generalized Markup Language) is the first known standard for how to specify a document markup language or tag set. SGML is based somewhat on earlier generalized markup languages developed at IBM, including General Markup Language (GML) and ISIL. SGML is not in itself a document language, but a description of how to specify one. It is a Document type definition (DTP).

SGML is based on the idea that documents have structural and other semantic elements that can be described without reference to how such elements should be displayed. The actual display of such a document may vary, depending on the output medium and style preferences. Some advantages of documents based on SGML are:

● They can be created by thinking in terms of document structure rather than appearance characteristics (which may change over time).

● They will be more portable because an SGML compiler can interpret any document by reference to its document type definition (DTD).

● Documents originally intended for the print medium can easily be re-adapted for other media, such as the computer display screen.

12.2.5.2 Hypertext Markup Language

Documents on the web are generally coded in a markup language called Hypertext Markup Language, or HTML. HTML is derived from Standardized Markup Language (SGML)-a standard for adding tags that identify elements in a document such as headings, subheadings, chapters, paragraphs, and appendices. Like SGML, HTML is also used to describe the structure of a document. A number of HTML codes embedded in the document text explicitly describe the text, providing information to the web client about how to interpret it. For example, if the text has a heading code surrounding it, as in <h1>This is a title</h1>, then the web client knows that it should display this text by itself on one line in a large, bold face font. Other types of information that HTML code provides the web client include:

● whether text is bolded or italicized

● where line breaks and paragraph breaks are

● whether a picture or graphic should be inserted, and where

● whether a certain word or phrase on the page is a link to other documents

HTML is based on SGML (Standard Generalized Markup Language), a much bigger document-processing system. You can think of HTML as being a sub-set of SGML. The primary focus of HTML is the **content** of the document, **not its appearance.** It is a language for describing structured documents. Most documents have common elements, for example, titles, paragraphs, lists. If you define the set of elements before you start writing, you can label the parts of the document with the appropriate name. These elements are labeled through the use of HTML **tags.**

HTML does not describe the appearance or layout of the document the reason being that different browsers (Netscape, Mosaic, etc.) will "read" the HTML information differently. The same document may appear differently in each browser. The number one prevailing rule of designing documents for the web is: **Don't design**

your documents based on what they look like in one browser. Focus on providing clear, well-structured content that is easy to read and understand.

12.2.5.3 Dynamic HTML

The Dynamic HTML is a collective term for a combination of new Hypertext Markup Language (HTML) tags and options, that will let you create web pages more animated and more responsive to user interaction than previous versions of HTML. Much of dynamic HTML is specified in HTML 4.0. Simple examples of dynamic HTML pages would include (1) having the color of a text heading change when a user passes a mouse over it or (2) allowing a user to "drag and drop" an image to another place on a web page. Dynamic HTML can allow web documents to look and act like desktop applications or multimedia productions.

The features that constitute dynamic HTML are included in Netscape Communications' latest Web browser, Navigator 4.0 (part of Netscape's Communicator suite), and by Microsoft's browser, Internet Explorer 4.0. While HTML 4.0 is supported by both Netscape and Microsoft browsers, some additional capabilities are supported by only one of the browsers. The biggest obstacle to the use of dynamic HTML is that, since many users are still using older browsers, a web site must create two versions of each site and serve the pages appropriate to each user's browser version.

12.2.5.4 Extensible Markup Language (XML)

XML (Extensible Markup Language) is a flexible way to create common **information** formats and share both the format and the **data** on the World Wide Web, intranets, and elsewhere. For example, computer makers might agree on a standard or common way to describe the information about a computer product (processor speed, memory size, and so forth) and then describe the product information format with XML. Such a standard way of describing data would enable a user to send an intelligent agent (a programme) to each computer maker's web site, gather data, and then make a valid comparison. Any individual or group of individuals or companies that wants to share information in a consistent way can use XML.

XML, a formal recommendation from the World Wide Web Consortium **(W3C),** is similar to the language of today's web pages, the Hypertext Markup Language **(HTML)**. Both XML and HTML contain **markup** symbols to describe the contents of a page or file. HTML, however, describes the content of a web page (mainly text and graphic images) only in terms of how it is to be displayed and interacted with. For example, the letter "p" placed within markup tags starts a new paragraph. XML describes the content in terms of what data is being described. For example, the word "phonenum" placed within markup tags could indicate that the data that followed was a phone number. This means that an XML file can be processed purely as data by a programme or it can be stored with similar data on another computer or, like an HTML file, that it can be displayed. For example, depending on how the application in the receiving computer wanted to handle the phone number, it could be stored, displayed, or dialed.

XML is "extensible" because, unlike HTML, the markup symbols are unlimited and self-defining. XML is actually a simpler and easier-to-use subset of the Standard Generalized Markup Language (SGML), the standard for how to create a document structure. It is expected that HTML and XML will be used together in many web applications. XML markup, for example, may appear within an HTML page.

Early applications of XML include Microsoft's Channel Definition Format **(CDF),** which describes a **channel,** a portion of a web site that has been downloaded to your hard disk and is then updated periodically as information changes. A specific CDF file contains data that specifies an initial web page and how frequently it is updated. Another early application is ChartWare, which uses XML as a way to describe medical charts so that they can be shared by doctors. Applications related to banking, e-commerce ordering, personal preference profiles, purchase orders, litigation documents, part lists, and many others are anticipated.

12.2.5.5 Virtual Reality Modeling Language (VRML)

VRML, often pronounced "ver-mull," is Virtual Reality Modeling Language, the open standard for virtual reality on the Internet. You can use VRML to create three dimensional worlds, representations of information, and games. AS an open standard, no one particular company controls the VRML specification, that is, the language definition. Theoretically, anybody can use VRML to write software or worlds without having to license technology from others. Using VRML, you can build a sequence of visual images into Web settings with which a user can interact by viewing, moving, rotating, and otherwise interacting with an apparently 3-D scene. For example, you can view a room and use controls to move the room as you would experience it if you were walking through it in real space.

To view a VRML file, you need a VRML viewer or browser, which can be a plug-in for a Web browser you already have. Among viewers you can download for the Windows platforms are blaxxun's CC Pro, Platinum's Cosmo Player, WebFX, WorldView, and Fountain. Whurlwind and Voyager are two viewers for the Mac. Virtual reality refers to an immersive environment, an environment that you feel you are inside of. You can attain this immersive feeling with computers using 3D graphics and audio. Sounds in a virtual world can be specialised so that they sound louder when you are closer to them.

When virtual reality happens on the Internet, new possibilities arise for distributed, networked virtual environments. In HTML, in-line images let you include graphics from anywhere on the Web on your Web page. In VRML, you can have in-line parts of a virtual world, so that a chair in a VRML world can be a VRML. URL on a server in France, while the garden comes from a server in Japan, and the soundtrack is a URL. On a server in England. In addition, hyperlinks from an object in a VRML, world can lead to another URL on the Web, which could be another VRML world, an HTML page, or any other URL!

Uses of VRML

There are many applications for VRML, varying in focus from VRML's open 3D file format, to its networking capabilities, to its multimedia nature. Here are some applications for which people are currently using VRML:

- Computer-aided design (CAD),
- Scientific simulations
- Games
- Data visualization
- Distributed, multi-user environments
- Social computing

- User interfaces to information

- Financial applications

- Product marketing and advertising

- Education

- Entertainment.

12.2.6 Search Engines

The Internet can be said to be the most exhaustive, important and useful source of information. Data on almost all the topics on earth are hosted in millions of servers connected to the Internet around the world. There is, however, no set policy for hosting the data or no centralised database for organization and searching the desired information. This makes the Internet most unorganized source of information. With the availability of excessive information it has become very difficult for the common user to find the precise information, required. There are, however, a variety of search tools for resource discovery on the Web. These include apart from others the search engines. The search engines can be defined as the tool created for finding, classifying and storing information about various Websites on the Internet. These can help in locating information of relevance on a particular subject by using various search methods.

A search engine is a service that indexes, organizes, and often rates and reviews Web sites. It helps you find the needle that one Web site you've got to see in the Internet haystack. Different search engines work in different ways. Some rely on people to maintain a catalogue of Web sites or pages. Some use software to identify key information on sites across the Internet. Some combine both types of service. So when you search their "holdings," you're bound to get different results.

Search engines are online utilities that quickly search thousands of Web documents for an entered word or phrase. Although there are some subscription-based search engines, most operate off profits from advertisements. It should be noted that no single search engine has the contents to every Web page on the Internet. Instead, each search engine has the contents to only the Web pages that are manually entered into that search engine by the Web page operators. This is why you might get different results from different search engines.

Search engines are usually accessed through Web browser software. Each search engine provides different searching options and has its own look. Search engines also differ greatly in the number of resources they allow you to search. Some search engines have both searching and browsing capabilities.

12.2.6.1 How do you Find Information on the Web?

The Web is quickly becoming a vast information space that needs search tools to find information efficiently. In the same way that Veronica grew out of a need to be able to search an ever-growing system of gophers, search tools also have been developed to search the Web. The most common form of Web search tools are what are known as World Wide Web robots, wanderers, and spiders. These systems move through the network, automatically searching various elements of HTML documents (e.g., titles) looking for key words and building indexes. Users interact with the search tools using a Web browser. Through the form feature, users enter search terms, limit the number of hits, and provide other information that the robot uses in its search. Once activated, the robots search their indexes-or in some cases,

HTML pages the network in real time-and then dynamically build an HTML page containing links to relevant pages. Some examples of Web robots include the World Wide Web Worm, the WebCrawler Index, Lycos, the Jumpstation Robot, and Aspider.

The robot form of search engines are intensive users of system resources-both network bandwidth and server CPU cycles-because they are constantly roving the network, connecting to Web servers, building indexes and doing real time searches.

12.2.6.2 Categories of Search Engines

Major search sites generally provide more than a search engine for finding Web sites. They also allow you to look up information such as recent news stories, newsgroup postings, reference material (such as dictionary entries and maps), and e-mail addresses, street addresses, and telephone numbers of businesses and individuals. Search tools can be categorized as:

- Webcrawlers

- Subject directories

- Web databases: indexes

i) *Web Crawlers*

A Web Crawler (sometimes called a spider) is an automated search engine. When someone submits a resource(website) to the search engines it sends a small, but powerful programme back to the site of the submission which scans the site for more resources. If it finds additional resources, it catalogues them as well as the original resource. The advantage of this is you can find information in a site, even if the site is not dedicated to that specific type of information. For example, locating information on Airplanes in a site about NASA space activities.

The main drawback to this type of search engine is the tendency to catalogue too much information. A couple of examples of the best crawlers are:

- HotBot   (http://www.hotbot.com/)
- Infoseek   (http:///ultra.infoseek.com/)
- Alta Vista   (http://www.algtavista.digital.com/)

ii) *Subject Directories*

Directories are the Yellow Pages of the Internet. They contain only that information which has been submitted to them. Here you will find vast listings of resources, but if the person submitting the information didn't include all of the description of his/her site, you may not find what you are looking for.

- YAHOO!   (http://www.yahoo.com)
- Linkstar   (http://www.linkstar.com/)
- InfoHiway   (http://www.infohiway.com/way/index.html)

Subject directories are manually compiled by the staff of the directory and by users who submit entries. Web databases are compiled by software "robots" or "intelligent agents" that roam the Web and collect information for the databases. Both approaches to search engines have created effective products. The search engines highlighted here are currently popular or have historical interest. However, new search engines may be worth using, so be watching for them.

Directories, such as Yahoo! Are good at identifying general information. They group websites under similar categories, such as internet tutorials, English universities and Paris museums. The results of your search will be a list of websites related to the subject you are searching for. If you are interested in locating the site for the Louvre museum, for instance, try using a directory.

iii) *Web Database indexes*

But let's say you want more specific information, such as biographical information about Leonardo da Vinci. Web indexes are the way to go, because they search all the contents of a website. Indexes use software programmes called spiders or robots that scour the Internet, analyzing millions of web pages and newsgroup postings, indexing all of the words.

Indexes like Alta Vista and Lycos find individual pages of a website that match your search, even if the site itself has nothing to do with what you are looking for. You can often find unexpected gems of information this way, but be prepared to wade through a lot of irrelevant information too. Search results are usually ranked in order of relevancy-the number of times the search terms you used appear in a document, or how closely the document appears to match a concept you have entered. This is a much more thorough way to locate what you want.

Searching on the World Wide Web can be confusing. A myriad of search engines exist, often with little or no documentation, and many of these search engines work differently from the standard commercial search engines we normally use. There are also many directories that attempt to organize the Internet by subject, and, today, there are many search engines that combine directory and keyword search capability.

12.2.6.3 How to Search

Browsing a directory is a simple matter of following the links for the topic of interest. Searching either a directory or the portion of the web that a search engine covers works very much the same in almost all search engines. The basic format is that of a dialogue box, pane, or line where search terms can be entered followed by options to either submit or clear the search.

Once the search request is received, the search engine searches its own indexed database first, then, based on design, sends out spiders or other robots to add to the database. Results are sent back to the searcher, some annotated extensively, with links to the sources retrieved.

Full featured search engines also have options to expand or limit searches in a variety of ways. For example, in Lycos, the basic search assumes a boolean "or", which means that two or more terms will return results if any of the terms occur in documents indexed by Lycos. To obtain documents containing all the terms in a search, the Enhance Your Search option must be chosen and adjustments made to the default options.

12.2.6.4 Choosing a Search Engine

Choosing a search engine depends on the results you're looking for, though there are some criteria that may be useful. These criteria include:

- **Browsability—** how easy is it to understand the results? Do you receive enough information from the retrieved results to make a decision about the usefulness of the results?

- **Customizability**— can you construct a sufficiently detailed search so as to eliminate or greatly reduce irrelevant results?

- **Relevance**— no matter how browsable or customizable, are the results returned relevant to your search?

For example, searching for some information on the Native American squash blossom design using WebCrawler will bring relevant results, but either OpenText or InfoSeek would be better first choices because they both give more information to help you determine relevancy.

12.2.6.5 Functioning of Search Engine

How a search engine organizes its findings varies greatly between programmes. It is common for search engines to offer indexes of their contents to users. When we go to a search engine site, we can peruse the index rather than entering search terms. These indexes may not have a human overseeing them; the robot might just automatically assign a new site to a place in the index. We cannot rely on the index to contain cross-references.

Because of the automatic inclusion, search engines are good for bulk return but not qualitative content. We may have to search several engines to find a site, because one search engine cannot cover the entire Internet although some have come close.

Keywords: A search engine will give us a form to fill out. We enter in words as we would like it to look for these words which might be called key words or search terms. However, in a broader sense, key words are any search terms. However, in a broader sense, keywords are any search term we put in. The search engine then looks through its database listing for any matches. It might look in the text of a Web page, the title of a Web page, or simply the address. It then gives us a list of sites that match our request.

Boolean Logic: One way to avoid this is by using Boolean logic. Many search engines allow us to use "and", "or", and "not" to narrow or broaden a request. To find sites that contain all our key words, our search term would look like this: "University and California and Los Angeles". We would still get a lot of irrelevant sites, but at least not as many as before.

We would use "and" to narrow a search, "or" to broaden it, and "not" to refine it. "Cowboys not football" would help us to find sites that deal with the Old West heroes and not the football team, which is also named cowboys. Search request terms can become quite complex, but most search engines are not designed to accommodate the longer ones.

Plus and Minus Signs: Plus and minus signs are another way in which we can refine our research. Some search engines use these signs to indicate which words must be in a document and which ones must not. For example, "+cow-boys-football" would tell the software that we want sites that include the word "cowboy" but not ones that have the word "football".

Search Results: Search engines usually give us a list of best matches to least likely matches in descending order. The software has been designed to rank the sites in likely order of relevance. It does this by assigning a number to each word in a Web site title, address, and perhaps to all the words in a document. It's then a matter of

matching numbers to our key words. Some programmes are more sophisticated than others.

Even if the search engine we are using does not allow limiting factors, such as Boolean logic, we will generally get the sites that are closest to our request first. This helps; however, we still may have to go through several dozen-site listings, depending on how specific our original key words were.

Result Lists: Generally, we will receive a list of 15 sites. At the bottom of the list will be an option to click for more hits. The software may have a limit to how many sites it retrieves. It might give us the first 100 sites it found that matched our key words. There may still be other sites we would be interested in, however, they did not get listed. Some search engines allow us to change the amount of sites listed.

As search engines allow for specific requests that let us fine-tune our research, they will become more valuable. Many people use search engines but sift through countless lists of irrelevant sites. If we don't find one useful site in the first 50 listed, chances are that the search request we put in is not going to give us the results we are looking for. We should start all over again.

Result Descriptions: Some search engines offer descriptions of sites to help us determine whether the link is worth following. Others will include the first few paragraphs of a document. We might be able to choose the level of description we get with a listing. Many search engines simply give us a listing with little explanation.

The following table lists major search engines, directories, and archives with brief follow-on help for using them well. There have been major changes in most of the search engines in the past few months. To take full advantage of them, be sure to read their help files regularly. There is a great deal of information available on the web that the major search engines cannot find.

Name	URL	Comment
Altavista	http://www.altavista.com	250+ million pages and six million articles from 50,000 Usenet newsgroups. Full text search of pages. Good for academic research.
Excite	http://www.excite.com	130+ million pages and directories. Full text search.
Hotbot	http://www.hotbot.com	110+ million pages, WWW, Usenet newsgroups, most keywords on pages indexed.
Infoseek	http://www.infoseek.com	100+ million pages, web sites, but also images, newsgroups, faqs, news, companies, phone numbers, and email addresses, current news and Directories.
Lycos	http://www.lycos.com	Directory, reviews of the best Web sites, browse-able by

		category and rated on a scale of 1-50.
Northern Light	http://www.northernlight.com	Search web sites, articles or a special collection (over 1 million articles from books, magazines, journals, newspapers etc). Sorts results by relevance, and removes duplicate links.
Webcrawler	http://www.webcrawler.com	Search web sites, usenet, gopher, ftp. Searches only titles, keywords and URL's.
Yahoo	http://www.yahoo.com	Directory, useful for browsing, general searching and locating popular sites. 14 broad categories (such as computers/ internet, news and media, regional, society and culture).

12.2.7   Meta Search Engines (Search Engines for Search Engines)

Metadata in its broadest sense is data about data. The familiar library catalogue record could be described as metadata in that the catalogue record is 'data about data'. Similarly database records from abstracting and indexing services are metadata (with a difference variation on location data). However the term metadata is increasingly being used in the information world to specify records, which refer to digital resources available across a network. This means a metadata record refers to another piece of information capable of existing in a separate physical form from the metadata record itself. Metadata also differs from traditional catalogue data in that the location information is held within the record in such a way as to allow direct document delivery from appropriate application software, in other words the record may well contain detailed access information and the network address(s).

A metasearch engine queries two or more search engines at the same time. It sends the results from the search engines back to you. Metasearch engines do not have their own database or index web pages. They send their queries to search engines. The results from each search engine are combined. Duplicates (pages that are the same) are removed or merged. Pages are sometimes ranked in order of importance. Metasearch engines are good for finding information quickly. They are best when keywords are used. You can use metasearch engines to find out which search engines have the information you want. Metasearch engines do not return all the pages found. They only return the top ten or so pages. This means you can miss good pages using a metasearch engine. When you use a metasearch engine, you tell it what search engines to use. Metasearch engines use simple queries.

Most of these engines work on the principle of taking the highest ranking hits from the various engines they pass their requests on to, limiting that selection either by number of bits or sometimes by time elapsed in response. Results can then usually be grouped either by site, keyword, or other parameters, allowing the user to easily analyse what has appeared where and how it has been ranked. This is a very easy

way of broadening coverage but minimizing the effect of massive recall killing off precision: key: sites will hopefully appear in more than one system and thus be flagged up, but lesser-known sites should still appear on at least one system near the top if truly relevant. Among the most popular meta searchers are: all4one, All-in-one, MetaCrawler, SavySearch, northerlight.com and cyber411, etc.

12.3   INTERNET APPLICATIONS

The Internet is the largest, must vast, most complex but unorganised learning source in the world. Through the Internet one can find knowledge resources that allow one to study virtually any discipline imaginable. Not only that, one can communicate quickly and effectively with others who are also interested in the same discipline. Teachers, students, and other educators can share ideas instantly across vast distances.

Individuals, companies, and institutions use the Internet in many different ways. Business Houses use the Internet to provide product information, online support service, etc. Companies carry out online trading, including advertising, selling, buying, distributing products, and providing after-sales services. Institutions use the Internet for voice and video conferencing and other forms of communication that allow people to telecommute, or work from a distance.

The use of Internet services over the Internet has greatly speeded communication among companies, coworkers, and individuals. Media and entertainment companies use the Internet to broadcast audio and video, including live radio and television programmes; to offer online chat and to offer online news and weather programmes. Scientists and scholars use the Internet to communicate with colleagues, to perform research, to distribute lecture notes and course materials to students, and to publish papers and articles. Individuals use the Internet for communication, entertainment, finding information, and to buy and sell goods and service.

A variety of Internet tools have been developed over the years to make effective use of these services. These include Telnet, Archie, Gopher, WAIS and the World Wide Web (WWW). A few of the services have been discussed in brief in this unit.

1)   **Electronic Mail (E-mail)**

Electronic Mail, or e-mail, is a fast, easy and inexpensive way to communicate with other Internet users around the world. It is the most popular and widely used service of the Internet. It is one of the fastest, most economical and convenient methods of communication. Most of the good libraries worldwide have e-mail addresses and for many people, especially scientists and researchers, e-mail is the main reason for getting on to the Internet or an online service.

E-mail can be used not only to exchange correspondence with friends, but to transfer documents, obtain electronic copies of books, subscribe to electronic news services or journals, and to obtain information about anything that is stored on a computer. We can even search databases using e-mail. It is possible to attach different kinds of files such as MS Word, Word Perfect, Windows WordPad, Notepad or spreadsheet files to our e-mail.

Basic E-mail functions

There are a few basic functions in e-mail, and almost all mailers handle them. These

are:

— **Read**

— **Compose** (new messages)

— **Reply** (to messages you've received)

— **Forward** (messages you've received)

— **Refile** (save the message away)

— **Delete**

Other handy functions allow you to work more efficiently:

— **Include or attach** (remember human names for common addresses)

— **Sort you mail** by any given category; sender, date, subject and so on.

One of the more notable conveniences is the speed with which e-mail travels. E-mail messages travel on high-speed networks and usually arrive at their destination within minutes of being sent. As a result, Internet users have come to refer to the regular postal service as *snail mail* when comparing it to the speed at which e-mail travels. Unlike a paper letter, e-mail can be stored in file form on a computer, which in turn enables you to pull it into a word processor for editing and printing. E-mail can be forwarded easily to another person without the inconvenience and cost of making a paper copy, stuffing it in an envelope, affixing a stamp, addressing it, and taking it to the post office. Further, one e-mail message can be directed easily to multiple recipients - no more printing out multiple paper copies and mailing each one separately.

E-mail Address

A popular point of entry for many Internetters is electronic mail. Each person who uses e-mail receives an e-mail address and mail box. This helps in sending and receiving mail and to retrieve mail from the server. Addresses vary from site to site in their appearances. An email address is easy to spot. It always has that at symbol (@) in the middle of it. For example, we know at a glance that ravilss@hotmail.com is an email address. In most email addresses, everything following the @ symbol is the domain address of a company, ISP, or other organization. The part before the @ is the name or user ID of a particular organization or user. In the address chadha@sansad.nic.in, the chadha identifies the individual, sansad identifies the machine the account is on, the nic identifies the domain, or institution where the machine is located, and in identifies the country.

2) **Listservs**

E-mail also provides a platform for sharing information quickly, and on a grand scale by using listservs. Listservs are electronic groups that typically centre around a broad topic such as Digital Libraries or Reference Service, etc. Listservers of IFLA with the name IFLA-l, Digilib-l, LIBJOBS-l are good example of this. Every e-mail message sent to the listserv is distributed to all members of that listserv, which is potentially hundreds or thousands of people. It does not cost anything to subscribe to a listserv, but simply requires that the user send an e-mail message to the appropriate address with the message: subscribe (listserv) Firstname Lastname. Each Listserver has one address where a user sends requests to subscribe, un-subscribe, search the

archives, etc., and another address to send actual questions or responses to the readers of the list. Apart from organising discussions, job announcements and conference announcements are popular usage of Listservs.

3) Newsgroups

Another Internet service similar to listservs are newsgroups. News Groups are like an international bulletin board. Each group is a forum for a different subject, where you can post questions and answers. There are many thousands of groups covering just about any area of interest. The difference between the two is that when you join a group, the mail is no longer automatically deposited into your mailbox. You must go to the newsgroup yourself to read it. Some listservs can also be accessed as a newsgroup. A good analogy to a newsgroup is a bulletin board you go to it, as opposed to having mail delivered to your desk. The benefit is that your mailbox does not get as cluttered (some groups post 5-20 messages per day) and you can more easily regulate how often you read the messages. A drawback is that you have to remember to go out and look for the information.

There are hundreds of newsgroup communities. They centre around topics such as computing, news, recreation, social, and "alternative" topics. Each newsgroup name begins with a code that identifies the type of newsgroup that is. For instance:

comp.human-factors is a newsgroup dealing with the human factors of computing.
soc.college.teaching-asst is a social newsgroup for college teaching assistants
alt.fan.jimmy-buffett is an alternative newsgroup for fans of Jimmy Buffet
clari.biz.market.otc a business newsgroup for counter stock market exchange.

There are newsgroups dealing with virtually every topic under the sun (although new groups appear every day). The only problem you may encounter is that is up to whoever provides your Internet access to determine which newsgroups will be made available on your system (this is comparable to the problem encountered with basic cable television.) The decision is often based on system capability, but may also be an issue of censorship. Some of the "alt" groups are truly for adult audiences.

4) Usenet

Usenet is a collection of more than 8,000 newsgroups, or discussion groups, on every conceivable subject. For example, some newsgroups are self-help groups for victims of cancer or sexual abuse, and others give the latest in gossip about show business personalities. Anyone can contribute a message, called an article, to a Usenet newsgroup or post a reply, known as a follow-up post, to an existing article. With the aid of a newsreader (a programme designed to access Usenet newsgroups), you can read an entire threat-all the replies to an interesting article.

The system is intended for exchange of information in an informal way. Anyone can post new messages to the group and reply to other messages. News groups are arranged in a loose hierarchical order covering about 5,000 subjects. About half of these are related to computing, the rest are for recreational subjects, professional discussion and trivia. To use Usenet you need a news viewer and access to an NNTP server. Most Internet service providers have such a server, as do most large academic institutions.

5) Remote Login

Remote Login is the ability of a computer user in one location to establish an on-line connection with another computer elsewhere. Once a connection is established with

a remote computer, users can use that remote system as if their computer were a hard-wired terminal of that system. Within the TCP/IP protocol suite, this facility is called "Telnet". Utilising Telnet, an Internet user can establish connections with a multitude of bibliographic databases (primarily library catalogues), campus information systems of various universities, full-text databases, data files (e.g., statistics, oceanographic data, meteorological data, geographic data, etc.), and other on-line services. Many of these systems are available for any Internet user to access and use without an account.

What makes this application truly remarkable is that ease and speed of access are not dependent upon proximity. An Internet user can connect to a system on the other side of the globe as easily as (and generally not much slower than) he or she can connect to a system in the next building. In addition, since many Internet users are not at present charged for their network use by their institutions, or at least are not charged by the level of their use, cost is often not a significant inhibit. For example: CARL -Colorado Alliance of Research Libraries operates a public access catalogue services. Offered are a number of library databases including searches for government periodicals, book reviews and access to other library databases around the country.

6) **Telnet**

Telnet is the service of the Internet that allows you to access remote computers outside your area. Many computers on the Internet are set up to allow Telnet access. Some require login names and passwords, but many do not have any restrictions. Telnet is a powerful Internet application that enables librarians to access other libraries' online catalogs worldwide. Accessing other libraries' catalogues can help you verify information for acquisitions, inter-library loan, and copy cataloguing. Some systems, such as CARL's UnCover (database.carl.org), enable you to search magazine article indexed online. The original idea behind Telnet was to let researchers from different institutions share resources with one another. Telnet allows remote login to host computers, and commonly is used to connect to electronic catalogues and databases at near and distant libraries. Hytelnet is a tool that helps you access the various sites through Telnet. The World Wide Web lets you access sites through Telnet and use FTP to retrieve documents you find.

Telnet, like most other services, is based on the client/server model. The client programme, running on your machine, initiates the connection with a server programme, running on a remote machine. Keystrokes are passed from your terminal directly to the remote computer just as though they were being typed at a terminal on the remote computer. Output from the remote computer is sent back and displayed on your terminal.

Many sites offer hytelnet, a programme that allows a user to look through a menu of options by type of system (library catalogue, gopher, etc.). Once the user has selected the site, hytelnet provides the telnet address, logon and logoff instructions for the remote site, and in some cases, will automatically connect the user to the site when prompted.

When telneting to other sites, you must be prepared to read help screens, and experiment. Many librarians have mastery over their local OPAC as well as some commercial services. However, there is no formal mechanism for mastering the ins and outs of the variety of OPACs, information systems, and gophers that exit in the great world of the Internet. Possible failings of opacs must also be considered: Certain

special collections may not be included, material obtained before a certain date may not be included, or level of cataloguing quality may vary from site to site, etc.

7) **File Transfer (FTP)**

File transfer is one of the most frequently used Internet applications, enabling you to copy files from over a thousand different archives around the world. You can think of these archives as libraries-electronic libraries housing digitized information. Although information in traditional paper-based libraries is stored in books and magazines, the file is the unit of storage in electronic libraries. These files hold such things as text, images, sound, and executable programmes.

FTP or File Transfer Protocol is used to transfer files between computers on the network. There are two types of FTP connections, anonymous and non-anonymous. Some sites enforce a strict FTP authorization that prohibits you from accessing files until you obtain a login name and password for their computer. Many other sites, the ones you will most often be connecting to, allow anonymous FTP, which provides unrestricted access to public files. Anonymous FTP access means that you don't have to be a registered user to connect to the remote host. You connect to an anonymous FTP server, then you would use anonymous as your login name and your e-mail address as a password. Non-anonymous, you will need a private login name and password. There are also two types of file transfers. ASCII and binary. ASCII is for text transfers only. Binary transfers are for transferring anything else. If in doubt, use binary (bin).

The set of conventions that govern how this file transfer process takes place on the Internet is called File Transfer Protocol, or FTP. It transfers files from one computer to another computer. Typically, FTP is used to transfer shareware, software upgrades, reports and other data. Types of files that can be transferred using FTP include virtually every kind of file that can be stored on a computer: text files, software programmes, graphic images, sounds, files formatted for particular software programmes (e.g., files with word processing formatting instructions), and others. Many computer administrators have set aside portions of their machines to offer files for anyone on the Internet to retrieve. The FTP protocol also enables you to list directories on remote machines and to transfer files in either direction. When a file is transferred, it is not actually moved; rather, the file is copied from one machine to another. The computer from which the file is copied is sometimes referred to as the source host and the computer to which the file is copied is called the destination host.

Like most of the Internet applications, the FTP is also based on a client/server model. You run a programme on your local machine called an FTP client, which in turn connects to another programme running on a remote machine called a server.

In order to provide a reasonable level of performance and to avoid overloading their system, FTP sites install user limits for anonymous FTP. When one site is busy, one or more archive sites around the world are usually available that can provide the same files. These archives are referred to as mirror sites.

The following procedure can be used for downloading a file from a remote server:

1) The remote computer's Internet address (either the domain name address or the IP address);

2) The pathname that tells you the path of directories you must follow to get from the root directory to the directory where the file resides; and

3) The file name itself.

8) **Gopher**

Gopher is an Internet system developed at the University of Minnesota (land of the Golden Gophers), which allows a site to create an invisible tunnel from itself to other sites. Gopher was originally developed in April of 1991 by a team of experts including Bob Alberti, Farhad Anklesaria, Paul Lindner, Mark McCahill, and Daniel Torry at the University's Microcomputer and Workstation Networks Center. The originators define Gopher as "a software following a simple protocol for burrowing (as a gopher does) through a TCP/IP Internet. The protocol and software follows a client-server model." Once on a gopher, you can work through a menu system to connect to other gophers that exist in North America and throughout the world. When ultimately connected to another gopher, the menu looks the same as the home, or initial gopher. There is no prompt to logon. Instead, all of the resources available on the remote gopher appear to reside at the point where you began your search.

The Internet Gopher is a simple protocol for building distributed information systems and organising access to Internet Resources. It uses a series of menus to organise and automate access to information and other on-line system wherever they reside on the Internet. The Internet Gopher client software presents users with a virtual information matrix (gopher space) that they can navigate by either browsing a hierarchical arrangement of items, or search by submitting queries. For browsing in gopherspace, the gopher client software presents the user with a list of items from which the user selects an item of interest. User can read or access an item without having to worry about domain names, IP addresses, etc. For example if you want to access the on line library catalogues at the University of California, rather than looking up the address and telnetting to it, you find an entry in a menu and select it. The gopher then "goes fer" it. You may be even able to arrange inter-library loans if library permits it through on-line catalogue. Alternatively a user can select an item where a query can be put. User is prompted for what works to search for and a full text search is done by the server.

The resources on a gopher system vary from site to site. Generally gophers include full text documents (such as speeches, reference books, news items, subject bibliographies, campus publications, etc.), campus or institutional information for that location (class schedules, campus events, lists of campus organisations, job postings), access to library resources and databases, and access to systems other than itself. The resources are found by working through the menu system. Some gopher sites even allow the user to retrieve documents found at that site in a very easy manner. Even if your Internet connection is not a gopher site itself, it is possible to telnet to any number of gophers and use that gopher to tunnel to other gophers.

9) **Archie**

Archie is a file searching system that was developed at McGill University of Canada. It was originally formed to be a quick and easy way to scan the offerings of the many anonymous FTP sites that are maintained around the world. Currently it indexes about 12000 servers and over 2 million files. The Archie service is accessible through an interactive telnet session, e-mail queries and command line and X-window

clients. Each Archie server, independent of the others, collects data and stores it for future use. The data it collects is retrieved from anonymous FTP sites. Each Archie server stays in contact with anonymous FTP archive sites all around the world and maintains an up-to-data composite index of their holdings. Archie servers do not collect and store the actual files themselves; instead, they index only the directory names and file names.

Finding a file on the Internet by simply browsing through he directories of hundreds of FTP sites is physically impossible. Herein lies the purpose for learning how to use Archie: Archie servers provide a search mechanism for locating files. Anyone who has access to the Internet can search the indexes. Search results tell you which FTP server stores a particular file or directory and what the pathname is for locating it. Although each Archie server is a separate entity, they all perform the same service and even index close to the same FTP sites.

A query can be put to Archie to find filenames which contain a certain search string or whose description contains a certain word. It returns the actual filenames that meet the search criteria, and the name of servers containing those files. Then you can retrieve the desired file using FTP.

10) **Veronica**

Veronica and Jughead are named after the Archie of comic book fame. Veronica attempts to index the world of Gopher servers, much like Archie attempts to build a comprehensive index of all anonymous FTP servers. A central server periodically scans the complete menu hierarchies of Gopher servers appearing on an ever-expanded list. The resulting index is provided by the Veronica server and can be accessed by Gopher servers.

Veronica usually is a menu item on a Gopher or World Web server. There might be several Veronica sites available on a given Gopher or WWW server. This variety enables a user to select a not-too-busy Veronica server. After the user selects a Veronica, Gopher asks the user for their search words. A successful search formed the Veronica returns a new menu of Gopher items satisfying the conditions of the search.

The search engine supports sophisticated search strings specifically to target the desired information. The Veronica search engine permits limited Boolean and substring searches. Boolean searches enable the user to include and, or, and not, along with a hierarchy of parentheses, to control the search. Thus, if the search were the following, all references including both smoking and drinking would be included. Veronica is one of the most useful Internet information-hunting tools around. Unfortunately, because it's also the one that has the fewest servers, reaching a Veronica server can be very difficult.

11) **Wide Area Information Search (WAIS)**

WAIS (pronounced WAYS) stands for wide area information servers. Real-time connections are used to connect to other WAIS servers and to search their indexes. In combination with the World Wide Web, WAIS can be used as a back-end search engine. WAIS relies on indexed data collections, or libraries. These libraries are the collections that consist mostly of informational material. If molecular biology is your meat and drink, for example, several journals on the subject are available online through WAIS libraries. These data collections generally have been indexed

and made available by volunteers at academic sites. Commercial WAIS libraries, such as the Dow Jones Information Service, also are available.

WAIS was developed by Thinking Machines in collaboration with Apple Computer. Dow Jones, and KPMG Peat Marwickl had developed a generalized retrieval system for accessing data around the world, i.e., to provide a common interface to a multitude of Internet databases. WAIS works in two parts.

- The server side, which indexes WWW sites periodically and maintains this index for access by WAIS and WWW clients.

- The client side, which includes WAIS clients and WWW clients.

WAIS is a distributed information service that takes its input in a simple natural language format, offers indexed search for fast retrieval, and has a "relevance feedback" mechanism to permit the results of initial searches to bias future searches. WAIS let you search through Internet archives looking for articles based on keywords. It is really a tool for working with collections of data or databases. WAIS can select information looking at an index and present it to you regardless of its format. IT is similar to gopher in use but the difference is that you do not have to go through various menus. It does a complete search for you. A WAIS command is essentially: "find me items about this in that library". WAIS looks at all the documents in the library or libraries you gave it and tells you which documents are most likely to contain what you want. If you wish, WAIS then displays the documents for you. There are more than 500 free WAIS libraries on the network now.

WAIS search engines also support Boolean queries, supporting syntax such as "freed and green not blue," which would get all items containing both red and green, excluding all records with only blue. Partial requests, such as "hum*" might show "hum", "hummingbird," "humid." For users who don't know how to spell their search term, WAIS can optionally use the soundex and phonix search algorithms to "guess" correct matches.

How WAIS works? Wais is a distributed text searching system. It is based on a standard Z39.50 (a draft ANSI standard for requesting bibliographic information) that describes a way for one computer to ask another to do searches for it. When you request a search from a WAIS client, it contacts the server that handles the libraries/databases you suggested. It asks each server in turn, to search its index for a set of works. The server then sends you a list of documents that may be appropriate. WAIS has one really unique feature called "Relevance Feedback". It allows you to find articles that are similar to the articles you have already found.

12) **Finger**

Finger is an old Unix utility to examine the user login file and learn something about the user. It helps to know if the user is logged in to the system at that very moment. If the user has configured his account for interrogation by outside users, it can also reveal the user's e-mail address and other information. The command structure for finger is as follows:

Finger username@hostname

The username is optional and requests specific information about a user whose first or last name matches the username on the hostname system. If you are using the user's login name, the name match is case-sensitive, meaning "joe" is different from "Joe." If you are using the user's actual name (such as the GECOS file of/etc/password

of Unix systems), the search is case-insensitive. To be safe, you can use all lowercase and will find a match most of the time.

13) **WHOis**

Whois is several things: It is the name of a particular white pages directory, it is a general directory, and it is an application that accesses these directories. Using it, you can find email addresses, host computer names and domain names. Organizations and users register their names, host computers and organization domain names in a large registry database.

The original directory, or white pages, was maintained by the Defense Data network (DDN) at the Network Information Centre (NIC) and contained approximately 70,000 entries at its peak. The original listees included anyone with responsibility for running Internet networks,. Eventually, anyone who filled out a form could register into this database.

When the original network, ARPAnet, was decommissioned, the maintenance of the white pages and the registration was transferred to the InterNIC, the new Internet Information provider to maintain the Whois directory for non-military network and domain contacts. The military Whois directory is maintained at nic.ddn.mil and the non-military Whois directory is maintained by ds.internic.net.

14) **Messaging**

The messaging (also called *electronic messaging)* is the creation, storage, exchange, and management of text, images, voice, telex, fax, e-mail, paging, and Electronic Data Interchange (EDI) over a communications network.

In programming, messaging is the exchange of **message** (specially-formatted data describing events, requests, and replies) to a **messaging server,** which acts as a message exchange programme for client programmes. There are two major messaging server models: the point-to-point model and the publish/subscribe model. Messaging allows programmes to share common message-handling code, to isolate resources and interdependencies, and to easily handle an increase in message volume. Messaging also makes it easier for programmes to communicate across different programming environments (languages, compilers, and operating systems) since the only thing that each environment needs to understand is the common messaging format and protocol. IBM's **MQSeries** and Sun Microsystems Java Message Service (JMS) are examples of products that provide messaging interfaces and services.

15) **Instant Messaging**

Instant messaging (sometimes called IM or IMing) is the ability to easily see whether a chosen friend or co-worker is connected to the Internet and, if they are, to exchange messages with them. Instant messaging differs from ordinary **e-mail** in the immediacy of the message exchange and also makes a continued exchange simpler than sending e-mail back and forth. Most exchanges are text-only. However, some services allow attachments.

AOL first popularized instant messaging and, with its large membership, is likely to be the most popular service for some time to come. AOL's Instant Messenger can be used by AOL members but there is no requirement to be connected to the Internet through AOL. **An open source** alternative to AOL's instant messaging is called **Jabber.**

In order for IMing to work, both users (who must subscribe to the service) must be online at the same time, and the intended recipient must be willing to accept instant messages. (It is possible to set your software to reject messages.) An attempt to send an IM to someone who is not online, or who is not willing to accept IMs, will result in notification that the transmission cannot be completed. If the online software is set to accept IMs, it alerts the recipient with a distinctive sound, a window that indicates that an IM has arrived and allowing the recipient to accept or reject it, or a window containing the incoming message.

Under most conditions, IMing is truly "instant." Even during peak Internet usage periods, the delay is rarely more than a second or two. It is possible for two people to have a real-time online "conversation" by IMing each other back and forth. Once in a while, a person might receive an IM from someone while already engaged in a chat with someone else, and decide to carry on IM chats with both people independently and concurrently. This requires mental alertness to avoid the embarrassment of sending one IM companion a message intended for the other.

16) **Web Chat**

The Internet provides facility for a real time (on-line) communicating with one to one or one to many computers. This is popularly known as chat. On the Internet, chatting is talking to other people who are using the Internet at the same time you are. Usually, this "talking" is the exchange of typed-in messages requiring one site as the repository for the messages (or "chat site") and a group of users who take part from anywhere on the Internet. In some cases, a private chat can be arranged between two parties who meet initially in a group chat. Chats can be ongoing or scheduled for a particular time and duration. Most chats are focused on a particular topic of interest and some involve guest experts or famous people who "talk" to anyone joining the chat. Transcripts of a chat can also be archived for later reference.

Chats are also conducted on online services (especially America Online), by bulletin board services, and by Web sites. Several Web sites, notably Talk City, exist solely for the purpose of conducting chats. Some chat sites such as Worlds Chat allow participants to assume the role or appearance of an avatar in a simulated or virtual reality environment. One to one chat is commonly used to communicate with friends, relatives or co-workers. A chat can also be conducted using sound or sound and video, assuming you have the bandwidth access and the appropriate programming.

A large number of chat software like freetel, talkcity, netvoice, etc. are now available for transmitting voice, data pictures on the internet. The Internet Roundtable Society has developed WebChat, a real-time multimedia chat application for the Web. Users can incorporate images, video, audio, and hot links into their chat. Working with standard browsers and no additional special software, WebChat can be used for businesses that want to attract lots of visitors to their websites for marketing, real-time customer support, direct sales, corporate teleconferencing, and consumer education. You can try it out at the company's WebChat Broadcasting System.

Another popular programme is ICQ, short for 1 Seek You. The software can be downloaded for free from the ICQ website. It allows registered users to set up both public and private chat sessions, send written and voice messages and files. A useful feature is the ability to save the text generated by chat sessions. The ICQ website contains a comprehensive set of tutorials that explain how this sophisticated software works.

Genesis.Jive is another example of a Web-based chat system now being used on The GLOBE, an entertainment site. Developed by WebGenesis, this software allows parties in remote locations to Interact in real-time. It can also be used to conduct online classes or provide technical support. Users can share graphics during a session and automatically save transcripts of conversations.

17) **Internet Relay Chat (IRC)**

Internet Relay Chat (IRC) allows you to meet online with people of similar interests. IRC consists of a large number of channels (a channel is like a room where messages are exchanged) that contain group conversations. Channels can be created by anyone, and users can join channels and exchange messages. It is a text based mode of communication. Users often use cryptic codes and abbreviations to signify phrases of speech, so they do not have to type whole words.

Channels are located on IRC servers, which often talk to each other and share channels, so that messages on one server are relayed to the same channels on other IRC servers. To access IRC requires the use of an IRC client. One such client is available from http://www.mirc.co.uk/

IRC provides real-time communication with others on the Internet. Popularly known throughout the Internet as IRC, Internet Relay Chat offers the ability to communicate interactively with many thousands of people around the world about nearly any subject imaginable. Unlike any other Internet service, IRC enables you to hold a spontaneous conversation, in real time, with a group of individuals who may never physically meet.

In most cases there are no constraints on the topics that may be discussed. Topics range from the serious (an oft-mentioned example is the successful use of IRC to relay information into and out of Russia during the 1993 Boris Yeltsin coup; it was also used in the 1994 and 1995 earthquakes in California and Kobe, Japan), to the inevitable computer-speak, to the recreational topics of international society, such as sports, movies, books, gossip about the rich and famous, and political debate. Topics may be controversial or mainstream at the whim of the participants, and anyone can create a topic.

Internet Relay Chat works as a client/server application, and the IRC servers are interconnected to form, from the user's perspective, one large service. Each IRC user must use a client programme on their Internet host. You instruct the client to connect to at least one IRC server, preferably a server nearby to minimize network delays and inefficiencies. Internet Relay Chat clients are available for Microsoft Windows-equipped PCs with Winsock support, the Macintosh, IBM mainframes running VM, most UNIX systems, and VMS with a TCP/IP package. IRC clients must run on a computer attached directly to the Internet.

If you try to initiate a Talk session with someone, your screen will divide in half and the Talk daemon will tell you that it's trying to reach your friend. If the person is on a different host, you will see a note that the daemon is trying to talk to a remote daemon.

18) **Audio and Video Streaming**

Streaming audio and video is the ability to download video and audio from the Internet and have this play without having to wait for the entire file to be downloaded. This means streaming audio and video can support "live" events, and viewers can

watch or listen to events as they occur. Streaming audio and video requires special servers and software to be installed on your computer.

18A. **Streaming Audio**

The major problem associated with the traditional formats of waveform audio (such as wave files) is the file size. When you use a waveform audio file on a web page, the browser must download the entire file before it can be played.

For files of reasonable size, this leads to long delays for the receiver. Streaming audio addresses the issue of this delay by allowing the file to be played whilst it is still being downloaded. Special players are required for viewing and playing streaming audio clips.

Streaming audio is primarily associated with Real Networks. "Real Player" software supports the playing of streaming audio over the Internet. The advantages of using streaming audio are:

i) the use of a special codec that reduces the bandwidth requirements for audio file down loading

ii) instant playing of the audio clip

iii) suits limited bandwidth connections (i.e., a dial-up user with a 28.8Kbps modem can even make use of the facility effectively).

When a user selects a streaming audio clip, the streaming audio player contacts the server and establishes the bandwidth and delays between the user computer and the host server. A few seconds of the clip is then downloaded into temporary buffer storage before the clip begins to play. As the clip begins to play, more content is downloaded and placed into temporary storage. If there is a congestion delay in the Internet connection, the clip continues to play using the contents of the buffer storage. Hopefully, before the buffer storage is exhausted, the congestion delay will cease and thus the buffer storage will fill up again.

In this manner, the streaming player tries to achieve smooth playback of the clip. Without using storage, gaps or breaks in the audio would be evident when a congestion delay occurred. A special content provider programme is necessary for producing streaming audio clips. A limited free version can be downloaded from the Real Networks web site. In addition, to provide streaming audio clips from a server requires the use of a streaming server.

18B. **Streaming Video**

Streaming video is a sequence of "moving images" that are sent in compressed mode over the Internet and displayed by the viewer as they arrive. Streaming media is streaming video with sound. With streaming video or streaming media, a Web user does not have to wait to download a large file before seeing the video or hearing the sound. Instead, the media is sent in a continuous stream and is played as it arrives. The user needs a player, which is a special programme that decompresses and sends video data to the display and audio data to speakers. A player can be either an integral part of a browser or downloaded from the software maker's Web site.

The major streaming video and streaming media technologies include RealSystem G2 from RealNetwork, Microsoft Windows Media Technologies (including its Net Show Services and Theater Server), and VDO. Microsoft's technology offers

streaming audio at up to **96 Kbps** and streaming video at up to **8 Mbps** (for the NetShow Theater Server). Streaming video is usually sent from prerecorded video files, but can be distributed as part of a live broadcast "feed." In a live broadcast, the video signal is converted into a compressed digital signal and transmitted from a special Web server that is able to do multicast, sending the same file to multiple users at the same time. The advantages of using streaming video are:

i) the use of a special codec that reduces the bandwidth requirements for video

ii) instant playing of the video clip

iii) suits limited bandwidth connections (i.e., a dial-up user with a 28.8Kbps modem)

19) **Conferencing**

One of the most existing feature of the Internet is to communicate, talk, and see the groups of people in different locations around the world, without the expense of travel. Conferencing can take many forms, such as web chat, audio conferencing, video conferencing, multimedia conferencing, screen-sharing, etc. The conferencing programmes, such as the popular CU-SeeMe, allow workgroups to use the Internet to see each others' faces in small windows on the computer screen and to hear their voices through computer speakers. You can use the desktop video and audio simultaneously, use the audio alone, or just use the screen-sharing capability without either audio or video.

19A. **Audio Conferencing (Telephony)**

Telephony is normally associated with the transmission of voice between distant parties using a handheld device containing both a speaker or transmitter and a receiver. However, we believe that telephony does connote voice or spoken and heard information predominately. It is a temporarily dedicated point-to-point (rather than a broadcast) connection. With the transmission of digital signals over the Internet, it is now possible to have the following services.

- The ability to make a normal voice phone call (whether or not the person called is immediately available; that is, the phone will ring at the location of the person called)

- The ability to send fax transmissions at very low cost (at local call prices) through a gateway point on the Internet in major cities.

- The ability to leave voice mail at a called number.

Some companies that make products that provide or plan to provide or plan to provide these capabilities include: IDT Corporation (Net2Phone), Netspeak, NetXchange, Rockwell International, Vocal Tec. And Voxspeak.

19B. **Videoconferencing**

Videoconferencing is one of the most exciting areas of development in telecommunications, with applications ranging from business to government to education to home and family. Videoconferencing involves sending video signals as well as audio and computer data signals. Conferencing can be done one-to-one, one-to-many (called multicast), and many-to-many (called multipoint). While video conferencing we are able to talk as well as see the people sitting miles away as if all are discussing in one room facing each other. One of the most popular applications

is transmission on news from various locations by the TV news channels. It also has enormous potential for enhancing communications for small and mid-sized companies, as well as distance learning. However, because at the very minimum you need an ISDN connection for acceptable video quality, entry barriers are still high. As a result, desktop conferencing is not yet widely implemented for business and educational use. As bandwidth increases, we expect to see desktop videoconferencing blossom. The PictureTel and Vtel are two of the largest companies that sell videoconferencing equipment.

20) **The Virtual Corporation**

A virtual corporation can be defined as a temporary collaborative network of a group of professionals such as creative designers, manufacturers, suppliers, customers or other experts in marketing and finance linked by an interactive multimedia network to share skills, costs, production facilities, resources, and access to each other's markets.

The virtual corporation implies a competitive advantage or even survival in an increasingly interdependent global economy. It represents the ability to develop and deliver a real produce or service any time, any place, in any quantity but also faster than any competitor. Associated with this approach is also the need for a drastic change in the mindset of business executives and workers of traditional enterprises. They must with their marketing perspectives and work habits from mass production to mass customization of goods where customer service is the key driving factor.

21) **Collaborative Multimedia Computing**

Collaborative computing allows users to work together on a document or project, usually in real time, by taking advantage of underlying network communication systems. A number of software have been developed for collaborative computing allowing people to work together over networks. A few examples could be:

Application suits such as Microsoft Office and Exchange, Lotus Notes, and Novell Groupwise that provide messaging, scheduling, document coauthoring, rules-based message management, workflow routing, and discussion groups.

Videoconferencing applications that allow users to collaborate over local networks, private WANs or over the Internet. Internet collaboration tools that provide virtual meetings, group discussions, chat rooms, whiteboards, document exchange, workflow routing, and many other features. Multicasting is an enabling technology for groupware and collaborative work on the Internet that reduces bandwidth requirements. A single packet can be addressed to a group, rather than having to send a packet to each member of the group.

Self Check Exercise

1) Distinguish between web server and web browser.

2) Distinguish between Search engine and Metasearch engine.

....

....

....

12.4   SUMMARY

It is important to know the various information resources and services available in Internet or World Wide Web (www). The unit explained in detail the importance and functioning of World Wide Web, web servers, web browsers, mark-up languages such as SGML, HTML, XML, VRML. It is equally important to know about search engines, types of search engines, selecting and searching from a particular search engine including meta search engine are the added features which are discussed in detail.

12.5   ANSWERS TO SELF CHECK EXERCISES

1) A *computer* that delivers (serves up) *Web pages.* Every Web server has an *IP address* and possibly a *domain name.* For example, if you enter the *URL http:/www.pcwebopedia.com/index.html* in your browser, this sends a request to the server whose domain name is *pcwebopedia.com.* The server then fetches the page named index.html and sends it to your browser.

    Web browser, a *software application* used to locate and display Web pages. The two most popular browsers are *Netscape Navigator* and Microsoft *Internet Explorer.* Both of these are graphical browsers, which means that they can display graphics as well as text. In addition, most modern browsers can present *multimedia* information, including sound and *video,* though they require plug-ins for some formats.

2) *A program* that searches documents for specified *keywords* and returns a list of the documents where the keywords were found. Although *search engine* is really a general class of programs, the term is often used to specifically describe systems like *Alta Vista and Excite* that enable users to search for documents on the *World Wide Web* and *USENET newsgroups.*

    A *search engine* that queries other search engines and then combines the results that are received from all. In effect, the user is not using just one search engine but a combination of many search engines at once to optimize Web searching. For example, Dogpile is a metasearch engine.

12.6   KEYWORDS

Adobe Acrobat	:	Programmes developed by Adobe Systems, Inc for creating and distributing electronic documents.
Hyper media	:	An extension to hypertext that supports linking graphics, sound, and video elements in addition to text elements.
Hypertext	:	A special type of database system in which objects (text, pictures, music, programmes and soon) can be creatively linked to reach other.
Shockwave	:	A technology developed by macromedia, Inc that enables web pages to include multimedia objects.

Streaming	:	A technique for transferring data such that it can be processed as a steady and continuous stream.
Vector Graphics	:	Same as object-oriented graphics, refers to software and hardware that use geometrical formulas to represent images

12.7   REFERENCES AND FURTHER READING

Ablan, J.(1996). Web site administrator's survival guide. Indianapolis: Samsnet,. Arents, Hans C. Search engines, 1996.

Balas, J. Beyond Veronica and Yahoo! More Internet search tools. Computers in Libraries, 16(3) Mar.96, p.34-5, 38. 1996.

Bates, M.E. The Internet: part of a professional searcher's toolkit. Online, 21(1) Jan/Feb, 1997.

Beayulieu, M. Experiments on interfaces to support query expansion. Journal of Documentation, 53 (1), 1997.

Brandt, D.S. Relevancy and searching the Internet. Computers in Libraries, 16(8) Sep 96, 1996

Cole, C and Carter, S. Interview and online searching for undergraduates doing a social science assignment. Information Processing and management, 32(6) Nov, 1996.

Collins, B.R. Web Watch. Library Journal, 122(1) Jan. 1997.

Dong, L T Su, LT. Search engines on the World Wide Web and information retrieval from the Internet: a review and evaluation. Online and CD ROM Review, 21 (2) Apr. 97.

Eager, Bill and Antony, Tobin. Using the World Wide Web. Indianapolis, IN: Que, 1996.

Falk, H. World Wide Web search and retrieval. Electronic Library, 15(1) Feb, 1997.

Harrison, S. NHSWeb: a health intranet, Aslib proceedings, 49 (2), 1997.

Inside the World Wide Web, USA, New Ridwers Publishing, 1996.

Iryine, Martin. Web Works., New York: W.W. Norton, 1997.

Jain, V.K. Internet & Web Page Design. New Delhi, BPB Publications, 2000.

Koutnik, C. The World Wide Web is here: is the end of printed reference sources near? R_O, 36(3) Spring, 1997.

Liu, Jain. Understanding WWW search tools, 1996.

Watson, I. Searching the Web: a quick comparison of Infoseek, Lycos, Yahoo, Infor Market, and Webcrawler. Managing Informations, 3 (1) Jan, 1996.

Winship, Ian R. World Wide Web searching tools; An evaluation, 1995.

Gopi Garge and H. Krishnamurthy, Internet as a business tools, BIZNET, December 1996, pp 8-11 .

Marcus J Ranum, Internet Firewalls Frequently Asked Questins, downloaded from http:/www.clarke.net/pub/mir

Sharma, Aashish; Yurcik, William , The Gyandoot Digital Library Intranet. Bulletin of the American Society for Information Science , Apr/May2001, Vol. 27 Issue 4, Pp23-27.

Sharmila Shankar and Md. Atiq, E, E-business, Counting on Extranets, it magazine, pp 38-42.